TIME-LAPSE INVERSE THEORY

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF GEOPHYSICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Musa Maharramov
February 2016

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Biondo L. Biondi) Principal Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Mark D. Zoback)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Jon F. Claerbout)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Stewart A. Levin)

Approved for the University Committee on Graduate Studies

Abstract

Compaction in reservoir overburden can impact production facilities and lead to a significant risk of well-bore failures. Prevalent practice of time-lapse seismic processing of 4D data above compacting reservoirs relies on picking time displacements and converting them into estimated velocity changes and subsurface deformation. This approach relies on prior data equalization, and requires a significant amount of manual interpretation and quality control. In this thesis I develop methods for automatic detection of production-induced subsurface velocity changes from seismic data, and computational techniques of subsurface characterization from measurable surface deformation.

In the first part of my work I describe a time-lapse inversion technique based on a simultaneous regularized full-waveform inversion of multiple surveys. I provide a theoretical foundation of the proposed method, and analyze its sensitivity to a realistic 1-2% velocity deformation in the overburden. The method is applied in a study of overburden dilation above the Gulf of Mexico Genesis field and achieves a stable recovery of "blocky" negative velocity anomalies above compacting reservoirs. I propose a multi-scale extension of the method for recovering both long and short-wavelength velocity changes.

In the second part I describe a geomechanical model of overburden deformation in response to fluid extraction or injection. I provide a method for inverting pore pressure changes from noisy and sparse measurements of surface deformation. The method is applied to estimating the efficiency of Cyclic Steam Stimulation (CSS) of a heavy oil reservoir. Inverted subsurface pore pressure changes indicate a significant

heterogeneity of the propagating steam fronts. The method is extended to a study of sharp contrasts in reservoir attributes by using a new splitting algorithm for solving large-scale constrained regularized optimization problems.

Preface

The electronic version of this report[1] makes the included programs and applications available to the reader. The markings **ER**, **CR**, and **NR** are promises by the author about the reproducibility of each figure result. Reproducibility is a way of organizing computational research that allows both the author and the reader of a publication to verify the reported results. Reproducibility facilitates the transfer of knowledge within SEP and between SEP and its sponsors.

ER denotes Easily Reproducible and are the results of processing described in the paper. The author claims that you can reproduce such a figure from the programs, parameters, and makefiles included in the electronic document. The data must either be included in the electronic distribution, be easily available to all researchers (e.g., SEG-EAGE data sets), or be available in the SEP data library[2]. We assume you have a UNIX workstation with Fortran, Fortran90/95, Fortran 2003/2008, C, C++, X-Windows system and the software downloadable from our website (SEP makerules, SEPScons, SEPlib, and the SEP latex package), or other free software such as SU. Before the publication of the electronic document, someone other than the author tests the author's claim by destroying and rebuilding all ER figures. Some ER figures may not be reproducible by outsiders because they depend on data sets that are too large to distribute, or data that we do not have permission to redistribute but are in the SEP data library, or that the rules depend on commercial packages such as

[1] http://sepwww.stanford.edu/private/docs/sep160
[2] http://sepwww.stanford.edu/public/docs/sepdatalib/toc_html

Matlab or Mathematica.

CR denotes Conditional Reproducibility. The author certifies that the commands are in place to reproduce the figure if certain resources are available. The primary reasons for the CR designation is that the processing requires 20 minutes or more.

NR denotes Non-Reproducible figures. SEP discourages authors from flagging their figures as NR except for figures that are used solely for motivation, comparison, or illustration of the theory, such as: artist drawings, scannings, or figures taken from SEP reports not by the authors or from non-SEP publications.

Our testing is currently limited to Linux 2.6 (using the Intel Fortran 9*/200* compiler), but the code should be portable to other architectures. Reader's suggestions are welcome. For more information on reproducing SEP's electronic documents, please visit *http://sepwww.stanford.edu/research/redoc/*.

I use special notation for important statements and optional material.

Important points or statements that I would like to emphasize appear in sections like this one.

Optional material that can be skipped at first reading is marked as shown here and typed in a script-size font.

Acknowledgments

"it warn't no time to be sentimentering."

Mark Twain

I would like to thank my adviser, professor **Biondo L. Biondi**, and my second project adviser, professor **Mark D. Zoback**, for granting me an opportunity to be part of Stanford Geophysics, and for their invaluable guidance and mentorship. Special thanks go to the founder of Stanford Exploration Project, professor **Jon F. Claerbout**, for starting it all and being a constant inspiration in my own work. I would like to express my deep gratitude to **Stewart A. Levin**, **Mark A. Meadows**, and **Paul Segall** for their advice and contribution to various aspects of my research.

My thanks go to **Shuki Ronen**, **Robert Clapp**, **Clement Kostov**, **Joseph Stefani**, **Dimitri Bevc**, **John Etgen**, **Matthew Karazincir**, **Phuong Vu**, **Min Zhou**, **Jean Paul Gestel**, **Joseph Dellinger**, **Jean Virieux**, **Jack Dvorkin**, **Gary Mavko**, **James Mika**, **Olav Barkved**, and **David Nichols** for useful discussions and suggestions. I thank **Chevron**, **ExxonMobil**, and **BHP Billiton** for providing Genesis field seismic data used in Part I of my thesis, and permission to publish my work; affiliate members of **Stanford Exploration Project** for their support; **CEEES HPC** and **Dennis Michael** for providing computing resources.

And finally, I would like to thank my **family** for their patience and support.

Contents

Abstract **v**

Preface **vii**

Acknowledgments **ix**

I Seismic time-lapse analysis **1**

1 Introduction and overview I **3**

A unified approach to inverse time-lapse analysis 5

Seismic time-lapse analysis 6

Time-lapse full-waveform inversion 9

Part I overview 10

2 Time-lapse scattering theory **11**

Acoustic scattering 13

Born series 15

Rytov series 17

Slowness inversion 20
Algorithm 24

3 Simultaneous 4D FWI **27**

Time-lapse full-waveform inversion 29
Joint time-lapse full-waveform inversion 30
Simultaneous inversion and cross-updating 33
Total-variation model-difference regularization 36

4 Sensitivity analysis **43**

Application to reflection data 45
Synthetic model 45
Time-lapse inversion from reflection data 47
Sensitivity to cycle-skipping 50

5 Genesis case study **57**

Genesis Field 59
2D inversion 62
Global and 3D inversion 67
Discussion and perspectives 73

II Geomechanical time-lapse analysis **79**

6 From deformation to pressure change **81**

Introduction 83

Poroelastic deformation 85

Forward and inverse problems 86

7 Monitoring by pressure inversion **93**

Cyclic steam stimulation 95

Inversion of pressure from tilt data 99

8 Reservoir characterization **107**

Recovering pressure contrasts 109

Comparison of two methods 109

III Appendices **115**

A FWI Objective Functional **116**

Objective Functions 118

Adjoint-state method 119

Wave-equation modeling 120

Method 121

Adjoint-state recipe 124

Examples 126

Frequency-domain full-waveform inversion 126

Time-domain full-waveform inversion 130

Phase-only full-waveform inversion 135

B Compressive Conjugate Directions **143**

Introduction 146
Steered and Compressive Conjugate Directions 153
Convergence Analysis 157
Limited-memory Compressive Conjugate Directions Method 166
Trade-off between the number of iterations and problem condition number 168
Applications 170
Image Denoising 170
Inversion of Dilatational Point Pseudo-sources 173
Inversion of Pressure Contrasts 180
Discussion 184
Generalizations 184

C TV with constraints **189**
Introduction 192
Method 193
Results 198
Conclusions and Perspectives 200

D Useful functions and equations **203**
Mindlin's Elastostatic Tensor 206

Bibliography **209**

Index **221**

List of Tables

List of Figures

3.1 True velocity difference consists of a negative (−200 m/s) perturbation at about 4.5 km inline 800 m depth, and a positive (300 m/s) perturbation at 6.5 km inline, 1 km depth. [**ER**] 34

3.2 Model difference inverted from a 14 dB SNR synthetic with matching baseline and monitor acquisition geometries using (a) iterated parallel difference; (b) iterated sequential difference; (c) cross-updating; (d) regularized simultaneous inversion; (e) iterated double difference. [**CR**] 37

3.3 Model difference inverted from a 7 dB SNR synthetic with matching baseline and monitor acquisition geometries using (a) iterated parallel difference; (b) iterated sequential difference; (c) cross-updating; (d) regularized simultaneous inversion; (e) iterated double difference. [**CR**] 38

3.4 Model difference inverted from a clean synthetic for different baseline and monitor acquisition geometries and sources. [**CR**] 39

3.5 Target area of the baseline model inverted from a clean synthetic. [**CR**] 39

3.6 Target area of the baseline model inverted from a 7 dB SNR synthetic. In both cases the baseline model is reconstructed reasonably well, however, errors due to noise are comparable in magnitude to production-induced effects. [**CR**] 40

3.7 True velocity differences consist of a negative (−200 m/s) perturbation at about 4.5 km inline 800 m depth and a positive (300 m/s) perturbation at 6.5 km inline, 1 km depth. [**CR**] 40

3.8 Model difference inverted using iterated sequential difference. [**CR**] . 41

3.9 Model difference recovered using cross-updating. [**CR**] 42

3.10 Model difference recovered using the simultaneous inversion with a TV-regularized model difference. Note the higher accuracy and stability to random noise of the TV-regularized simultaneous inversion. [**CR**] . 42

4.1 The true baseline model. A flat reflector model to study the sensitivity of FWI of short-offset reflection data to small velocity perturbation in the overburden. [**ER**] 46

4.2 The true model difference is a combination of a positive +300 m/s velocity change in a target reflector at a depth of 3900 m, and a negative velocity change in the overburden above the reflector, peaking at −50 m/s. In this chapter I investigate the sensitivity of simultaneous time-lapse FWI to small-magnitude blocky velocity changes in the overburden. [**ER**] 46

4.3 Time shifts observed in common-midpoint gathers centered above the target reservoir (blue is baseline, red is monitor). Travel times of the monitor near-offset reflections traveling through the negative velocity anomaly of Figure 4.2 are slightly delayed. [**CR**] 47

4.4 The parallel difference method fails to resolve the long-wavelength velocity changes of Figure 4.2, and produces negative short-wavelength artifacts around the target reflector. [**CR**] 48

4.5 The cross-updated FWI method cross-equalizes the baseline and monitor model but still fails to resolve the long-wavelength overburden changes of Figure 4.2. [**CR**] 48

4.6 Simultaneous FWI with a total-variation model-difference regularization resolves the long-wavelength overburden changes of Figure 4.2, but underestimates the maximum change, depending on the regularization strength. [**CR**] 49

4.7 Simultaneous FWI using Tikhonov model-difference regularization with the long-wavelength inversion of Figure 4.6 supplied as a prior. Note that such a multiscale approach can now resolve the short-wavelength positive-velocity changes of Figure 4.2. Strong Tikhonov regularization results in underestimated velocity changes within the reservoir but correctly locates the anomalies. [**CR**] 49

4.8 True model difference for demonstrating the inversion of multiple overburden anomalies. [**ER**] 50

4.9 Inversion of the two long-wavelength overburden anomalies of Figure 4.8 using simultaneous time-lapse FWI with a total-variation model-difference regularization. [**CR**] 51

4.10 Inversion of the two long-wavelength overburden anomalies of Figure 4.8 starting from a wrong initial model and using weak regularization (a small regularization parameter). FWI cycle skipped, and the baseline and monitor inversions diverged, contaminating the difference with cycle-skipping artifacts. [**CR**] 52

4.11 Inversion of the two long-wavelength overburden anomalies of Figure 4.8 starting from the same bad initial model but using a stronger regularization (a larger regularization parameter). FWI still cycle skipped, however, the strong model-difference regularization kept baseline and monitor within the characteristic wavelength of the overburden anomalies. The resulting model difference inversion is qualitatively accurate (compare with Figure 4.9), albeit stronger regularization has resulted in underestimated velocity magnitudes. [**CR**] 52

4.12 (a) True monitor image. (b) Monitor migrated using the baseline velocity model. Note that overestimated velocity in the overburden results in a downward reflector shift in the right image. [**CR**] 54

4.13 (a) Monitor image migrated using the sum of the baseline model and inverted model difference of Figure 4.6. (b) Monitor migrated using the baseline velocity model. Note that reflector shift in the overburden has been significantly reduced in the left image. [**CR**] 55

5.1 The Genesis field (from Magesan et al. (2005)). [**NR**] 59

5.2 A north-south inline section of the baseline Genesis image produced by Chevron (vertical axis two-way travel time in seconds, horizontal axis inline meters). [**NR**] 60

5.3 Monitor and baseline image-difference obtained from the 3D time-migration images provided by Chevron which corresponds to the inline section of Figure 5.2. Production-induced changes stand out at approximately 3.5 s (wet Illinoisan sands) and 4 s two-way travel times—stacked Neb 1, 2, and 3 reservoirs. [**NR**] 60

5.4 Production-induced changes resulted in measurable time-shifts between the surveys. Shown here are time-shifts between the baseline (blue) and monitor (red) common-offset gathers, 1074 m offset. [**CR**] 61

5.5 Inverted baseline velocity model (m/s). FWI resolved fine model features and oriented them along the dip structure of the image in Figure 5.2 (vertical axis depth meters). [**CR**] 62

5.6 (a) Parallel difference and joint inversion results for (b) $\delta = 100$, (c) $\delta = 500$, and (d) $\delta = 1000$ in the target area. The parallel difference result is not interpretable because of the presence of artifacts. Increasing the regularization parameter δ results in gradual removal of most model differences except the negative velocity change in the overburden, peaking around the Illinoisan sands and near the top of the stacked reservoirs—see Figures 5.7(a) through 5.7(d). [**CR**] . . . 65

5.7 (a) Baseline target area and (b) estimated model difference for $\delta = 1000$. Close-up of (c) baseline target area and (b) estimated model difference for $\delta = 1000$. [**CR**] 66

5.8 Inverted velocity difference interleaved with migrated image difference for baseline and monitor. The strongest image differences correspond to the top of Neb 1 reservoir and the Illinoisan sands above. Peaks of negative velocity change correlate with the overburden above the reservoir and the sands. [**CR**] 66

5.9 Trajectories of the Genesis production wells at the time of the monitor survey. Failed wells are shown in red, block 205 wells are plotted blue, block 160 wells are plotted magenta, and block 161 wells are plotted black. [**CR**] 67

5.10 Local (target-oriented) inversion described in the previous section plotted against a projection of well paths. Wells plotted in red are known to have failed. [**CR**] 68

5.11 Global velocity change inversion plotted against a projection of well paths. [**CR**] 69

5.12 Global velocity change inversion using an improved starting model, plotted against a projection of well paths. This result appears to be in a better agreement with the trajectories of the failed wells closest to the 2D inversion plane. [**CR**] 70

5.13 Velocity changes exceeding negative 50 m/s estimated by the 3D phase-only simultaneous FWI using only 60 shots. [**CR**] 71

5.14 Inversion of velocity changes in excess of -30 m/s in the overburden using (a) 4200-shot 2.5D experiment (b) 60-shot 3D experiment. Approximate depth of 3550 m. [**CR**] 72

5.15 Inversion of velocity changes in excess of -30 m/s in the overburden using (a) 4200-shot 2.5D experiment (b) 60-shot 3D experiment. Approximate depth of 3650 m. [**CR**] 74

5.16 Inversion of a velocity change in the overburden using (a) 4200-shot 2.5D experiment (b) 60-shot 3D experiment. Approximate depth of 3750 m. [**CR**] 75

5.17 Inversion of velocity changes in excess of -30 m/s in the overburden using (a) 4200-shot 2.5D experiment (b) 60-shot 3D experiment. Approximate depth of 3875 m. [**CR**] 76

7.1 Injection well trajectories. The portions of well trajectories that are within the reservoir are shown in red. [**CR**] 96

7.2 Inverted induced pore pressure change (color scale) and differential tilt measurements (arrows) for $\epsilon = 10^{-3}$ at the beginning of Phase 1 of Cycle 1, after 2 weeks of injection. The observed differential tilts are shown in green, modeled tilts are in red. [**CR**] 97

7.3 Inverted induced pore pressure change (color scale) and differential tilt measurements (arrows) for $\epsilon = 10^{-3}$ at the end of Phase 1 of Cycle 1, after 24 days of injection. [**CR**] 98

7.4 Inverted induced pore pressure change (color scale) and differential tilt measurements (arrows) for $\epsilon = 10^{-3}$ at the beginning of Phase 2 of Cycle 1, after 31 days of injection. [**CR**] 100

7.5 Inverted induced pore pressure change (color scale) and differential tilt measurements (arrows) for $\epsilon = 10^{-3}$ after 38 days of injection. [**CR**] . 101

7.6 Inverted cumulative pore pressure change and differential tilt measurements for $\epsilon = 10^{-3}$ during Phase 2 of Cycle 1. Day 46 of the injection. [**CR**] 102

7.7 Inverted cumulative pore pressure change and differential tilt measurements for $\epsilon = 10^{-3}$ during Phase 2 of Cycle 1. Day 60 of the injection. Note the increase of pressure from Figure 7.6 to Figure 7.7 in the northeast of the computational domain. This matches the geometry of injection wells within the reservoir in Figure 7.1. [**CR**] 104

7.8 Cumulative pore pressure change with $\epsilon = 10^{-3}$ for (a) 14 (b) 38 (c) 60 days into the injection cycle. [**CR**] 104

7.9 Cumulative pore pressure change with $\epsilon = 5 \times 10^{-3}$ for (a) 14 (b) 38 (c) 60 days into the injection cycle. [**CR**] 105

7.10 Cumulative pore pressure change with $\epsilon = 10^{-3}$ for (a) 14 (b) 38 (c) 60 days into the injection cycle. [**CR**] 105

8.1 Reservoir permeability. [**CR**] 110

8.2 Locations of water injection wells. [**CR**] 111

8.3 Noisy uplift measurements on (a) day 20, (b) day 40, (c) day 60. [**CR**] 111

8.4 (a) True induced pore pressure change (color scale) on day 20. (b) Estimated induced pore pressure change inverted using (6.7,6.8). [**CR**] 112

8.5 (a) True induced pore pressure change (color scale) on day 60. (b) Estimated induced pore pressure change inverted using (6.7,6.8). [**CR**] 113

8.6 (a) True induced pore pressure change (color scale) on day 60. (b) Estimated induced pore pressure change inverted using (6.7,6.8). [**CR**] 113

8.7 Estimated induced pore pressure change on day 40 using TV-regularized inversion (6.9,6.8) with bound constraints. Note that the permeability barrier of Figure 8.1 is well approximated by a level curve of the estimated pressure distribution. [**CR**] 114

A.1 Starting FWI model obtained by smoothing true model of Figure 4.1 using a 1920 m window. [**ER**] 139

A.2 A gradient computed using (A.62),(A.69) for $\omega = 3$ Hz, first iteration of conjugate gradients. Note that first arrival energy dominates the gradient, however, non-zero cross-correlations (A.62) are beginning to emerge at depth. [**CR**] 139

A.3 A gradient of (A.61) computed using (A.62),(A.69) for $\omega = 3$ Hz, tenth iteration of conjugate gradients. Note the long-wavelength periodic cross-correlations (A.62) at depth. [**CR**] 140

A.4 A gradient of (A.61) computed using (A.62),(A.69) for $\omega = 7.4$ Hz, first iteration of conjugate gradients. Note the short-wavelength periodic cross-correlations (A.62) at depth. [**CR**] 140

A.5 A gradient of (A.61) computed using (A.62),(A.69) for $\omega = 7.4$ Hz, tenth iteration of conjugate gradients. The model is sufficiently close to the true solution for the gradient to point to the solution. [**CR**] . . 141

A.6 Final inversion result for $\omega = 30$ Hz. Note that model updates for different frequencies constructively interfered to produce a quantitatively accurate approximation of the true model shown in Figure 4.1. [**CR**] 141

B.1 (a) Clean image [**NR**]; (b) Noisy image contaminated with Gaussian noise with $\sigma = 15\%$ of maximum amplitude [**CR**]; (c) Image denoised using Algorithm 4 with $\alpha = 10,\ \lambda = 1$ and memory size $m = 50$ [**CR**]; (d) Image denoised using Algorithm 5 with $\alpha = 10,\ \lambda = 1,\ N_c = 1$ [**CR**]. 172

B.2 Performance of Algorithm 4 with $m = 20$ versus Algorithm 5 with varying N_c for (a) $\lambda = 1$ [**CR**]; (b) $\lambda = 100$ [**CR**]; (c) $\lambda = 1000$ [**CR**]; (d) $\lambda = 10000$ [**CR**]. 174

B.3 (a) A spiky true pseudosources [**CR**]; (b) the resulting true (black) and noisy (red) surface displacements [**CR**]. 175

B.4 Inversion results for CCD (red), RCG (blue), FISTA (green) after 100 operator and adjoint applications for (a) $\lambda = .05$ [**CR**]; (b) $\lambda = 0.1$ [**CR**]; (c) $\lambda = 1$ [**CR**]; (d) $\lambda = 100$ [**CR**]. Note that FISTA does not use λ and the same FISTA results are shown in all plots but using different vertical scales. Improving condition number of (B.63) is accompanied by slower convergence. Compressive Conjugate Directions method most accurately resolves the spiky model at early iterations, and performs well when (B.63) is ill-conditioned. 177

B.5 Inversion results for CCD (red), RCG (blue), FISTA (green) after 1000 operator and adjoint applications for (a) $\lambda = .05$ [**CR**]; (b) $\lambda = 0.1$ [**CR**]; (c) $\lambda = 1$ [**CR**]; (d) $\lambda = 100$ [**CR**]. Note that FISTA does not use λ and the same FISTA results are shown in all plots but using different vertical scales. Compressive Conjugate Directions method still retains its advantage in resolving the spiky model at earlier iterations. *Asymptotically* faster convergence of FISTA kicks in when $\lambda = 100$ with a well-conditioned (B.63), when the ADMM convergence is slowed—compare with Figure B.7(d). 178

B.6 Convergence curves for CCD (solid red), RCG (dashed), FISTA (solid green) for (a) $\lambda = .05$ [**CR**]; (b) $\lambda = 0.1$ [**CR**]; (c) $\lambda = 1$ [**CR**]; (d) $\lambda = 100$—compare with Figures B.5(a),B.5(b),B.5(c),B.5(d) [**CR**]. . . 179

B.7 Convergence curves for CCD (solid red), ADMM with exact solver (blue), FISTA (green) for (a) $\lambda = .05$; (b) $\lambda = 0.1$; (c) $\lambda = 1$; (d) $\lambda = 100$. Limited-memory Compressive Conjugate Directions with $m = 100$ achieves convergence rate comparable to ADMM with exact minimization of (B.63). 181

B.8 (a) A blocky true pressure model (MPa) [**CR**]; (b) the resulting surface displacements (mm) with added random Gaussian noise with $\sigma = 15\%$ of data amplitude [**CR**]. 182

B.9 Inversion results after (a) 100 iterations (operator and adjoint applications) of CCD with $\lambda = 10$ [**CR**]; (b) 100 iterations of RCG with $\lambda = 10$ [**CR**]; (c) 1000 iterations of CCD with $\lambda = 10$ [**CR**]; (d) 1000 iterations of RCG with $\lambda = 10$ [**CR**]. In all tests, CCD is the limited-memory Compressive Conjugate Directions method of Algorithm 4; RCG is ADMM with restarted Conjugate Gradients of Algorithm 5 showing the most accurate model reconstruction among the outputs for different N_c–see Figures B.10(a),B.10(b),B.10(c),B.10(d). 183

B.10 Convergence rates for CCD and RCG with various N_c for (a) $\lambda = 5$ [**CR**]; (b) $\lambda = 10$ [**CR**]; (c) $\lambda = 50$ [**CR**]; (d) $\lambda = 100$ [**CR**]. 185

C.1 (a) True and noisy uplift observations. Random Gaussian noise with $\sigma = 15\%$ of maximum clean data amplitude was added to the clean data. (b) True model exhibits a highly compartmentalized "blocky" behavior. [**ER**] 199

C.2 (a) Unconstrained TV-regularized inversion. The algorithm tries to fit the data by allowing negative relative pressure changes. (b) Bound constrained TV-regularized result. Note that enforcing lower bounds resulted in a more accurate shape matching of the true model. [**ER**] . 201

C.3 (a) Direct imposition of the bound constraints at each iteration of the unconstrained solver resulted in a qualitatively and quantitatively wrong inversion. (b) Convergence of TV-regularized inverted models with bound constraints. The method quickly resolves both sharp contrasts and active bounds as only a few initial curves out of 1000 iterates lie outside a small neighborhood of the final curve. [**ER**] 201

Part I

Seismic time-lapse analysis

Chapter 1

Introduction and overview of part I

"Off we go!"

Yuri Gagarin

This first chapter provides a summary of research presented in Part I of the thesis related to seismic imaging applications. I begin with a conceptual description of time-lapse inverse theory and its relation with traditional seismic inversion methods. A brief overview of physical mechanisms of production-induced subsurface model change is followed by a summary of the existing time-lapse seismic techniques based on full-waveform inversion, and an overview of the subsequent Part I chapters.

A UNIFIED APPROACH TO INVERSE TIME-LAPSE ANALYSIS

This work is dedicated to developing a systematic theory for solving *time-lapse* inversion problems that I subsequently refer to as *time-lapse inverse theory*. The term "time-lapse" relates to separate observations of physical phenomena taken at discrete time intervals. Inverse time-lapse theory concerns itself with estimating or *inverting* changes in the underlying physical models from such discrete observations.

Problems of time-lapse or "4D" seismic imaging and reservoir geomechanics that arise in the Petroleum industry provide important applications for the developed theory, and are the primary target of my work as an applied geophysicist. However, many fundamental concepts, constructs and ideas that I have developed in the course of this study, as well some mathematical, algorithmic and computational byproducts of this research, are applicable beyond the limits of exploration seismology and reservoir geomechanics. I envisage ubiquitous applications of these concepts to diverse problems of acoustic and electromagnetic inverse scattering, imaging sciences and large-scale numerical optimization. While exploration geophysicists are the primary and key audience of this work, I have attempted to make my narrative accessible to specialists in applied mathematics and mathematical physics. Although a considerable part of this work is dedicated to providing a robust and systematic theoretical background for the proposed inversion techniques, I have structured the material in such a way that a motivated practitioner can go straight to the examples and case studies, as well as the nuts and bolts of specific algorithms as required for immediate practical application.

If we think of seismic time-lapse analysis as an estimation of *changes* in subsurface model parameters[1] that occurred between two separate seismic experiments, the "inverse time-lapse theory" can be simply regarded as a subset of the inverse acoustic scattering theory. Indeed, classical scattering theory addresses the problem

[1] as a result of petroleum production, fluid injection or environmental phenomena

of estimating properties of a "scatterer"—a perturbation in the background model—from incident and scattered wavefields. If production-induced subsurface changes are regarded as a scatterer, then the 4D analysis simply becomes a problem of inverse scattering theory. If so, why develop a new "time-lapse" inverse theory?

The answer is quite simple. In practical time-lapse applications of applied geophysics and beyond, the background model is not known accurately. Moreover, errors in our best estimates of the background model can be of the same magnitude as, or even exceed, the time-lapse effects that we seek to estimate. Is it even possible to estimate time-lapse changes when their magnitudes can be easily masked by the effects of measurement noise or errors in the background model? The main product of this research is a systematic theory of inverting small (and spatially bounded) time-lapses changes from noisy and limited observations. The proposed "time-lapse inverse theory" differs from direct inverse theory by placing emphasis on accurate estimation of *relative model changes* while ignoring errors in the background and perturbed models. I provide a toolkit of robust inversion techniques for accurate inversion of time-lapse changes, and demonstrate them on a series of synthetic and field-data examples.

☞ The proposed "time-lapse inverse theory" differs from direct inverse theory by placing emphasis on accurate estimation of *relative model changes* while ignoring errors in the background and perturbed models.

SEISMIC TIME-LAPSE ANALYSIS

At this point it is worthwhile to remind the reader of some key physical mechanisms that relate fluid production/injection or environmental changes to changes in the reservoir or overburden rock properties. First and the most obvious effect is the effect of *fluid substitution*, when water or other fluids replace produced hydrocarbons, or gas is released out of liquid hydrocarbons below the bubble point (Johnston, 2013).

Fluid substitution changes the density and bulk modulus of the saturated rock, but to a first order does not affect the shear modulus (Nur, 1982; Mavko et al., 2009). As a result, fluid substitution may significantly alter the acoustic, or pressure, wave propagation velocities within a hydrocarbon reservoir undergoing production or injection, but has a much smaller effect on the shear wave propagation velocity. Changes in velocity contrasts at the reservoir cap rock or trap affect reflectivity and translate into detectable changes in the amplitude of seismic signal and amplitude-versus-offset (AVO) signatures (Dvorkin et al., 2014). Velocity changes within the reservoir and in the overburden as a result of fluid migration[2], if sufficiently large in magnitude and spatial extent, yield detectable time shifts of seismic signals. If elastic properties of the rock frame, porosity and fluid saturations before and after productions are known, the effect of fluid substitution (a change in saturation) on the subsurface velocities can be accurately predicted[3] using the Gassman equation (Mavko et al., 2009; Dvorkin et al., 2014). Acoustic velocity and rock density increase as a result of water substituting for oil or gas, and decrease as a result of gas replacing oil or water. Shear wave velocities are relatively unaffected. Acoustic velocity changes resulting from fluid substitution may be significant and reach hundreds of meters per second, even for relatively small changes in fluid saturation, such as a mere 3% increase in gas saturation (Nur, 1982).

Changes in reservoir rock properties as a result of *stress changes* induced by production is another important mechanism producing time-lapse effects (Holt et al., 2005). It should be noted that velocity dependence on the confining stress and pore (fluid) pressure does not lend itself to easy analytical treatment and often exhibits hysteresis[4]. For compacting reservoirs and overburdens undergoing *dilation* above compacting reservoirs, it is more convenient to relate velocity and travel-time changes to strain—i.e., deformation. Unlike the effects of fluid substitution that are accurately quantifiable using the Gassman equation, the existing velocity-strain relations are empirical such as the R-factor method (Hatchell and Bourne, 2005; Herwanger, 2008). In

[2]e.g., fluids escaping from the reservoir
[3]to a first order
[4]i.e., dependence on the stress path

the *absence of counteracting effects*, acoustic and shear velocities, and rock density increase within a compacting reservoir, with an increasing effective stress and dropping pore pressure. Both acoustic and shear wave velocities typically decrease in a reservoir under water injection, with a decreasing effective stress and increasing pore pressure (Zoback, 2010). In a dilating, or stretching, overburden, rock density, acoustic and shear velocities decrease. Changes in acoustic velocity due to compaction or dilation are assumed proportional to strain and typically peak at 1-3% of the unperturbed velocity magnitude, reaching tens of meters per second.

Of course, time-lapse velocity changes are not limited to the effects of changing stress and fluids. Thermal and chemical processes play an important role in enhanced oil recovery (EOR) and result in significant induced changes of hydrocarbon fluid and reservoir rock properties (Nur et al., 1984; Johnston, 2013). However, regardless of what caused subsurface velocity changes, at the resolution limit of current seismic imaging methods, production-induced velocity changes can be classified as either short-wavelength "spiky" velocity model changes on a reservoir level, or long-wavelength "blocky" changes—typically within the overburden or underburden—that can be geometrically extensive but smaller in magnitude.

☞ At seismic resolution, subsurface velocity changes can be classified as either short-wavelength "spiky", occurring on a reservoir level, and long-wavelength "blocky", affecting large areas of the overburden. Automated image-difference techniques exist for recovering reservoir-level changes. The main objective of Part I of this thesis is to develop an automated method for recovering long-wavelength overburden effects, and propose a multi-scale extension of the method.

TIME-LAPSE FULL-WAVEFORM INVERSION

Detectability of changes in the subsurface properties induced by production was predicted in rock physics literature, and a time-lapse seismic method for the detection of fluid substitution effects in reservoir rocks was proposed by Nur (1982), leading to a systematic use of imaging attributes for predicting changes in the subsurface fluid parameters (Lumley, 1995). Prevalent practices of time-lapse seismic processing rely on picking time displacements and changes in reflectivity amplitudes between migrated baseline and monitor images, and converting them into impedance changes and subsurface deformation (Johnston, 2013). This approach requires a significant amount of manual interpretation and quality control. An alternative approach is based on using the high-resolution power of full-waveform inversion (Sirgue et al., 2010a) to reconstruct production-induced changes from wide-offset seismic acquisitions (Routh et al., 2012; Zheng et al., 2011; Asnaashari et al., 2012; Raknes et al., 2013; Maharramov and Biondi, 2014c; Yang et al., 2014). However, while potentially reducing the amount of manual interpretation, time-lapse FWI is sensitive to repeatability issues (Asnaashari et al., 2012), with both coherent and incoherent noise potentially masking important production-induced changes. A *linearized* joint inversion technique for time-lapse imaging that is more robust to repeatability issues was previously proposed by Ayeni (2011). In chapter 3 of this thesis I describe a joint time-lapse FWI (Maharramov and Biondi, 2013, 2014c) that addresses repeatability issues by jointly inverting multiple survey vintages with a model-difference regularization. This joint inversion approach is then extended to include edge-preserving total-variation (TV) model-difference regularization (Maharramov and Biondi, 2014e; Maharramov et al., 2015b). The new method is shown to achieve a dramatic improvement over alternative techniques by significantly reducing oscillatory artifacts in the recovered "blocky" model difference.

PART I OVERVIEW

Chapter 2 describes a time-lapse scattering theory that provides a theoretical foundation for the proposed joint regularized time-lapse FWI techniques operating on different scales. Chapter 3 describes the actual joint FWI methods and algorithms, and presents the results of synthetic experiments. Chapter 4 presents a sensitivity analysis of recovering small-magnitude long-wavelength velocity anomalies using a phase-only joint time-lapse FWI, and provides further synthetic experiments. In addition to recovering long-wavelength "blocky" changes in the overburden, I demonstrate a simultaneous recovery of production effects in both overburden and reservoir using a "cartoon + texture" model decomposition technique (Meyer, 2001; Maharramov and Biondi, 2015a). In chapter 5, the TV-regularized simultaneous inversion technique is applied to the Gulf of Mexico Genesis Field data, and is demonstrated to achieve a stable recovery of production-induced subsurface velocity changes (Maharramov et al., 2015a).

Chapter 2

Time-lapse scattering theory

"The beginning of wisdom is to call things by their right names."

Confucius

"Time-lapse inverse scattering theory" introduced in this chapter focuses on recovering changes in a physical model without accurate knowledge of the model background. More specifically, I study the feasibility of recovering low and high-wavenumber components of model perturbation using the traditional Born and Rytov scattering approximations, and establish a connection between the Rytov approximation and phase-only full-waveform inversion (FWI). I provide a theoretical justification for applying a regularized simultaneous time-lapse FWI to problems of applied seismology. I discuss the method's sensitivity to realistic production effects in seismic data, and its stability with respect to inaccurate starting models.

ACOUSTIC SCATTERING

In this chapter I define a "time-lapse inverse scattering theory" as a study of the effects that slowness perturbations have on acoustic wave propagation, and the inversion of a slowness perturbation from scattered wavefields without accurate knowledge of the background (unperturbed) slowness. I study different asymptotic representations of wavefields scattered by short and long-wavelength perturbations, and describe implications for the choice of a numerical method for inverting the perturbations. A theoretical sensitivity analysis of inversion accuracy to errors in the background model is followed by the description of a solution algorithm. This algorithm forms a basis for the numerical methods described in chapter 3.

Assuming a known background slowness $s(\mathbf{x}), \mathbf{x} \in \mathbb{R}^{31}$ and a slowness perturbation $\delta s(\mathbf{x})$, the total wavefield component $u(\mathbf{x})$ for frequency ω satisfies the Helmholtz equation

$$\left[\Delta + \omega^2 (s(\mathbf{x}) + \delta s(\mathbf{x}))^2\right] u(\mathbf{x}) \;=\; -f(\mathbf{x}),\; \mathbf{x} \in D \subset \mathbb{R}^3, \tag{2.1}$$

where $f(\mathbf{x})$ is the seismic source component for frequency ω. The total wavefield is the sum of the incident and scattered wavefields

$$u(\mathbf{x}) \;=\; u_I(\mathbf{x}) \;+\; u_S(\mathbf{x}), \tag{2.2}$$

where the incident wavefield u_I satisfies the Helmoholtz equation with the unperturbed slowness:

$$\left[\Delta + \omega^2 s^2(\mathbf{x})\right] u_I(\mathbf{x}) \;=\; -f(\mathbf{x}). \tag{2.3}$$

Note that for the well-posedness of (2.1) and (2.3) we need to impose an additional condition on the solution, such as the Sommerfeld radiation condition for a homogeneous medium (Colton and Kress, 1998). Physically, such a condition requires that the total field be outgoing at infinity. We will assume that equations (2.1) and (2.3) are solved in a domain $D \subset \mathbf{R}^3$, and absorbing boundary conditions (Engquist and

[1] acoustic slowness is the reciprocal of acoustic velocity

Majda, 1977) are applied along the domain boundary, ensuring an outgoing propagation of the wavefields.

For time-lapse problems we consider slowness perturbations $\delta s(\mathbf{x})$ with a support wholly contained in the interior of D. If $G(\mathbf{x},\mathbf{y})$ is Green's function for the unperturbed Helmholtz equation (2.3) in D and absorbing boundary conditions on ∂D, then equation (2.1) is equivalent to the Lippmann-Schwinger integral equation

$$u_S(\mathbf{x},\mathbf{y}) \;= \omega^2 \int_D G(\mathbf{x},\mathbf{y}) \left[(s(\mathbf{y})+\delta s(\mathbf{y}))^2 - s^2(\mathbf{y})\right] \left[u_I(\mathbf{y}) + u_S(\mathbf{y})\right] d\mathbf{y}, \qquad (2.4)$$

or, equivalently,

$$u_S(\mathbf{x},\mathbf{y}) \;= \omega^2 \int_{\operatorname{supp}\delta s} G(\mathbf{x},\mathbf{y}) \left[(s(\mathbf{y})+\delta s(\mathbf{y}))^2 - s^2(\mathbf{y})\right] \left[u_I(\mathbf{y}) + u_S(\mathbf{y})\right] d\mathbf{y}. \qquad (2.5)$$

The expression in the first square brackets is a perturbation of the "slowness squared" denoted from now on as δs^2. Note that

$$\delta s^2(\mathbf{x}) \;=\; (s(\mathbf{x})+\delta s(\mathbf{x}))^2 - s^2(\mathbf{x}) \;=\; 2s(\mathbf{x})\delta s(\mathbf{x}) + [\delta s(\mathbf{x})]^2 \;\approx\; 2s(\mathbf{x})\delta s(\mathbf{x}). \qquad (2.6)$$

The incident wavefield $u_I(\mathbf{x})$ in (2.4),(2.5) is assumed known.

Our modeling of the scattered wavefield does not rely on solving (2.5) but requires solving the Helmholtz equation in the entire modeling domain. However, here we briefly discuss potential advantages of solving the Lippmann-Schwinger equation instead of the Helmholtz equation. Three advantages of solving (2.5) for time-lapse problems instead of solving (2.1) with absorbing boundary conditions are immediately evident. First, it suffices to solve the Lippmann-Schwinger equation in the domain of support of δs that in practical applications is much smaller that D (e.g., compaction effects are limited to reservoirs and reservoir overburden and underburden). After a discretization, (2.5) becomes a system of linear equations with a dense modeling operator, and the dimension of the model space is determined by the size of supp δs over the computational grid. Second, a solution to (2.5) automatically satisfies absorbing

boundary conditions along ∂D because the unperturbed Green's function $G(\mathbf{x}, \mathbf{y})$ already satisfies those conditions. Third, once the scattered field is computed inside the support of δs, equation (2.5) can be used to compute its values outside of the perturbation—e.g., at surface receivers.

However, discretization of (2.5)[2] is a dense linear system, and its numerical properties are highly dependent on the spectral content (smoothness) of δs (Duan and Rokhlin, 2009). "Sparsifying" preconditioners for (2.5) are an area of active research (see Ying (2015) for homogeneous backgrounds) and merit an investigation as a potentially useful technique for forward modeling of scattered wavefields for spatially bounded perturbations. Another challenge of using (2.5) is that it explicitly contains a Green's function for problem (2.3). However, spatial boundedness of one of the arguments allows practical application of precomputed Green's functions (Etgen, 2012). In (2.5) both source and receiver arguments belong to the support of perturbation δs, making use of precomputed Green's functions feasible for compact targets. Computation of the scattered wavefield $u_S(\mathbf{x})$ outside of $\operatorname{supp} \delta s$ can be computationally equally efficient as wavefield measurements are made only at receiver locations.

BORN SERIES

Assuming that $\delta s^2 = O(\epsilon)$ where ϵ is the characteristic magnitude of a model perturbation, and formally representing the scattered wavefield as a *Born scattering series*

$$u_S(\mathbf{x}) \;=\; u_S^{(1)}(\mathbf{x}) \;+\; u_S^{(2)}(\mathbf{x}) \;+\; \ldots, \tag{2.7}$$

where

$$u_S^{(k)}(\mathbf{x}) \;=\; O(\epsilon^k),\; k = 1, 2, \ldots, \tag{2.8}$$

[2]using quadratures similar to Duan and Rokhlin (2009) for handling singularities at $\mathbf{x} = \mathbf{y}$

we obtain

$$\begin{aligned} u_S^{(1)}(\mathbf{x}) &= \omega^2 \int_{\text{supp}\,\delta s} G(\mathbf{x},\mathbf{y})\delta s^2(\mathbf{y}) u_I(\mathbf{y}) d\mathbf{y}, \\ u_S^{(i+1)}(\mathbf{x}) &= \omega^2 \int_{\text{supp}\,\delta s} G(\mathbf{x},\mathbf{y})\delta s^2(\mathbf{y}) u_S^{(i)}(\mathbf{y}) d\mathbf{y},\ i = 0, 1, \ldots \end{aligned} \tag{2.9}$$

Equations (2.9) will help us analyze the strengths of the Born series (2.7) in relating the diffracted wavefield $u_S(\mathbf{x})$ to $\delta s(\mathbf{x})$ for short-wavelength model perturbations, and its limitations for long-wavelength model perturbations. Assuming without a loss of generality, a homogeneous background $s(\mathbf{x}) = s_0$, and a constant finite δs, for a 3D medium we have

$$G(\mathbf{x},\mathbf{y}) = \frac{\exp(i\omega s_0|\mathbf{x}-\mathbf{y}|)}{4\pi|\mathbf{x}-\mathbf{y}|}. \tag{2.10}$$

An incident plane wave propagating along axis $x = x^1$ is given by

$$u_I(\mathbf{x}) = \exp(i\omega s_0 x). \tag{2.11}$$

For a perturbation in a *sufficiently small domain far from receiver locations*,

$$\sup\,|(\text{supp}\,\delta s)| \ll |x|\,, \tag{2.12}$$

the denominator of (2.10) is asymptotically a constant factor if $\mathbf{y} \in \text{supp}\,\delta s$, and from (2.9) and (2.6) we obtain

$$\begin{aligned} u_S^{(1)}(x,0,0) &\approx \frac{\delta s^2 \cdot \omega^2}{4\pi\,|x-x_0|} \int_{\text{supp}\delta s \cap \mathbb{R}^1} \exp(i\omega s_0(x-y^1)) \exp(i\omega s_0 y^1) dy^1 \\ &= \frac{L \cdot \delta s^2 \cdot \omega^2}{4\pi\,|x-x_0|} \exp(i\omega s_0 x) \approx \frac{L \cdot s_0 \cdot \delta s \cdot \omega^2}{2\pi\,|x-x_0|} \exp(i\omega s_0 x) \end{aligned} \tag{2.13}$$

where

$$L = |\text{supp}\delta s \cap \mathbb{R}^1(x)|,\ x_0 \in \text{supp}\delta s \cap \mathbb{R}^1(x), \tag{2.14}$$

is length of the model perturbation along axis $\mathbb{R}^1(x)$, and x_0 is an arbitrary point

within the support of δs. Note that (2.12) means that

$$|x_0| \ll |x|\,,\ L \ll |x|\,. \tag{2.15}$$

The second inequality in (2.15) can be understood as a "spikiness" of the model perturbation as it is concentrated in a relatively small subset, compared to the characteristic model dimensions, such as the distance from a source of the incident wave (i.e., a subsurface reflector) to a surface receiver.

Equation (2.13) means that the first-order Born scattering under our assumptions only accounts the amplitude effects but not the phase of the scattered wavefield. Indeed, phase changes (and time delays) accumulate in (2.7) through the effect of the denominator in (2.10), requiring progressively many terms to account for a phase delay or advance in the scattered (transmitted) wavefield. However, transmission through a constant perturbation δs of length L would cause a phase change proportional to $L\delta s$, therefore any technique based on the truncated Born scattering would be suboptimal for relating long-wavelength, or "blocky", velocity perturbations to measured scattered wavefields. This is a well-known limitation of diffraction tomography (Wu and Toksöz, 1987) that is inherited by any full-waveform inversion using the standard *time-domain* L_2 misfit function (Fichtner, 2011). On the other hand, the Born series is a very good scattering approximation for small-wavelength large-amplitude perturbations as, again, demonstrated by equation (2.13): such perturbations have a first-order effect on amplitude of the scattered wavefield.

RYTOV SERIES

The Rytov scattering series (Rytov et al., 1989; Ishimaru, 1978; Woodward, 1989) is based on an asymptotic phase expansion of the scattered wavefield and linearly relates phase changes with magnitudes of the slowness change to a first order, as does the initial approximation of full-waveform inversion of phase misfits (Fichtner, 2011). To demonstrate that, let us again study the propagation of a planar incident wave (2.11) through a perturbation δs. Under the Rytov scattering series we represent the

sum of the incident and scattered wavefields as

$$u_I(\mathbf{x}) \; + \; u_S(\mathbf{x}) \; = \; \exp\left[\phi^{(0)}(\mathbf{x}) \; + \; \phi^{(1)}(\mathbf{x}) \; + \; \phi^{(2)}(\mathbf{x}) + \; \ldots\right], \tag{2.16}$$

where

$$\phi^{(k)}(\mathbf{x}) \; = \; O\left(\epsilon^k\right), \; k = 0, 1, 2, \ldots, \tag{2.17}$$

and ϵ is again the characteristic magnitude of the perturbation. Note that in (2.16) phase terms $\phi^{(k)}(\mathbf{x})$ may have both nonzero real and imaginary parts, with the real parts determining solution amplitude. Substituting (2.16),(2.17) into (2.1) and assuming that

$$u_I(\mathbf{x}) = \exp\left[\phi^{(0)}(\mathbf{x})\right], \tag{2.18}$$

after grouping together the terms of the same order of magnitude we obtain:

$$\Delta\phi^{(0)}(\mathbf{x}) \; + \; |\nabla\phi^{(0)}(\mathbf{x})|^2 \; + \; \omega^2 s^2(\mathbf{x}) \; = \; -f(\mathbf{x}), \tag{2.19}$$

and

$$\Delta\phi^{(1)}(\mathbf{x}) \; + \; 2\langle\nabla\phi^{(0)}(\mathbf{x}), \phi^{(1)}(\mathbf{x})\rangle \; + \; |\nabla\phi^{(1)}(\mathbf{x})|^2 + \omega^2\delta s^2(\mathbf{x}) \; = \; 0, \tag{2.20}$$

where only the first two terms of the Rytov series (2.16) are used. Assuming, for simplicity but without loss of generality, a constant background s_0 and a constant finite perturbation δs, the phase change for a transmitted plane wave traveling through a perturbation δs of characteristic dimension L is approximately proportional to $L\delta s^2/(2s_0) \approx L\delta s$. Indeed, the plane-wave solution (2.11) satisfies (2.19) outside of supp f. Assuming a heterogeneity along the axis x only, hence

$$\phi^{(1)}(\mathbf{x}) \; = \; \phi^{(1)}(x),$$

we obtain

$$\frac{d^2}{dx^2}\phi^{(1)}(x) \; + \; \left[\frac{d}{dx}\phi^{(1)}(x)\right]^2 \; + \; 2i\omega s_0\frac{d}{dx}\phi^{(1)}(x) \; + \; \omega^2\delta s^2(x) \; = \; 0. \tag{2.21}$$

If slowness perturbations are small in magnitude

$$|\delta s| \ll 1 \Rightarrow \left|\frac{\delta s^2}{s_0}\right| \ll 1, \tag{2.22}$$

we can drop the second term in (2.21) as it is of order $O(\epsilon^2)$ to obtain

$$\frac{d^2}{dx^2}\phi^{(1)}(x) + 2i\omega s_0 \frac{d}{dx}\phi^{(1)}(x) + \omega^2 \delta s^2(x) = 0. \tag{2.23}$$

Furthermore, assuming that the perturbed phase is *slowly changing in space*, we can drop the first term to obtain the following equation that governs a strictly imaginary phase[3]

$$\begin{aligned} \phi^{(1)}(x) &\approx i\omega \int_{-\infty}^{x} \frac{\delta s^2(\eta)}{2s_0} d\eta = i\omega \int_{\operatorname{supp}\delta s} \frac{\delta s^2(\eta)}{2s_0} d\eta \\ &\approx i\omega \int_{\operatorname{supp}\delta s} \frac{2s_0 \delta s(\eta)}{2s_0} d\eta = i\omega \int_{\operatorname{supp}\delta s} \delta s(\eta) d\eta. \end{aligned} \tag{2.24}$$

Note that the discarded term in (2.23) contributes to the real phase component and thus accounts for amplitude effects. Let us now analyze the order of magnitude of amplitude effects that can be estimated by solving the full equation (2.23). Solving the linear equation with constant coefficients, we obtain

$$\frac{d}{dx}\phi^{(1)}(x) = -\omega^2 \int_{-\infty}^{x} e^{-2i\omega s_0(x-\eta)} \delta s^2(\eta) d\eta, \tag{2.25}$$

and after integration by parts,

$$\begin{aligned} \frac{d}{dx}\phi^{(1)}(x) &= -\frac{\omega^2}{2i\omega s_0} \int_{-\infty}^{x} \delta s^2(\eta) de^{-2i\omega s_0(x-\eta)} \\ &= \frac{i\omega}{2s_0}\delta s^2(x) - \frac{i\omega}{2s_0} \int_{-\infty}^{x} e^{-2i\omega s_0(x-\eta)} \left[\delta s^2(\eta)\right]' d\eta \\ &= \frac{i\omega}{2s_0}\delta s^2(x) - \frac{1}{4s_0^2}\left[\delta s^2(x)\right]' \end{aligned}$$

[3]thus producing a phase-only Rytov approximation

$$
\begin{aligned}
&+ \frac{1}{4s_0^2}\int_{-\infty}^{x} e^{-2i\omega s_0(x-\eta)}\left[\delta s^2(\eta)\right]'' d\eta \\
&= \frac{i\omega}{2s_0}\delta s^2(x) - \sum_{k=1}^{+\infty}\frac{i\omega}{2s_0\left(2i\omega s_0\right)^k}\left[\delta s^2(x)\right]^{(k)},
\end{aligned}
\tag{2.26}
$$

where we assume that the derivatives of squared slowness perturbation ultimately become negligible, $[\delta s^2]^{(k)} \sim 0$. The first term in the right-hand side of (2.26) yields the first-order phase-shift of (2.24) after integration. The odd terms of the series in (2.26) are phases with nonzero real parts and contribute to the amplitude of the scattered wave. Note that significant oscillations of model perturbation δs^2 may result in large spikes of the slowness derivative and thus affect the scattered amplitude. On the other hand, oscillatory model perturbations with zero average have negligible effect on the phase and, by extension, travel time of the scattered wave—see (2.24).

SLOWNESS INVERSION

We can now see that the *phase-only* Rytov approximation is applicable when the slowness perturbations are sufficiently small (2.22) and slowly varying in space, but possibly spread over a finite domain. The Born approximation, on the other hand, works best for perturbations within small domains, or scattered diffractors (compare with Slaney et al. (1984)). Indeed, the Rytov approximation with a strictly imaginary phase is equivalent to transmission travel-time analysis: the integral in (2.24) simply yields the travel-time delay as a result of a slowness perturbation δs.

It should be noted that while the requirement of perturbation "spikiness" (2.15) is critical for the accuracy of the Born approximation, the requirement of "blockiness" and small variation of the slowness perturbation is not required for the accuracy of the Rytov approximation with a full complex phase. Indeed, solving the second-order ordinary differential equation (2.23) in (2.26) accounts for both amplitude and phase effects of perturbation oscillations. However, in the absence of significant perturbation oscillations there is a simple linear relationship between the slowness perturbation and phase change (time-shift) (2.24) that matches the linearized travel-time tomography

approximation.

For blocky model changes affecting primarily travel-times, such as those due to the effects of overburden dilation, a method that inverts a slowness perturbations from the observed phase change (time-shift) may adequately resolve the perturbation without using amplitude information. For inverting such blocky model perturbations, in the absence of reliable amplitude information in the observed data, I propose to use a phase-only full-waveform inversion:

$$\|\exp i\arg \mathbf{d}_m \;-\; \exp i\arg \mathbf{u}(\mathbf{m}_m)\|_2^2 \;\rightarrow\; \min, \tag{2.27}$$

where arg in (2.27) stands for the complex argument; $\mathbf{d}_m$ and $\mathbf{u}$ are single-frequency components of the observed and predicted datasets, $\mathbf{m}_m$ is the "monitor"[4] slowness model. We seek a slowness perturbation

$$\delta s \;=\; \mathbf{m}_m \;-\; \mathbf{m}_b, \tag{2.28}$$

that minimizes the quadratic misfit of the observed and predicted wavefield phases (2.27). In (2.28), $\mathbf{m}_b$ is a known "baseline"[5] model (see chapter 3).

The first iteration of the frequency-domain FWI with a phase-only misfit function is equivalent to the linear phase-only Rytov inverse scattering approximation of model updates. Subsequent iterations take into account the nonlinearity of the phase as a function of slowness. The optimization problem (2.27) is solved in the frequency domain.

If amplitude information in the observed wavefield measurements is accurate, and our wave propagation is dynamically accurate[6], then we can use "phase-plus-amplitude" frequency-domain full-waveform inversion,

$$\|\mathbf{d}_m \;-\; \mathbf{u}(\mathbf{m}_m)\|_2^2 \;\rightarrow\; \min. \tag{2.29}$$

[4]post-production, or after subsurface changes occurred
[5]pre-production, or before subsurface changes occurred
[6]i.e., produces accurate amplitudes

A slowness perturbation (2.28) is nonzero only in the areas affected by production. Other *a priori* information may be available about δs such as its "spikiness" (changes within the reservoir) and "blockiness" (stretching of the overburden). We can combine the objective functions (2.27) and (2.29) with a regularization or penalty term that promotes desirable features and penalizes undesirable ones:

$$\|\exp i\arg \mathbf{d}_m \;-\; \exp i\arg \mathbf{u}(\mathbf{m}_m)\|_2^2 \;+\; \mathbf{P}(\mathbf{m}_m \;-\; \mathbf{m}_b) \;\to\; \min, \tag{2.30}$$

and

$$\|\mathbf{d}_m \;-\; \mathbf{u}(\mathbf{m}_m)\|_2^2 \;+\; \mathbf{P}(\mathbf{m}_m \;-\; \mathbf{m}_b) \;\to\; \min\,. \tag{2.31}$$

where $\mathbf{P}$ is some penalty function of δs (see chapter 3).

The phase-only inversion (2.27) or (2.30) has the advantage of not using the potentially unreliable dynamic wavefield information. Another significant advantage of the phase-only inversion is that in *time-lapse problems* of inverting long-wavelength small-magnitude model perturbations, Rytov inverse scattering and phase-only FWI are less sensitive to errors in the background model. For significant phase changes, phase wraps around the wavelength when exceeding it. Fitting peaks and trough of the modeled and observed scattered wavefields (ignoring the amplitude information) then results in an ambiguity of the total phase change: the phase change can be resolved only within an integer multiple of the incident wavelength. This results in the well-known phenomenon of cycle skipping in FWI: unless the FWI starting slowness model is accurate within the wavelength of the incident wave, the model cannot be resolved from signal phase information alone.

However, for time-lapse problems the phase change due to a compact velocity anomaly of a *sufficiently small magnitude* is only a fraction of the wavelength. Indeed, after translating to the time domain, time shifts due to dilation in overburden peak at about 10 ms (Rickett et al., 2006; Maharramov and Biondi, 2015b; Maharramov et al., 2015a)—i.e., about a third of the period for a 30 Hz signal (see chapter 5). Therefore, phase changes (equivalently, time delays) of scattered wavefields for small-magnitude long-wavelength perturbations that are of interest for us can still be

translated into slowness changes using (2.24) if there is sufficient illumination of the anomaly. Moreover, those slowness changes are accurate to a first order *regardless* of errors in the background slowness. However, errors in the background *velocity* model will result in errors in the estimated velocity perturbation. For a slowness s and the corresponding velocity model v, we have:

$$\delta s \;=\; \delta\left(\frac{1}{v}\right) \;\approx\; -\frac{\delta v}{v^2}. \tag{2.32}$$

If v_R and v_W and the right and wrong velocity models and δs is a slowness perturbation estimated using phase fitting, then

$$\delta v_W \;=\; -\delta s v_W^2 \;=\; -\delta s v_R^2\left(\frac{v_W}{v_R}\right)^2 \;=\; \delta v_R\left(\frac{v_W}{v_R}\right)^2, \tag{2.33}$$

where δv_R and δv_W are velocity perturbations estimated for the correct and wrong background velocity models, respectively. Equation (2.33) means that even with a wrong background velocity a qualitative information on the magnitude of *velocity* perturbation can still be extracted from the scattered wavefield. Note that location of the perturbation is determined by the illumination pattern of incident wavefields. Poor target illumination results in the ambiguity of the characteristic dimension L of the anomaly versus the perturbation magnitude δs as the two enter into (2.24) as a product. For example, lack of reflectors above the velocity anomaly results in an ambiguous vertical extent of the anomaly.

Note that our reliance on (2.24) requires transmission of the incident wave through the anomaly. For reflection seismic data, this implies that the baseline model in (2.27) contains a model contrast at an *approximately* correct position of actual reflectors below and above the perturbation zone. In the presence of diving-wave[7] signal this requirement is not necessary.

[7] refracted

☞ Phase-only full-waveform inversion (2.27) may produce a *qualitatively* accurate estimate of blocky, non-oscillatory production-induced model changes even when starting from a wrong baseline model. However, if using reflection seismic data, the baseline model should have slowness contrasts to generate synthetic reflections around the anomaly.

ALGORITHM

The regularized FWI (2.30) and (2.31) requires that the baseline and monitor models be close in some norm. Both models may contain errors and artifacts due to insufficient or noisy data, numerical errors, or insufficient solver iterations. While imposing a penalty $\mathbf{P}(\delta s)$ promotes similarity of the monitor model to the baseline model, and penalizes unwanted differences, cross-equalization of the two models can be more efficient if *both* models are allowed to change.

☞ Simultaneously inverting both baseline and monitor models with a model-difference regularization promotes cross-equalization of the two models, with a better constrained model filling in the gaps in the other model. A model-difference regularization is chosen in such a way as to promote desirable features in the model difference while penalizing the undesirable ones.

To implement a practical time-lapse inversion method using phase-only FWI, we can invert two models (unperturbed baseline and perturbed monitor) simultaneously, imposing a model-difference regularization. The latter is required to create a common "background" model for both inversions making the application of (2.24) possible.

Note that baseline and monitor inversions may still cycle-skip, but another advantage of imposing a model-difference regularization is that the two models can be forced to be either equally accurate or equally inaccurate for wavelengths greater than the characteristic wavelength of the inverted perturbation.

For blocky, long-wavelength anomalies I suggest imposing blockiness-promoting total variation regularization (Maharramov and Biondi, 2014e), while for the recovery of short-wavelength features I use an L_2 Tikhonov model-difference regularization (Maharramov and Biondi, 2013), as described in more detail in the next chapter.

Chapter 3

Simultaneous time-lapse full-waveform inversion

"Each problem that I solved became a rule which served afterwards to solve other problems."

Rene Descartes

In this chapter I describe a technique for reconstructing subsurface velocity model changes from time-lapse seismic data using full-waveform inversion (FWI) that implements the algorithm outlined in chapter 2. The technique is based on simultaneously inverting multiple survey vintages, with a model difference regularization using a variety of norms, including the total variation (TV) seminorm. I compare this new joint inversion method with existing techniques, and demonstrate successful recovery of production-induced model changes from noisy synthetic data.

TIME-LAPSE FULL-WAVEFORM INVERSION

Full-waveform inversion may be defined as solving the following optimization problem (Tarantola, 1984; Virieux and Operto, 2009)

$$\|\mathbf{d} - \mathbf{u}(\mathbf{m})\|_2^2 \ \rightarrow \ \min, \tag{3.1}$$

where $\mathbf{d}$ denotes observation data, $\mathbf{u}$ is the solution of a forward-modeling problem

$$\mathbf{D}(\mathbf{m})\mathbf{u} \ = \ \phi, \tag{3.2}$$

where $\mathbf{D}$ is the forward-modeling operator that depends on a model vector $\mathbf{m}$ as a parameter, and ϕ is a source. The minimization problem (3.1) is solved with respect to either both the model $\mathbf{m}$ and source ϕ or just the model. In the frequency-domain formulation of the acoustic waveform inversion, the forward-modeling equation (3.2) becomes

$$-\omega^2 u - v^2(x^1, \ldots, x^n)\Delta u \ = \ \phi(\omega, x^1, \ldots, x^n), \tag{3.3}$$

where ω is a temporal frequency, n is the problem dimension, and v is the acoustic wave propagation velocity. Values of the slowness $s = 1/v$ at all the points of the modeling domain constitute the model parameter vector $\mathbf{m}$. The direct problem (3.3) can be solved in either frequency or time domain (Virieux and Operto, 2009). The inverse problem (3.1) is typically solved using a multiscale approach, working from low to high frequencies, supplying the output of each frequency inversion to the next step (Fichtner, 2011).

FWI applications in time-lapse problems seek to recover induced changes in the subsurface model using multiple datasets from different acquisition vintages. For two surveys sufficiently separated in time, I call such datasets (and the associated models) *baseline* and *monitor*.

Time-lapse FWI can be conducted by separately inverting the baseline and monitor models (*parallel difference*), or by inverting them sequentially with, e.g., the

baseline supplied as a starting model for the monitor inversion (*sequential difference*). Another alternative is to apply the *double-difference* method, with a baseline model inversion followed by a monitor inversion that solves the following optimization problem,

$$\| \left(\mathbf{u}_m - \mathbf{u}_b\right) - \left(\mathbf{d}_m - \mathbf{d}_b\right) \|_2^2 \ \rightarrow \ \min, \tag{3.4}$$

by changing the monitor model (Watanabe et al., 2004; Denli and Huang, 2009; Zheng et al., 2011; Asnaashari et al., 2012; Raknes et al., 2013). The subscripts in equation (3.4) denote the baseline and monitor surveys, and $\mathbf{d}$ denotes the observed data. For a phase-only inversion, in all of the subsequent equations the modeled and observed data differences should be replaced with the corresponding "phase differences"

$$\mathbf{u} \ - \ \mathbf{d} \implies \exp i\arg \mathbf{u} \ - \ \exp i\arg \mathbf{d}, \tag{3.5}$$

where arg is the complex argument function of frequency domain wavefields. Note that unlike the traditional phase-only inversion (Fichtner, 2011), we evaluate exponent of the phase to avoid phase discontinuity[1].

JOINT TIME-LAPSE FULL-WAVEFORM INVERSION

In all of these techniques, optimization is conducted with respect to one model at a time, albeit of different vintages at different stages of the inversion. Implementing the algorithm of chapter 2, I propose to invert for the baseline and monitor models *simultaneously* by solving either one of the following two optimization problems:

$$\alpha\|\mathbf{u}_b - \mathbf{d}_b\|_2^2 + \beta\|\mathbf{u}_m - \mathbf{d}_m\|_2^2 + \tag{3.6}$$

$$\gamma\| \left(\mathbf{u}_m - \mathbf{u}_b\right) - \left(\mathbf{d}_m - \mathbf{d}_b\right) \|_2^2 + \tag{3.7}$$

$$\alpha_1\|\mathbf{W}_b\mathbf{R}_b(\mathbf{m}_b - \mathbf{m}_b^{\text{PRIOR}})\|_2^2 + \tag{3.8}$$

$$\beta_1\|\mathbf{W}_m\mathbf{R}_m(\mathbf{m}_m - \mathbf{m}_m^{\text{PRIOR}})\|_2^2 + \tag{3.9}$$

$$\delta\|\mathbf{WR}(\mathbf{m}_m - \mathbf{m}_b - \Delta\mathbf{m}^{\text{PRIOR}})\|_2^2 \ \rightarrow \ \min, \tag{3.10}$$

[1] effectively normalizing the wavefield by its amplitude

or

$$\alpha\|\mathbf{u}_b - \mathbf{d}_b\|_2^2 + \beta\|\mathbf{u}_m - \mathbf{d}_m\|_2^2 + \tag{3.11}$$

$$\gamma\|\left(\mathbf{u}_m - \mathbf{u}_b\right) - \left(\mathbf{d}_m - \mathbf{d}_b\right)\|_2^2 + \tag{3.12}$$

$$\alpha_1\|\mathbf{W}_b\mathbf{R}_b(\mathbf{m}_b - \mathbf{m}_b^{\mathrm{PRIOR}})\|_1 + \tag{3.13}$$

$$\beta_1\|\mathbf{W}_m\mathbf{R}_m(\mathbf{m}_m - \mathbf{m}_m^{\mathrm{PRIOR}})\|_1 + \tag{3.14}$$

$$\delta\|\mathbf{W}\mathbf{R}(\mathbf{m}_m - \mathbf{m}_b - \Delta\mathbf{m}^{\mathrm{PRIOR}})\|_1 \;\rightarrow\; \min, \tag{3.15}$$

with respect to both the baseline and monitor models $\mathbf{m}_b$ and $\mathbf{m}_m$. Problem (3.6-3.10) describes time-lapse FWI with an L_2 regularization of the individual models (3.8,3.9) and model difference (3.10) (Maharramov and Biondi, 2014d). The second formulation (3.11–3.15) involves an L_1-regularization of the individual models and their difference (Maharramov and Biondi, 2014e; Maharramov et al., 2015b). The terms (3.11) correspond to separate baseline and monitor inversions, the term (3.12) is the optional double difference term, the terms (3.13) and (3.14) are optional separate baseline and monitor inversion regularization terms (Aster et al., 2011), and the term (3.15) represents regularization of the model difference. In (3.13–3.15), $\mathbf{R}$ and $\mathbf{W}$ denote regularization and weighting operators respectively, with the subscript denoting the survey vintage where applicable. If $\mathbf{R}$ is the *gradient magnitude* operator

$$\mathbf{R}f(x,y,z) \;=\; \sqrt{f_x^2 + f_y^2 + f_z^2}, \tag{3.16}$$

then (3.13–3.15) become *total-variation (TV) seminorms*. The latter case is of particular interest in this work as the minimization of the L_1 norm of gradient may promote "blockiness" of the model-difference, potentially reducing oscillatory artifacts (Rudin et al., 1992; Aster et al., 2011).

A joint inversion approach has been applied earlier to the linearized waveform inversion (Ayeni and Biondi, 2012). In Maharramov and Biondi (2013, 2014d,b), a simultaneous full-waveform inversion problem (3.6,3.10) was studied with a single model difference L_2 regularization term (3.10).

An implementation of the proposed simultaneous inversion algorithm requires solving a nonlinear optimization problem with twice the data and model dimensions of problems (3.1) and (3.4). The model difference regularization weights $\mathbf{W}$ and, optionally, the prior $\Delta\mathbf{m}^{\text{PRIOR}}$ may be obtained from prior geomechanical information. For example, a rough estimate of production-induced velocity changes can be obtained from time shifts (Hatchell and Bourne, 2005; Barkved and Kristiansen, 2005) and used to map subsurface regions of expected production-induced perturbation and optionally provide a difference prior. However, successfully solving the L_1-regularized problem (3.11-3.12) is less sensitive to choice of the weighting operator $\mathbf{W}$. For example, I show below that the TV-regularization using (3.16) with $\mathbf{W} = 1$ recovers non-oscillatory components of the model difference, while the L_2 approach would result in either smoothing or uniform reduction of the model difference.

In addition to the fully simultaneous inversion, Maharramov and Biondi (2013, 2014d) proposed and tested a *cross-updating* technique that offers a simple but remarkably effective approximation to minimizing the objective function that involves only the terms (3.6) and (3.10), while obviating the difference regularization and weighting operators $\mathbf{R}$ and $\mathbf{W}$. This technique consists of one standard run of the sequential difference algorithm, followed by a second run with the inverted monitor model supplied as the starting model for the second baseline inversion,

$$\begin{aligned}\mathbf{m}_{\text{INIT}} \rightarrow &\text{baseline inversion} \rightarrow \text{monitor inversion} \rightarrow \\ &\text{baseline inversion} \rightarrow \text{monitor inversion},\end{aligned} \tag{3.17}$$

and computing the difference of the latest inverted monitor and baseline models. Process (3.17) can be considered as an approximation to minimizing (3.6) and (3.10) because non-repeatable footprints of both inversions are propagated to both models, canceling out in the difference. Both the simultaneous inversion and cross-updating minimize the model difference by tackling model artifacts that are in the null space of the Fréchet derivative of the forward modeling operators. The joint inversion minimizes the effect of such artifacts on the model difference by either minimizing the model difference term (3.10) in the simultaneous inversion, or by propagating

these artifacts to both models in cross-updating (3.17). Note that this process is not guaranteed to improve the results of the baseline and monitor model inversions but is only proposed for improving the model difference. Maharramov and Biondi (2014d,b) demonstrated a significant improvement of model difference recovery by both the L_2-regularized target-oriented simultaneous inversion and cross-updating compared to the parallel, sequential and double difference techniques. The simultaneous inversion and cross-updating yielded qualitatively similar results within the inversion target. The simultaneous inversion allows a hybrid approach with a non-zero double-difference term—e.g., (3.7) that was studied by Maharramov et al. (2015c) but is not further discussed in this thesis.

SIMULTANEOUS INVERSION AND CROSS-UPDATING

In this section I will demonstrate the simultaneous inversion (3.6,3.10) and cross-updating (3.17) methods, and compare them with earlier time-lapse FWI techniques. Marmousi velocity model, specified on a 384×122 grid with a 24 m grid spacing is used as a baseline. Production-induced velocity changes are modeled as a negative −250 m/s perturbation at about 4.5 km inline 800 m depth, and a positive 300 m/s perturbation at 6.5 km inline, 1 km depth shown in Figure 3.1. While I invert the whole Marmousi model, only model differences for the section between the approximate inline coordinates 3.6 km and 7.2 km to the depth of approximately 1.6 km are shown here. The inversion is carried out in the frequency domain for

$$3.0,\ 3.6,\ 4.3,\ 5.1,\ 6.2,\ 7.5,\ 9.0,\ 10.8,\ 12.8,\ \text{and } 15.5 \text{ Hz},$$

where the frequencies are chosen based on the estimated offset to depth range of the data (Sirgue and Pratt, 2004). The baseline acquisition has 192 shots at a depth of 16 m with a 48 m spacing, and 381 receivers at a depth of 15 m with a 24 m spacing. The minimum offset is 48 m. The source function is a Ricker wavelet centered at 10.1 Hz. Absorbing boundary conditions are applied along the entire model boundary, including the surface (thus suppressing multiples). A smoothed true model is used as a starting model for the initial baseline inversion (and for the initial monitor inversion

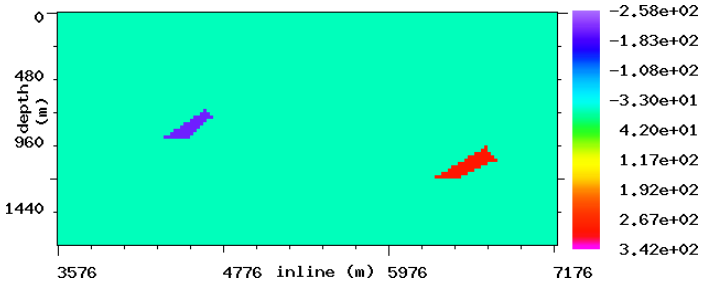


Figure 3.1: True velocity difference consists of a negative (−200 m/s) perturbation at about 4.5 km inline 800 m depth, and a positive (300 m/s) perturbation at 6.5 km inline, 1 km depth. [**ER**] jfwi/. truediff

in the parallel difference). The smoothing is performed using a triangular filter with a 20-sample half-window in both vertical and horizontal directions. Random Gaussian noise is added to the noise-free synthetic data to produce noisy data sets with 14 dB and 7 dB signal-to-noise ratios. Noisy monitor data sets are generated for the model perturbation of Figure 3.1, using the same acquisition geometry and source wavelet. Results of model difference inversion from the 14 dB and 7 dB synthetic data sets using various methods are shown in Figures 3.2 and 3.3, respectively. The simultaneous inversion objective function contains only terms (3.6) and (3.10) with no difference prior, i.e., $\Delta\mathbf{m}^{\text{PRIOR}} = 0$. The model-difference regularization weights $\mathbf{W}$ in (3.10) are set to 1 outside approximately .5 km from the centers of the velocity anomalies (Figure 3.1), tapering to zero within a smaller radius of the anomalies. The two terms in (3.6) are of the same magnitude and therefore α and β are set to 1. Parameter δ is chosen constant at 10^{-5} but can be different for different acquisition source and geometry parameters. The result of the initial baseline inversion is supplied as a starting model for both $\mathbf{m}_b$ and $\mathbf{m}_m$ in the simultaneous inversion. In all the inversions, up to 10 iterations of the nonlinear conjugate gradients algorithm (Nocedal and Wright, 2006) are performed for each frequency. Neither regularization nor model priors are used in single-model inversions (i.e., in the cross-updating, parallel, sequential, and double difference methods).

The results of applying cross-updating to the two noisy data sets are shown on Figures 3.2c and 3.3c, respectively. The corresponding simultaneous inversion results

are shown on Figures 3.2d and 3.3d. Since the problem (3.1) is nonlinear, supplying the result of the highest frequency inversion back to the lowest frequency and repeating the whole inversion cycle for all frequencies may result in achieving a better data misfit. In repeated cycles, lower-frequency inversions usually terminate earlier but higher frequencies still deliver model updates. For an objective comparison of the joint inversion with the parallel, sequential and double-difference methods, the effects of insufficient iteration count are reduced by performing an extra cycle of baseline and monitor inversion in each of the latter methods (we can call this approach "iterated" parallel, sequential and double difference). The results of applying the iterated parallel difference to the two noisy data sets are shown on Figures 3.2a and 3.3a. The results for the iterated sequential difference are shown on Figures 3.2b and 3.3b. The double-difference inversions are shown on Figures 3.2e and 3.3e. The double difference method yields the worst results for noisy data, and this is consistent with earlier tests of the method on noisy data (Asnaashari et al., 2012). The sequential difference delivers consistent improvement over the parallel difference, while the cross-updating delivers a significant improvement over the sequential method. The simultaneous inversion and cross-updating appear to yield similar results where $\mathbf{W} \approx 0$. The results of inverting the model difference from the noise-free synthetic are not shown as all the methods perform well in this case. Joint inversion, either by cross-updating or simultaneous inversion, demonstrates robustness with regard to uncorrelated noise in the data.

Cross-updating keeps the baseline and monitor data spaces separate, and the method is robust with respect to changes in acquisition geometry and source parameters between surveys. Figure 3.4 demonstrates cross-updating with different surveys. The monitor survey in this case has shot positions shifted by 24 m to the right, with shot and receiver depths now changed to 12 and 20 m, respectively. The new monitor Ricker source peak frequency is changed to 12.1 Hz. To isolate the effects of survey acquisition changes from the effects of random noise, the model difference is inverted from a clean synthetic. The result of Figure 3.4, in good agreement with the true perturbation of Figure 3.1, demonstrates the robustness of cross-updating with respect to non-random survey repeatability issues.

The new simultaneous inversion and cross-updating techniques provide robust alternatives to the existing time-lapse FWI methods. Applying the cross-updating method to synthetic data sets with variable amount of noise achieved a significant reduction of artifacts in the model difference compared to the parallel, sequential, and double difference methods. However, choosing the weighting operator $\mathbf{W}$ in the simultaneous inversion has to be controlled by prior knowledge of where production--induced velocity changes are likely to occur. A constant weight $\mathbf{W}$ with an L_2 model-difference regularization term (3.10) may result in an excessive penalization and smoothing of production-induced effects, destroying the blockiness of the model difference. In the next section we described a TV regularization approach that addresses this issue. The cross-updating method offers an attractive alternative to the regularized simultaneous inversion as it does not require additional regularization parameters.

In addition to achieving better results than the double difference method for noisy synthetics, cross-updating is less sensitive to repeatability issues that are due to differences in acquisition geometry and sources. The latter may require a cross-equalization of different data vintages (Ayeni and Biondi, 2012) prior to double differencing while the simultaneous inversion and cross-updating do not require data cross-equalization as these methods operate in the model space.

TOTAL-VARIATION MODEL-DIFFERENCE REGULARIZATION

In this section I will demonstrate the simultaneous time-lapse FWI with a total-variation model-difference regularization applied to noisy synthetic data.

As in the previous section, the same background Marmousi velocity model is used as a baseline. Production-induced velocity changes are modeled as a negative -200 m/s perturbation at about 4.5 km inline 800 m depth, and a positive 300 m/s perturbation at 6.5 km inline, 1 km depth, illustrated in Figure 3.7. While the whole Marmousi model is inverted, only model differences for the section between

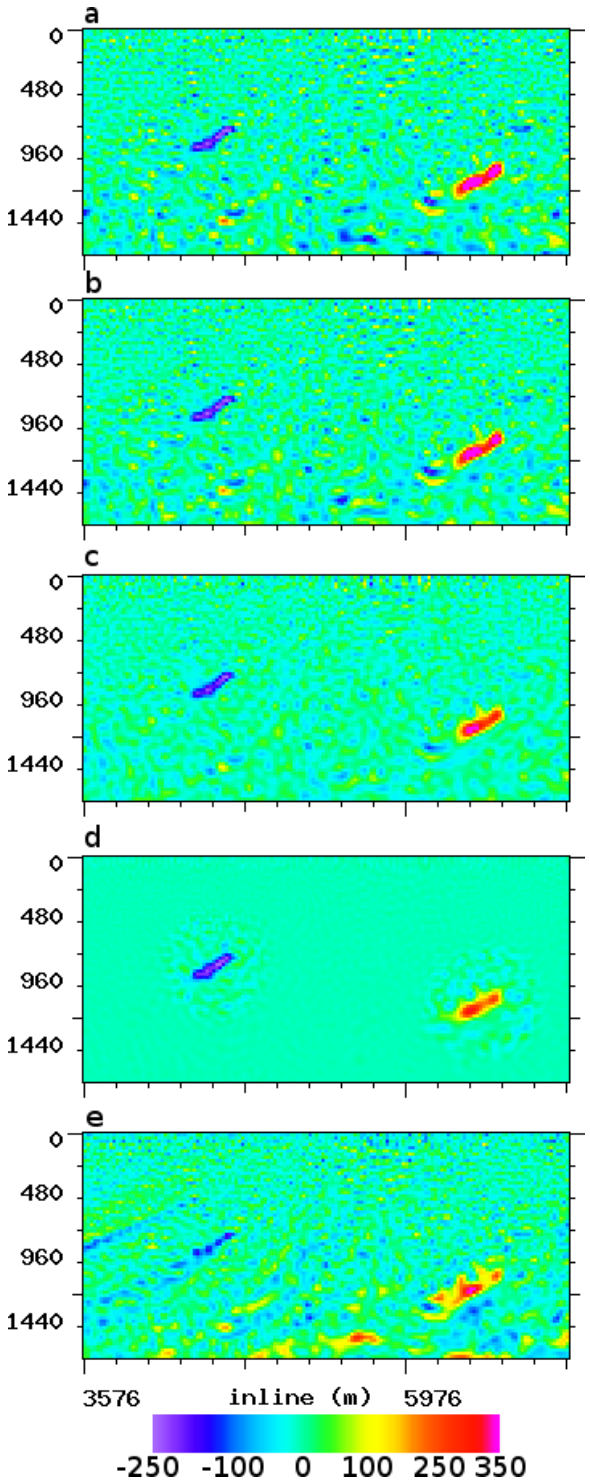


Figure 3.2: Model difference inverted from a 14 dB SNR synthetic with matching baseline and monitor acquisition geometries using (a) iterated parallel difference; (b) iterated sequential difference; (c) cross-updating; (d) regularized simultaneous inversion; (e) iterated double difference. [**CR**] jfwi/. n2

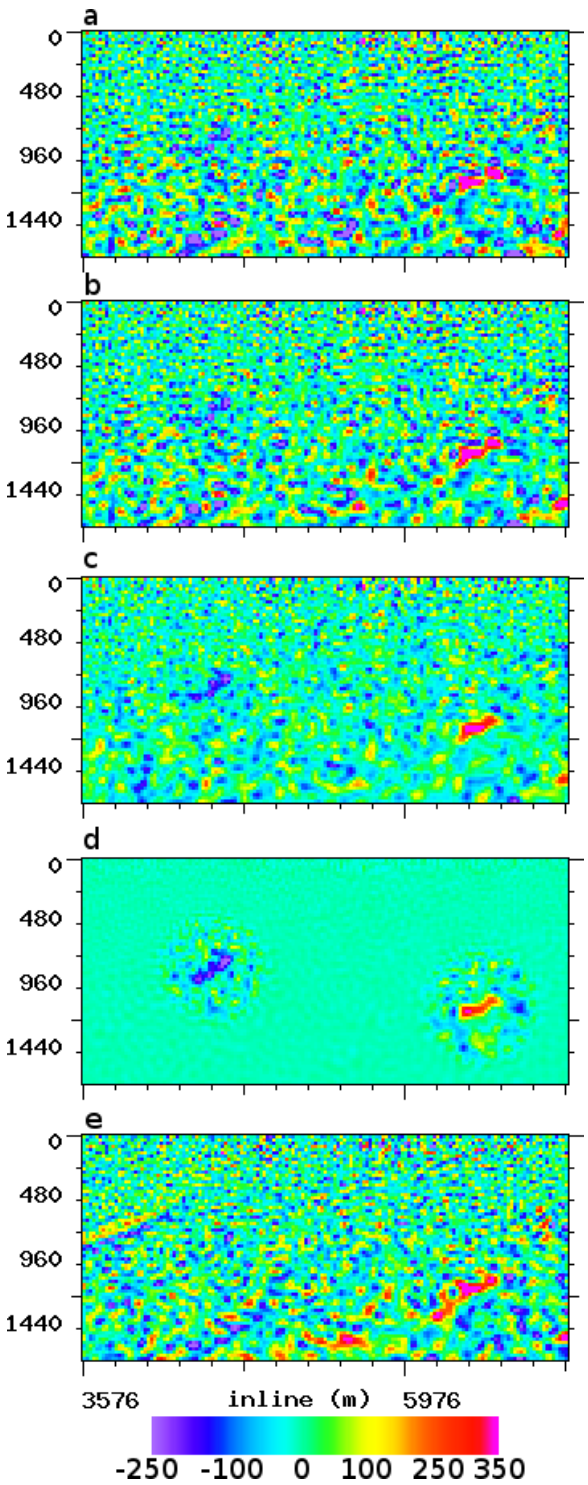


Figure 3.3: Model difference inverted from a 7 dB SNR synthetic with matching baseline and monitor acquisition geometries using (a) iterated parallel difference; (b) iterated sequential difference; (c) cross-updating; (d) regularized simultaneous inversion; (e) iterated double difference. [**CR**] jfwi/. n4

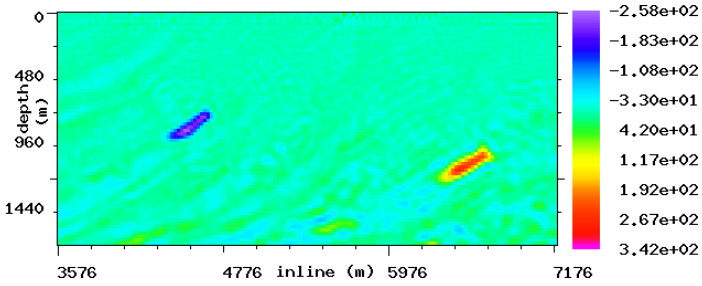


Figure 3.4: Model difference inverted from a clean synthetic for different baseline and monitor acquisition geometries and sources. [**CR**] jfwi/. repeatx

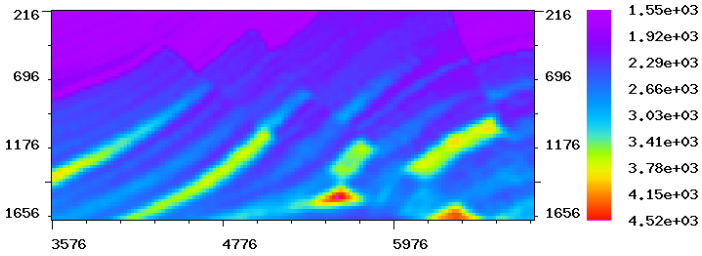


Figure 3.5: Target area of the baseline model inverted from a clean synthetic. [**CR**] jfwi/. zcleanbase

the approximate inline coordinates 3.5 km and 7 km to the depth of approximately 1.7 km are shown here. The inversion is carried out in the frequency domain for 3.0, 3.6, 4.3, 5.1, 6.2, 7.5, 9.0, 10.8, 12.8, and 15.5 Hz, where the frequencies are chosen based on the estimated offset to depth range of the data (Sirgue and Pratt, 2004). The baseline acquisition has 192 shots at a depth of 16 m with a 48 m spacing, and 381 receivers at a depth of 15 m with a 24 m spacing. The minimum offset is 48 m. The source function is a Ricker wavelet centered at 10.1 Hz. Absorbing boundary conditions are applied along the entire model boundary, including the surface (thus suppressing multiples). A smoothed true model is used as a starting model for the initial baseline inversion. The smoothing is performed using a triangular filter with a 20-sample half-window in both vertical and horizontal directions.

Random Gaussian noise is added to the noise-free synthetic data to produce a noisy dataset with 7 dB signal-to-noise ratio. The results of baseline model inversion from the clean and 7 dB SNR synthetic data are shown in Figures 3.5 and 3.6.

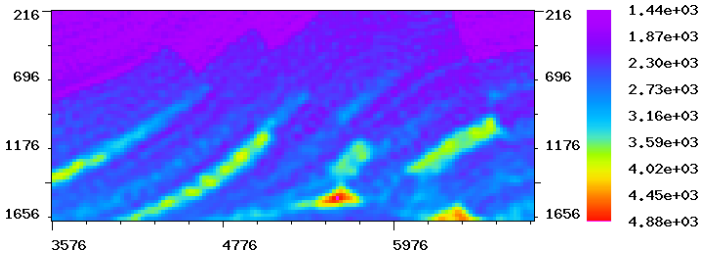


Figure 3.6: Target area of the baseline model inverted from a 7 dB SNR synthetic. In both cases the baseline model is reconstructed reasonably well, however, errors due to noise are comparable in magnitude to production-induced effects. [**CR**] jfwi/. zn4base

The noisy monitor dataset is generated for the model perturbation of Figure 3.7, using the same acquisition geometry and source wavelet. Results for model difference inversion from the 7 dB SNR synthetic datasets using various methods are shown in Figures 3.8, 3.9, and3.10. The model-difference regularization weights $\mathbf{W}$ in (3.15) are set to 1 everywhere in the modeling domain. The two terms in (3.11) are of the same magnitude and therefore α and β are set to 1. Parameter δ is set here to 10^{-6} but can be varied for different acquisition source and geometry parameters. The result of the initial baseline inversion is supplied as a starting model for both $\mathbf{m}_b$ and $\mathbf{m}_m$ in the simultaneous inversion.

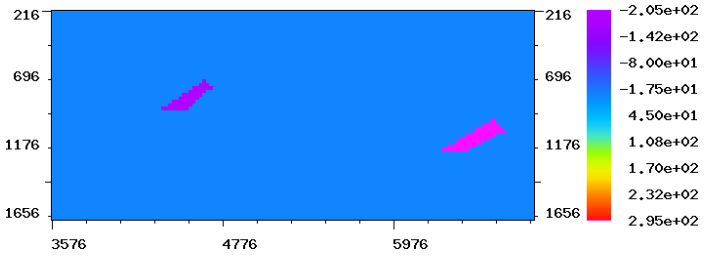


Figure 3.7: True velocity differences consist of a negative (-200 m/s) perturbation at about 4.5 km inline 800 m depth and a positive (300 m/s) perturbation at 6.5 km inline, 1 km depth. [**CR**] jfwi/. z4true

In all the inversions, up to 10 iterations of the nonlinear conjugate gradients algorithm (Nocedal and Wright, 2006) are performed for each frequency. However,

because the L_1 component (3.15) of the objective function is not smooth, differentiation of (3.16) may result in an overflow where the gradient magnitude is very small. To avoid this, we use gradient smoothing similar to "Iteratively Re-weighted Least Squares" (Aster et al., 2011): whenever the gradient magnitude (3.16) is below some threshold value ϵ, division by (3.16) is substituted with division by ϵ. The threshold $\epsilon = 10^{-5}$ was chosen in our experiments to be less than $|\Delta v/v^2|$, with v as the baseline model velocity within the target area, and Δv as a lower bound on production-induced velocity changes. Note that alternative solution techniques such as ADMM (see appendix B) for TV-regularized problems can be employed in TV-regularized (time-lapse) FWI (Maharramov and Biondi, 2014a) where a fixed small ϵ-thresholding may adversely impact convergence. In appendix B, I describe a new computationally efficient technique for solving large-scale $L_1 - TV$ regularized problems ("Compressive Conjugate Directions").

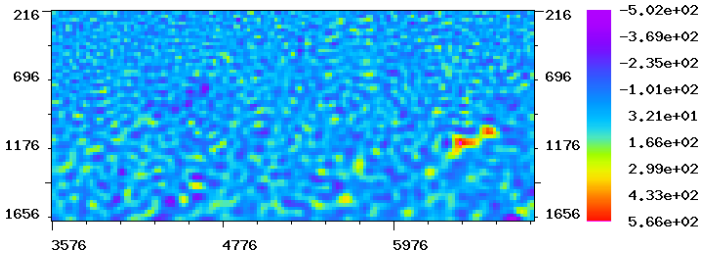


Figure 3.8: Model difference inverted using iterated sequential difference. [**CR**] jfwi/. zn4seqdiff

The result of applying iterated sequential difference (Maharramov and Biondi, 2014c) to the two datasets is shown in Figure 3.8. The corresponding cross-updating and TV-regularized simultaneous inversion results are shown in Figure 3.9,3.10. While cross-updating demonstrates certain robustness with regard to uncorrelated noise in the data and computational artifacts (note the quantitative improvement of the recovered difference magnitudes in Figure 3.9 compared with Figure 3.8), the TV-regularized simultaneous inversion (Figure 3.10) achieves a significant further improvement by reducing oscillatory artifacts and honoring the blocky nature of the model difference. Note, however, that imposing a model-difference regularization results in underestimated magnitudes of velocity perturbations. Maharramov and

Biondi (2015a) discuss a mathematical technique for improving amplitude recovery of TV-regularized inversion by using a "cartoon+texture" model decomposition technique (Meyer, 2001).

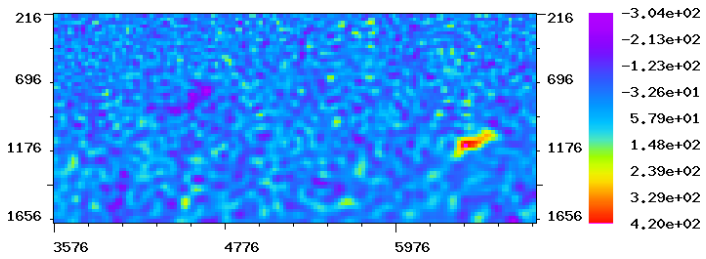


Figure 3.9: Model difference recovered using cross-updating. [**CR**] jfwi/. zn4x

The new TV-regularized simultaneous inversion technique is a more robust further development of the joint inversion methods described in the previous section. Use of TV regularization in the simultaneous inversion allows robust recovery of blocky production-induced changes, and penalizes unwanted model oscillations that may mask useful production-induced changes.

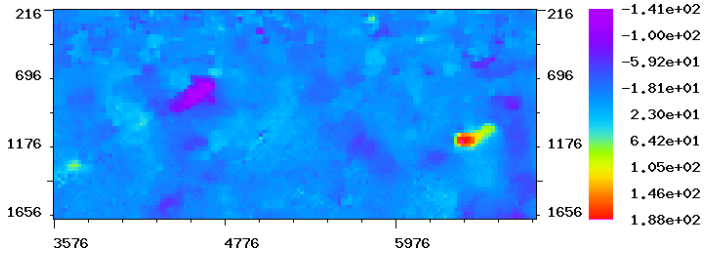


Figure 3.10: Model difference recovered using the simultaneous inversion with a TV-regularized model difference. Note the higher accuracy and stability to random noise of the TV-regularized simultaneous inversion. [**CR**] jfwi/. zn4tv

Chapter 4

Sensitivity analysis of simultaneous phase-only reflection time-lapse FWI

"Necessity is the mother of invention."

Plato

Examples of time-lapse inversion that I provided in chapter 3 demonstrate applications of time-lapse FWI with model-difference regularization to inverting velocity changes from both phase *and* amplitude information. Large acquisition offsets compared to target depth meant that our data had both reflected and diving-wave energy. This chapter is dedicated to applications of a *phase-only* simultaneous FWI with a model-difference regularization to reflection-only seismic data, and studying its sensitivity to the wrong baseline and monitor starting models. Additionally, I demonstrate an application the multi-model technique of Maharramov and Biondi (2015a) to simultaneously inverting the effects of overburden dilation and reservoir compaction.

APPLICATION TO REFLECTION DATA

Why do we need a special chapter dedicated specifically to phase-only time-lapse FWI of reflection data? Examples in chapter 3 involved phase and amplitude full-waveform inversion using both reflection and refraction data, making the inversion very well-constrained and utilizing the full high-resolution power of FWI. Many important examples of time-lapse applications indeed involve similar inversion targets, and it was originally expected that applications like that would be the primary target of the developed joint time-lapse FWI technique.

However, the field time-lapse data obtained by Stanford Exploration Project for applying and testing my method forced me to change the original plan. Survey and target parameters of the field data precluded use of diving waves. Uncertainty in the starting model meant that sensitivity to inaccurate starting models and cycle skipping could be an issue. It was at this time that I developed the theory of chapter 2, aiming to adapt my inversion technique to the challenging dataset in hand. The theory demonstrated a sensitivity of phase-only inversion to small-magnitude blocky velocity changes in the overburden that were of particular interest for the field data in question, see (2.24). Qualitative accuracy of the inversion for wrong FWI starting models was established as well, see (2.33). The purpose of this chapter is to provide numerical corroboration of the theoretical findings of chapter 3, and lay the groundwork for field data analysis presented in chapter 5.

SYNTHETIC MODEL

Chapter 3 has demonstrated effective recovery of blocky velocity anomalies from long-offset acquisitions in the presence of noise and repeatability issues—see also Maharramov and Biondi (2014e,c). In this chapter I demonstrate the recovery of blocky anomalies in the more challenging case of phase-only inversion of reflection-only synthetic data. Conceptually our synthetic example is similar to the field data studied in chapter 5 and Maharramov and Biondi (2015b); Maharramov et al. (2015a).

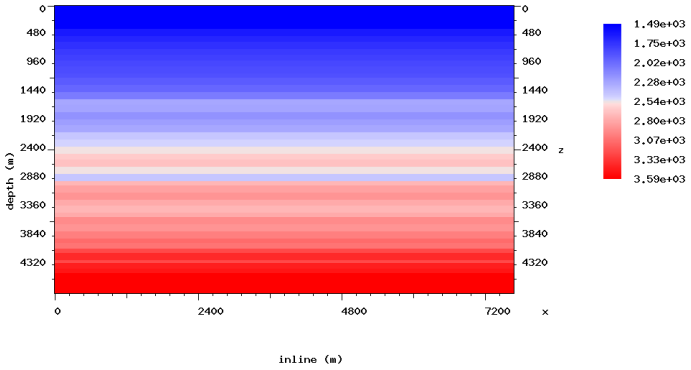


Figure 4.1: The true baseline model. A flat reflector model to study the sensitivity of FWI of short-offset reflection data to small velocity perturbation in the overburden. [**ER**] sensitivity/. 5truebase

As a baseline model I use the flat reflector model of Figure 4.1. The target reflector (reservoir) is located at a depth of 3900 m, the monitor (perturbed) model has two velocity anomalies—a positive +300 m/s change due to compaction and fluid substitution within the reservoir, and a blocky negative velocity change in the overburden above the reservoir, peaking at −50 m/s (see Figure 4.2). No physical reflector movement is prescribed.

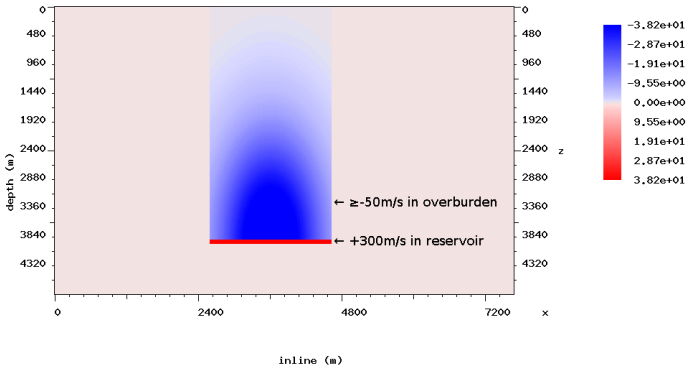


Figure 4.2: The true model difference is a combination of a positive +300 m/s velocity change in a target reflector at a depth of 3900 m, and a negative velocity change in the overburden above the reflector, peaking at −50 m/s. In this chapter I investigate the sensitivity of simultaneous time-lapse FWI to small-magnitude blocky velocity changes in the overburden. [**ER**] sensitivity/. 5truediff

For generating synthetic data I used a towed streamer acquisition geometry with a maximum offset of 5 km. Note that the relatively small maximum offset to target depth ratio for the model of Figure 4.1 means the target is illuminated only by reflection data. Figure 4.3 shows common-midpoint gathers above the center of the

target reservoir for the baseline (blue) and monitor (red) surveys. Signals that travel vertically through the anomaly are delayed the most, while larger-offset reflections are delayed less as they largely bypass the anomaly. Reflections above the anomaly show no time shift.

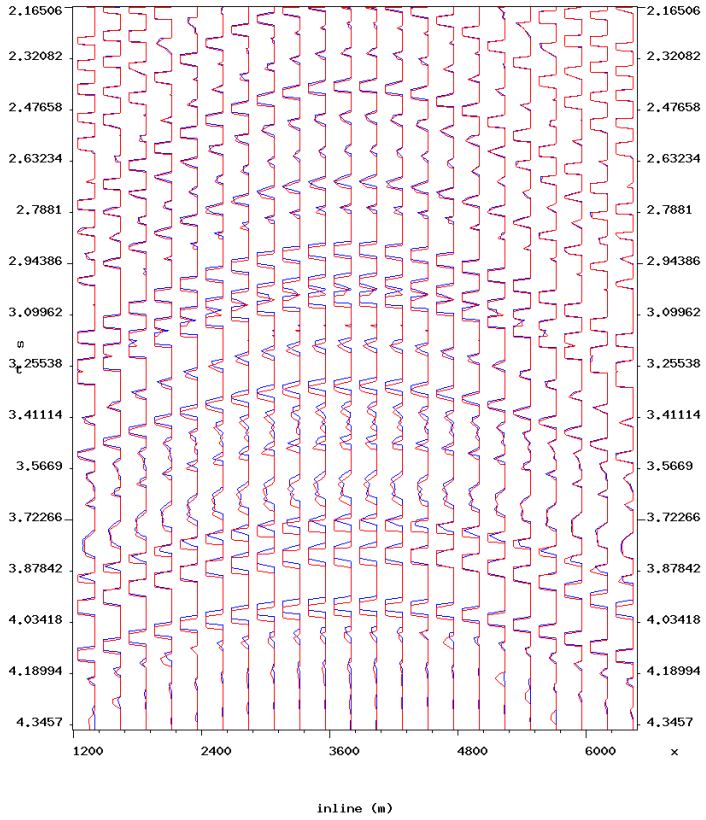


Figure 4.3: Time shifts observed in common-midpoint gathers centered above the target reservoir (blue is baseline, red is monitor). Travel times of the monitor near-offset reflections traveling through the negative velocity anomaly of Figure 4.2 are slightly delayed. [**CR**] sensitivity/. 5timeshifts

Figure 4.3 represents a typical overburden dilation effect on arrival times: timeshifts accumulate with distance traveled through the low-velocity anomaly, peaking at small but tangible delays to the order $\sim$ 10ms near the reservoir top (Johnston, 2013; Hatchell and Bourne, 2005).

TIME-LAPSE INVERSION FROM REFLECTION DATA

The results of parallel difference and cross-updating (see chapter 3) are shown in Figures 4.4 and 4.5. Note that neither method succeeds in recovering the blocky anomaly.

The FWI starting model used in these experiments was produced by smoothing the true model using a 1920 m smoothing window.

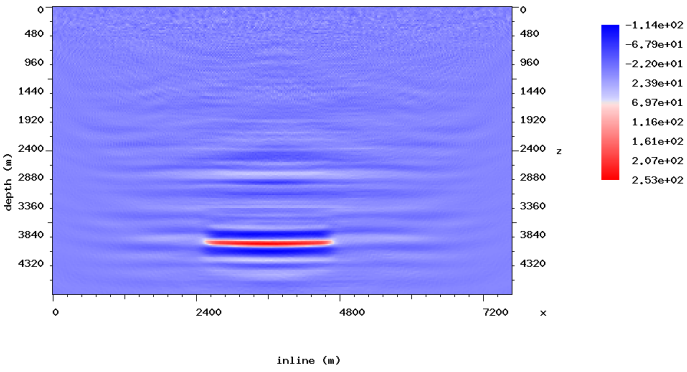


Figure 4.4: The parallel difference method fails to resolve the long-wavelength velocity changes of Figure 4.2, and produces negative short-wavelength artifacts around the target reflector. [**CR**] sensitivity/. 5pardiff

The result of simultaneous inversion with a total-variation model-difference regularization (3.11,3.15) is shown in Figure 4.6. The result is qualitatively accurate although peak magnitudes are underestimated due to regularization, as discussed in chapter 3 and by Meyer (2001); Maharramov and Biondi (2015a).

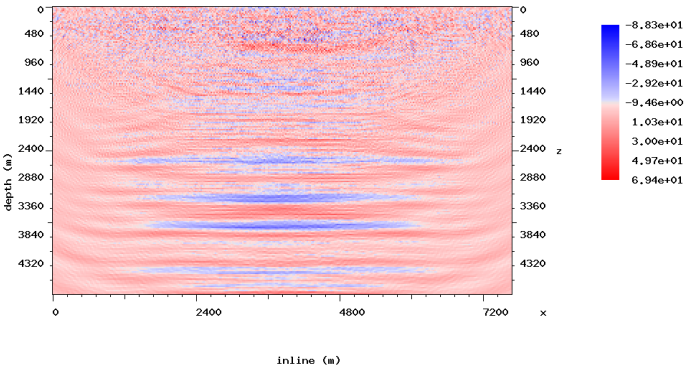


Figure 4.5: The cross-updated FWI method cross-equalizes the baseline and monitor model but still fails to resolve the long-wavelength overburden changes of Figure 4.2. [**CR**] sensitivity/. 5xdiff

To assess effectiveness of the inversion, in Figures 4.12(a) and 4.13(b) I show monitor images migrated using the true monitor and true baseline models, respectively. Note that the overestimated velocities in the overburden result in a downward reflector shift in Figure 4.13(b). However, migrating the monitor data using the sum of the baseline model and the *inverted* blocky anomaly of Figure 4.6 results in the image of Figure 4.13(a): the downward shift of reflectors in the overburden is now significantly

reduced.

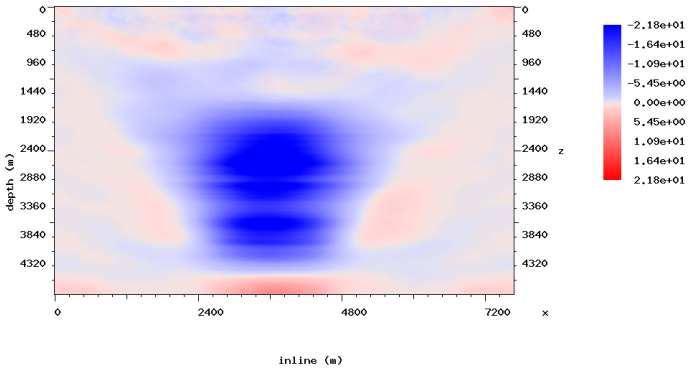


Figure 4.6: Simultaneous FWI with a total-variation model-difference regularization resolves the long-wavelength overburden changes of Figure 4.2, but underestimates the maximum change, depending on the regularization strength. [**CR**] sensitivity/. 5tvinverted

To recover the short-wavelength changes within the reservoir, I supplied the result of Figure 4.6 as a model-difference prior to inversion (3.6,3.10), with a Tikhonov regularization of the model difference. Note that the resulting model features both long and short-wavelength velocity perturbations. And again, the reservoir perturbation is underestimated due to the strong regularization. Note the leakage of a small-amplitude velocity anomaly below the reservoir. This is the result of a lack of strong reflectors below the reservoir— see Figure 4.3.

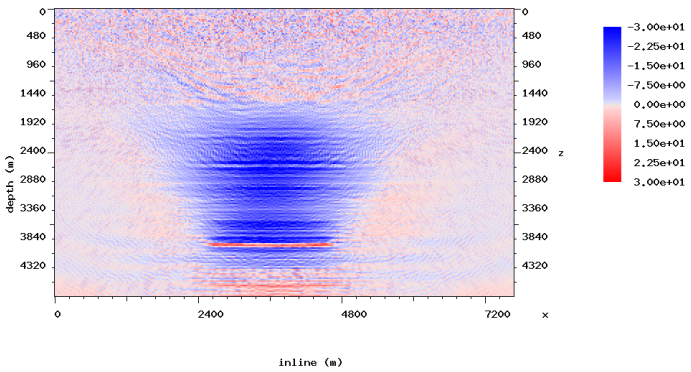


Figure 4.7: Simultaneous FWI using Tikhonov model-difference regularization with the long-wavelength inversion of Figure 4.6 supplied as a prior. Note that such a multiscale approach can now resolve the short-wavelength positive-velocity changes of Figure 4.2. Strong Tikhonov regularization results in underestimated velocity changes within the reservoir but correctly locates the anomalies. [**CR**] sensitivity/. 30multi

SENSITIVITY TO CYCLE-SKIPPING

Time-lapse scattering theory predicts that a qualitatively accurate inversion of blocky anomalies from phase information can be achieved despite inaccurate FWI starting models—see (2.33). Here I study the method's sensitivity to starting models and cycle-skipping by inverting two isolated anomalies shown in Figure 4.8. I chose a more challenging test than the previous model of Figure 4.2 to make it more sensitive to the accuracy of our inversion: inversion artifacts may not only distort each anomaly but also violate their separation.

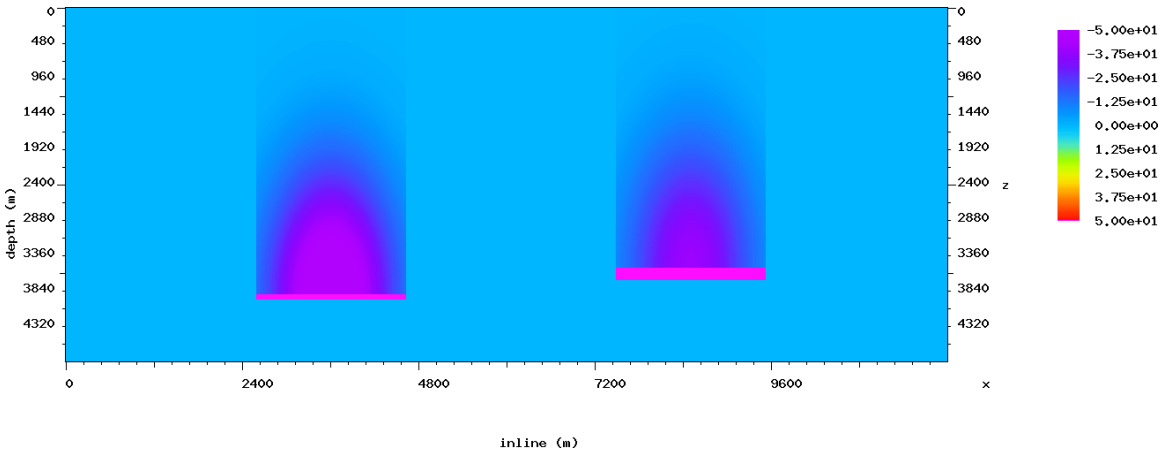


Figure 4.8: True model difference for demonstrating the inversion of multiple overburden anomalies. [**ER**] sensitivity/. twoanom

Figure 4.9 demonstrates the recovery of two separate overburden anomalies when the FWI starting velocity is a smoothed true velocity produced using a 1920 m smoothing window. As before, the inversion is in good agreement with the true difference. Here I invert only for the blocky component, ignoring the spiky component within the reservoir that is caused by fluid substitution and compaction.

The result of starting FWI from a wrong velocity (that contains a wrong high-wavenumber component) is shown in Figure 4.10. I deliberately used a weak regularization parameter for model-difference regularization to demonstrate the effect of diverging baseline and monitor models on the inverted model difference. Either one or both of the baseline and monitor inversions cycle-skipped, and the models

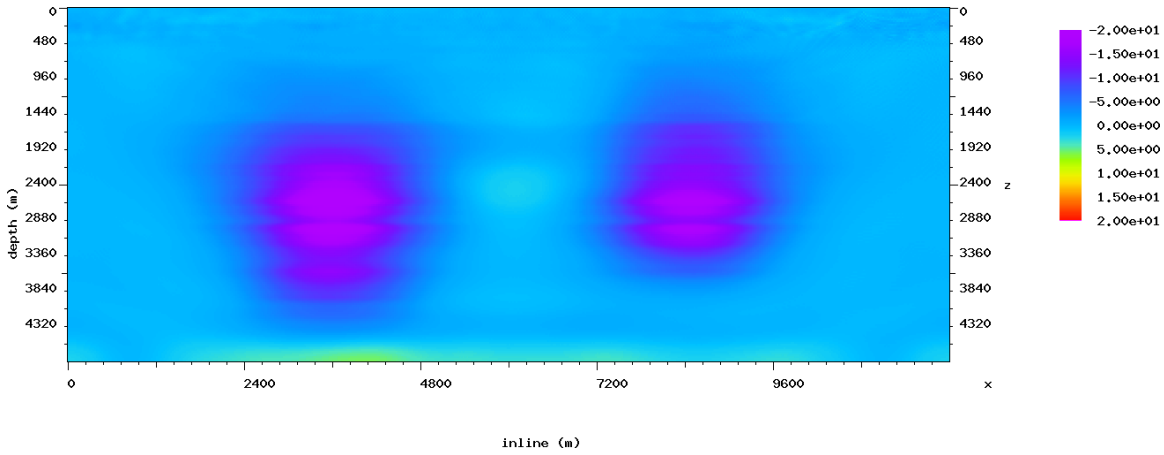


Figure 4.9: Inversion of the two long-wavelength overburden anomalies of Figure 4.8 using simultaneous time-lapse FWI with a total-variation model-difference regularization. [**CR**] sensitivity/. twoanominv

diverged from each other sufficiently far to contaminate the model difference with strong artifacts. The artifacts almost completely masked one of the anomalies and contaminated areas not affected by production with false positives.

Figure 4.11 shows the result of using a stronger TV regularization, as prescribed by chapter 2. By choosing a stronger model-difference regularization, we ensure a greater conformity between the two models, that the two models cycle-skip "in synchrony". As a result, we are still able to qualitatively recover the anomalies, although with strongly underestimated velocities—compare with equation (2.33).

Note that magnitude of the slowness change is underestimated as a result of the stronger model-difference regularization in (3.15). Adopting the recommendations of Meyer (2001); Maharramov and Biondi (2015a) can alleviate this problem so that *slowness* magnitudes can be inverted more accurately. However, because of (2.33), the wrong velocity background would still result in a quantitatively wrong velocity perturbation. Travel-time delays can be translated into accurate slowness changes even for wrong backgrounds as shown in (2.24)[1] but a quantitatively accurate estimate

[1]so long as the reflected signal is present in both predicted baseline and monitor data and differs by less than $2\pi/\omega$.

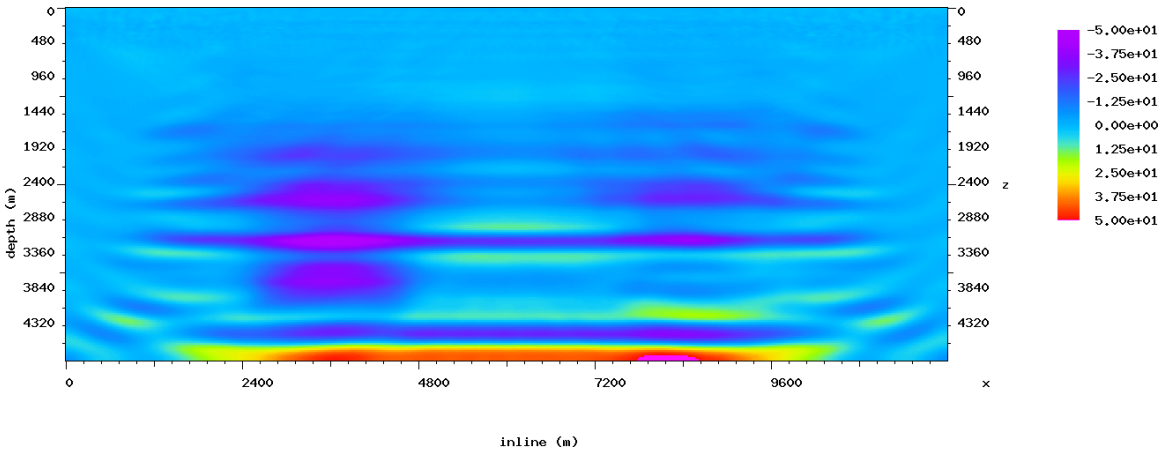


Figure 4.10: Inversion of the two long-wavelength overburden anomalies of Figure 4.8 starting from a wrong initial model and using weak regularization (a small regularization parameter). FWI cycle skipped, and the baseline and monitor inversions diverged, contaminating the difference with cycle-skipping artifacts. [**CR**] sensitivity/. twoanombadweak

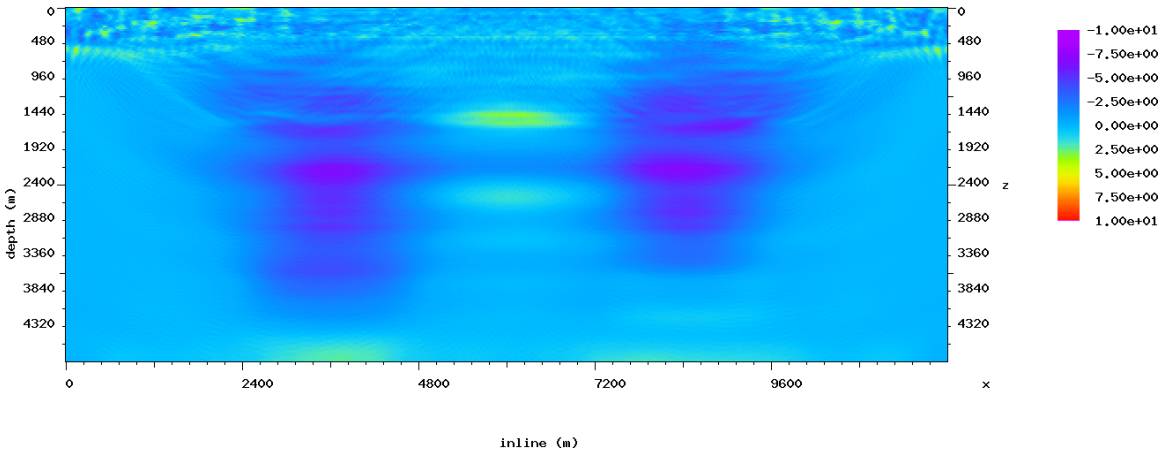


Figure 4.11: Inversion of the two long-wavelength overburden anomalies of Figure 4.8 starting from the same bad initial model but using a stronger regularization (a larger regularization parameter). FWI still cycle skipped, however, the strong model-difference regularization kept baseline and monitor within the characteristic wavelength of the overburden anomalies. The resulting model difference inversion is qualitatively accurate (compare with Figure 4.9), albeit stronger regularization has resulted in underestimated velocity magnitudes. [**CR**] sensitivity/. twoanombadstrong

of the velocity difference still requires knowledge of accurate background velocity.

Numerical experiments have born out the theoretical predictions of chapter 2 with regard to the sensitivity of phase-only time-lapse full-waveform inversion of reflection data to realistic production-induced time-shifts, and the method's stability with regard to inaccurate FWI starting models. In the next chapter I apply the method to field data for a reservoir where production was associated with a significant overburden dilation and reflection travel-time changes.

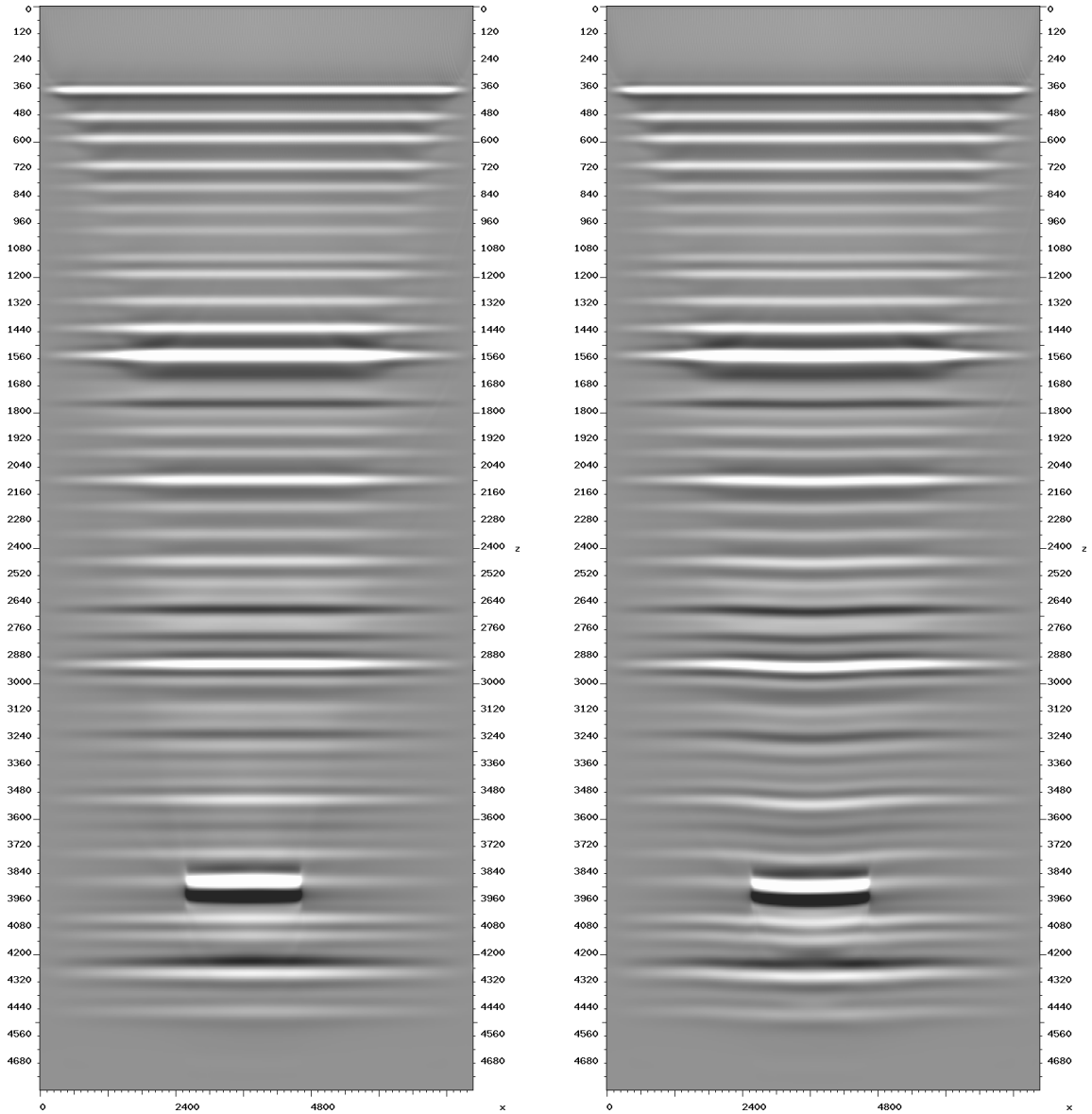


Figure 4.12: (a) True monitor image. (b) Monitor migrated using the baseline velocity model. Note that overestimated velocity in the overburden results in a downward reflector shift in the right image. [**CR**] sensitivity/. montrueimg,monwithbase

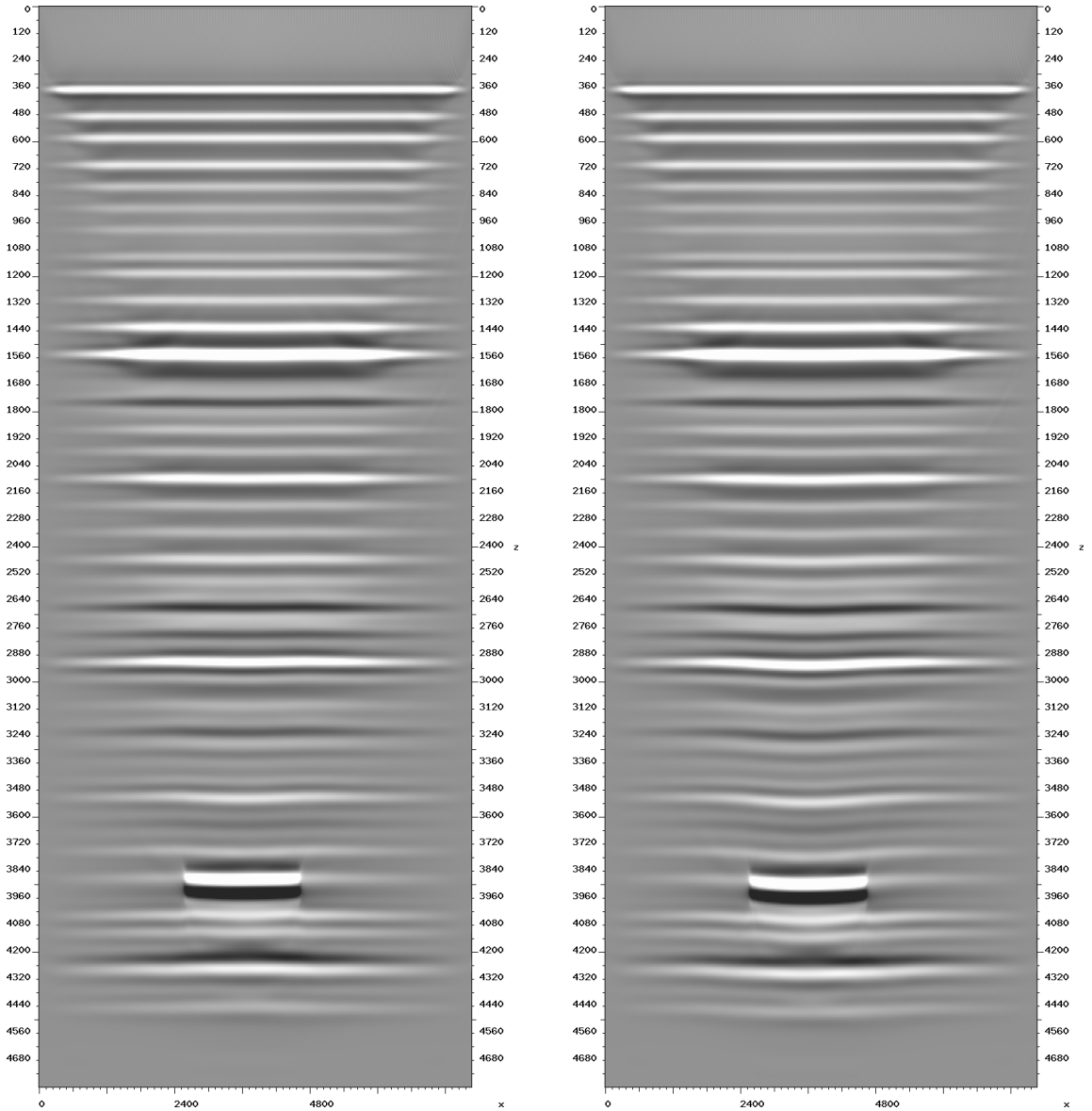


Figure 4.13: (a) Monitor image migrated using the sum of the baseline model and inverted model difference of Figure 4.6. (b) Monitor migrated using the baseline velocity model. Note that reflector shift in the overburden has been significantly reduced in the left image. [**CR**] sensitivity/. improved,monwithbase

Chapter 5

Case study: Gulf of Mexico Genesis Field

"I pass with relief from the tossing sea of Cause and Theory to the firm ground of Result and Fact."

Winston Churchill

This chapter presents a field data application of the method developed in chapters 2–4. Using phase-only simultaneous FWI with a total-variation model-difference regularization, I invert long-wavelength small-magnitude subsurface velocity changes induced by reservoir compaction at the Genesis field. I identify negative velocity changes associated with overburden dilation and demonstrate that the results are stable with respect to the amount of regularization applied. The results are consistent with well measurements and earlier estimates of time strain in the overburden.

GENESIS FIELD

The Genesis Field, operated by Chevron, is located 150 miles southwest of New Orleans in the Green Canyon area of the central Gulf of Mexico, in approximately 770-830 m of water (Magesan et al., 2005). Oil was found in several late Pliocene through early Pleistocene deep-water reservoirs. Most of the field's oil and gas reserves are in the early Pleistocene Neb 1, Neb 2, and Neb 3 reservoirs that are the primary subject of this study (see Figure 5.1).

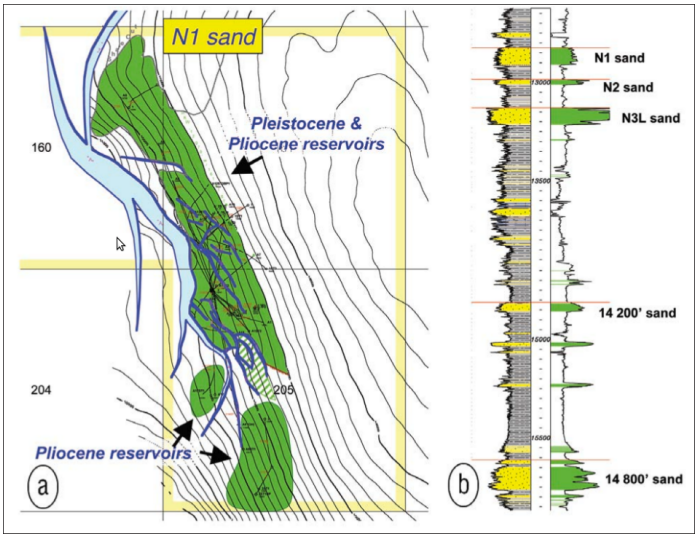


Figure 5.1: The Genesis field (from Magesan et al. (2005)). [**NR**] genesis/. map2

First oil production began in January 1999. A 3D seismic survey was shot in 1990, and a time-lapse 3D survey was shot in October 2002 with the aim of improving field management (Hudson et al., 2005; Magesan et al., 2005). Cumulative production from the field at the time of the monitor survey was more than 57 MMBO, 89 MMCFG, and 19 MMBW (Hudson et al., 2005).

In addition to fluid substitution effects, producing reservoirs compact, thereby increasing the depth to the top of the reservoirs and causing overburden dilation (Johnston, 2013). A time-lapse study performed by Chevron (Hudson et al., 2005) indicated significant apparent kinematic differences in the Pleistocene reservoir interval. Time shifts were observed both for the producing reservoirs and Illinoisan wet

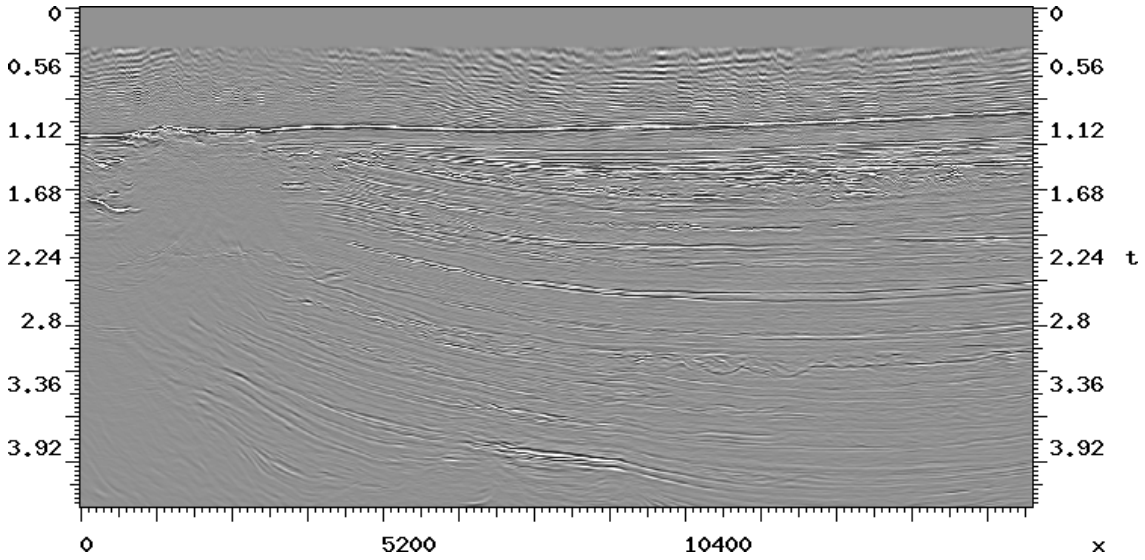


Figure 5.2: A north-south inline section of the baseline Genesis image produced by Chevron (vertical axis two-way travel time in seconds, horizontal axis inline meters). [**NR**] genesis/. cvxbase

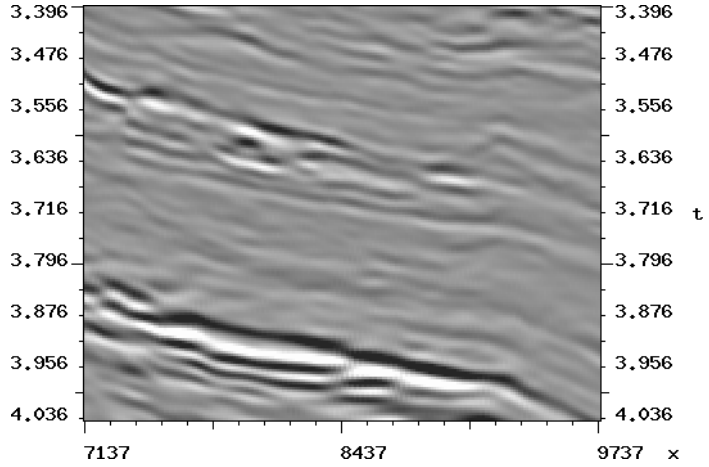


Figure 5.3: Monitor and baseline image-difference obtained from the 3D time-migration images provided by Chevron which corresponds to the inline section of Figure 5.2. Production-induced changes stand out at approximately 3.5 s (wet Illinoisan sands) and 4 s two-way travel times—stacked Neb 1, 2, and 3 reservoirs. [**NR**] genesis/. cvxdiff

sands above Neb 1 (see Figure 5.4). Kinematic differences were attributed to a time shift caused by subsidence at the top of the uppermost reservoir, subsidence of the overburden, and overburden dilation (Hudson et al., 2005).

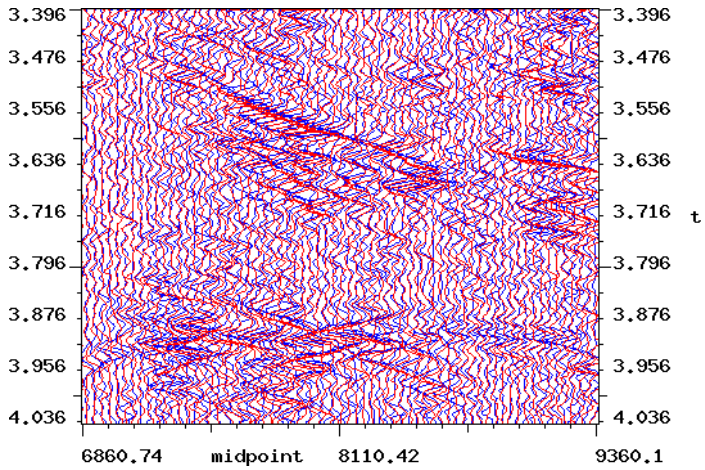


Figure 5.4: Production-induced changes resulted in measurable time-shifts between the surveys. Shown here are time-shifts between the baseline (blue) and monitor (red) common-offset gathers, 1074 m offset. [**CR**] genesis/. 6timeshifts

Processing parameters for the baseline and monitor surveys and subsequent time-lapse processing by Chevron were described by Magesan et al. (2005). The baseline survey had a maximum offset of 5 km, and the monitor survey had a maximum offset of 7.3 km. Both surveys used a bin size of 12.5 m by 37 m. For the purpose of time-lapse analysis, the acquired data had been subjected to pre-processing and imaging steps that included data equalization, spherical divergence correction, source and receiver statics, global phase rotation, time shift, amplitude scaling, global spectral matching, and cross-equalization (Magesan et al., 2005).

These pre-processed data were used by Chevron in Kirchhoff time migration of the baseline and monitor surveys to produce 3D images. A single inline section of the baseline image is shown in Figure 5.2. The corresponding monitor and baseline image difference is shown in Figure 5.3. As noted by Hudson et al. (2005), the image difference is contributed to by time shifts at the Illinoisan sands (upper event) and Neb 1 (lower event) in Figure 5.3—compare with Figure 1 of Hudson et al. (2005).

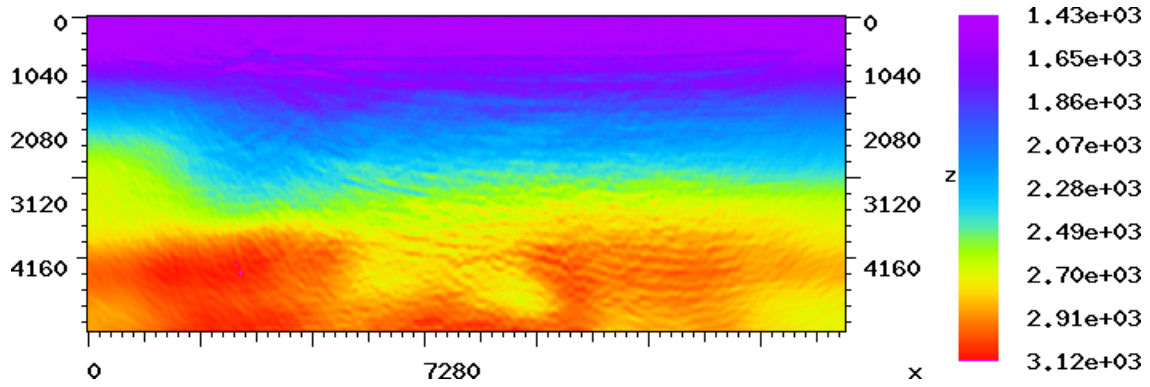


Figure 5.5: Inverted baseline velocity model (m/s). FWI resolved fine model features and oriented them along the dip structure of the image in Figure 5.2 (vertical axis depth meters). [**CR**] genesis/. 6base

2D INVERSION

The purpose of this experiment was to see whether joint regularized time-lapse FWI could resolve some of the production-induced model differences, thus providing additional insight into reservoir depletion patterns and optimal infill drilling strategies. As my first processing step, I performed separate baseline and monitor 2D full-waveform inversion of a single inline section. I extracted single north-south inlines corresponding to the image in Figure 5.2 from both surveys and sorted them into shot gathers with a minimum offset of 350 m and a maximum offset of 4,700 m. This provided 1,264 shots per survey with up to 175 receivers per shot. A frequency-domain 2D FWI (Sirgue et al., 2008, 2010b) was conducted over the frequency range of 3–30.7 Hz. Frequency spacings were selected using the technique of Sirgue and Pratt (2004). As noted above, the data provided to Stanford Exploration Project had undergone amplitude pre-processing that included a spherical divergence correction. Furthermore, accurate handling of the amplitudes in 2D FWI of 3D field data requires a 3D-to-2D data transformation (Auer et al., 2013). Because the data exhibited significant time-shifts at the reservoir level (Hudson et al., 2005) that can be readily observed even at large offsets (see Figure 5.4), I decided to use a "phase-only" inversion and ignored amplitude information in the data (Fichtner, 2011).

The result of baseline inversion is shown in Figure 5.5. To build a starting model for the FWI, I converted Chevron's RMS time-migration velocity model to an interval velocity using the Dix equation, and smoothed the result using a triangular filter with a 41-sample window. Observe that FWI succeeded in resolving fine features, and oriented them consistently along the dip structure of the time-migrated image in Figure 5.2. Close-up views of the model area covering both the Illinoisan sands and the reservoirs are shown in Figures 5.7(a) and 5.7(c).

The result of parallel differencing is shown in Figure 5.6(a). Although significant model changes appear to be concentrated around the target area, this result is not interpretable, neither qualitatively nor quantitatively, because it is contaminated with oscillatory artifacts and it overestimates the magnitudes of velocity perturbations. This result is consistent with my earlier assessment of conventional time-lapse FWI techniques tested on synthetic data—see chapter 3 and (Maharramov and Biondi, 2014c; Maharramov et al., 2015b).

Next, I computed the simultaneous, TV-regularized time-lapse full-waveform inversion formulation (3.11,3.15). I set $\alpha = \beta = 1$ and carried out multiple experiments with the value of the regularization parameter δ ranging from $\delta = 100$ to $\delta = 1000$. The weighting operator $\mathbf{W}$ was set to 1 inside the larger target area shown in Figures 5.6(a) through 5.7(b), and tapered off to zero outside.

The results of inverting the model difference for $\delta = 100$, 500 and 1000 are shown in Figures 5.6(b), 5.6(c), and 5.6(d), respectively. Gradual increase of the regularization parameter results in the removal of most model differences with the exception of a negative velocity perturbation in the overburden, peaking at approximately 3.6 km and 3.9 km depth (see Figures 5.7(b) and 5.7(d)). Such perturbations are consistent with overburden dilation due to the compaction of stacked reservoirs, with more significant dilation in the wet Illinoisan sands than the surrounding shales (Rickett et al., 2007). The zone of negative velocity change appears to extend upward into the overburden in a direction roughly orthogonal to the reservoir dip—see Figure 5.6(d). Two negative velocity changes at approximately 10 and 11.5 km inline persist with

increasing regularization, and may represent dilation effects associated with the production from deeper reservoirs—compare with Figure 3 of Rickett et al. (2007).

Figure 5.8 plots the inverted velocity change against difference of the RTM-migrated monitor and baseline images. Migration was conducted using the same (starting FWI) velocity for both baseline and monitor. The image difference is strongest at a 3.6 km depth, corresponding to the Illinoisan sands, and about 3.9 km—the overburden above Neb 1.

The estimated maximum negative velocity change of -45 m/s above the stacked reservoirs is consistent with the earlier estimates of time strain in the overburden (Rickett et al., 2007). Indeed, local time strain, physical strain and partial velocity change are related by the equation (Hatchell and Bourne, 2005)

$$\frac{d\tau}{dt} \approx \frac{\Delta t}{t} = \frac{\Delta z}{z} - \frac{\Delta v}{v}, \tag{5.1}$$

where τ, t, z, and v denote the observed time shift, vertical travel time, depth, and velocity. Assuming, following Hatchell and Bourne (2005), that

$$\frac{\Delta v}{v} = -R\frac{\Delta z}{z}, \tag{5.2}$$

where the factor R is estimated to be 6 ± 2 for the Genesis overburden (Hodgson et al., 2007), we obtain

$$\frac{\Delta v}{v} = -\frac{R}{R+1}\frac{\Delta t}{t} \approx -\frac{\Delta t}{t} \approx -\frac{d\tau}{dt}. \tag{5.3}$$

Maximum time strains in the Genesis overburden are estimated to be around +2% (Rickett et al., 2007), yielding the maximum negative velocity change of

$$\Delta v \approx -.02 \times 2800 \text{ m/s} = -56 \text{ m/s}, \tag{5.4}$$

where the estimated P-wave velocity of 2800 m/s at a 3.6 km depth was taken from the output of FWI.

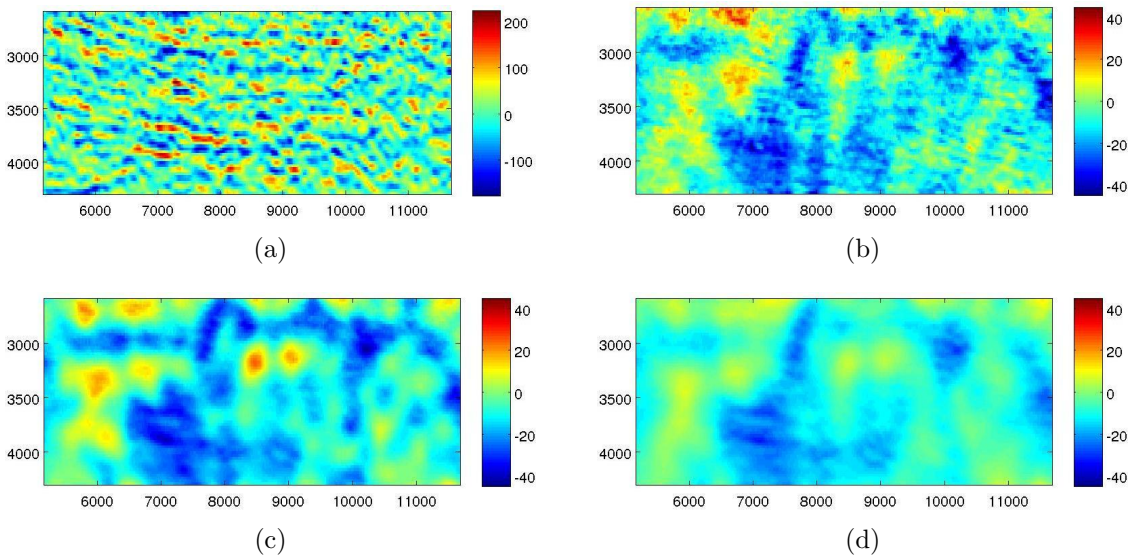


Figure 5.6: (a) Parallel difference and joint inversion results for (b) $\delta = 100$, (c) $\delta = 500$, and (d) $\delta = 1000$ in the target area. The parallel difference result is not interpretable because of the presence of artifacts. Increasing the regularization parameter δ results in gradual removal of most model differences except the negative velocity change in the overburden, peaking around the Illinoisan sands and near the top of the stacked reservoirs—see Figures 5.7(a) through 5.7(d). [**CR**] genesis/. 6pardiff,diff100,diff500,diff1000

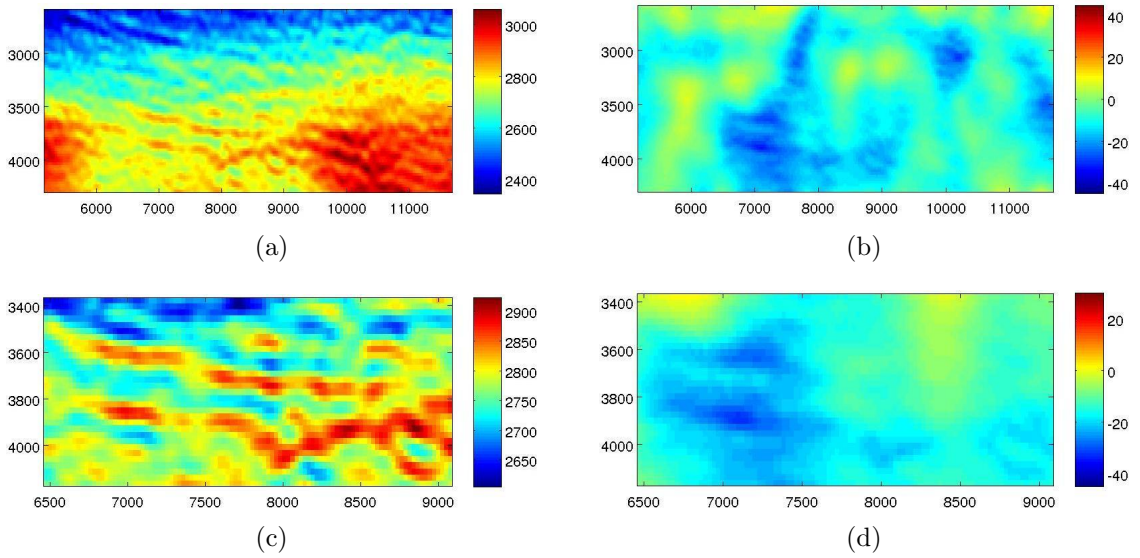


Figure 5.7: (a) Baseline target area and (b) estimated model difference for $\delta = 1000$. Close-up of (c) baseline target area and (b) estimated model difference for $\delta = 1000$. [**CR**] genesis/. baselargetarget,sdiff1000,basetarget,difftarget

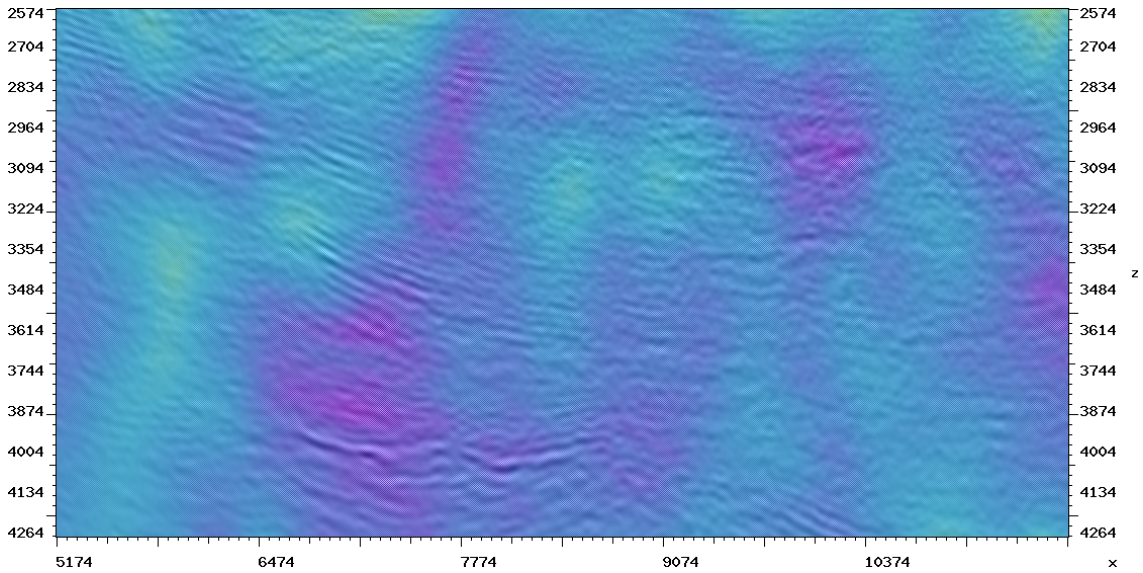


Figure 5.8: Inverted velocity difference interleaved with migrated image difference for baseline and monitor. The strongest image differences correspond to the top of Neb 1 reservoir and the Illinoisan sands above. Peaks of negative velocity change correlate with the overburden above the reservoir and the sands. [**CR**] genesis/. interleaved

GLOBAL AND 3D INVERSION

Figure 5.9 shows production well trajectories at the time of the monitor survey. Three wells, 205A18, 205A2, and 161A12STO3BP01 in blocks 205 and 161 (see Figure 5.1) have undergone failures. Note that the well paths of 205A18 and 161A12STO3BP01 have a significant curvature—see Figure 5.14(a)—that may have contributed to their failure. The 2D inversion described in the previous section was conducted along the constant crossline section shown in Figure 5.9, sufficiently close to wells 205A18, 205A2, and 205A8STO1 as to be a useful indicator of deformation associated with production from these wells.

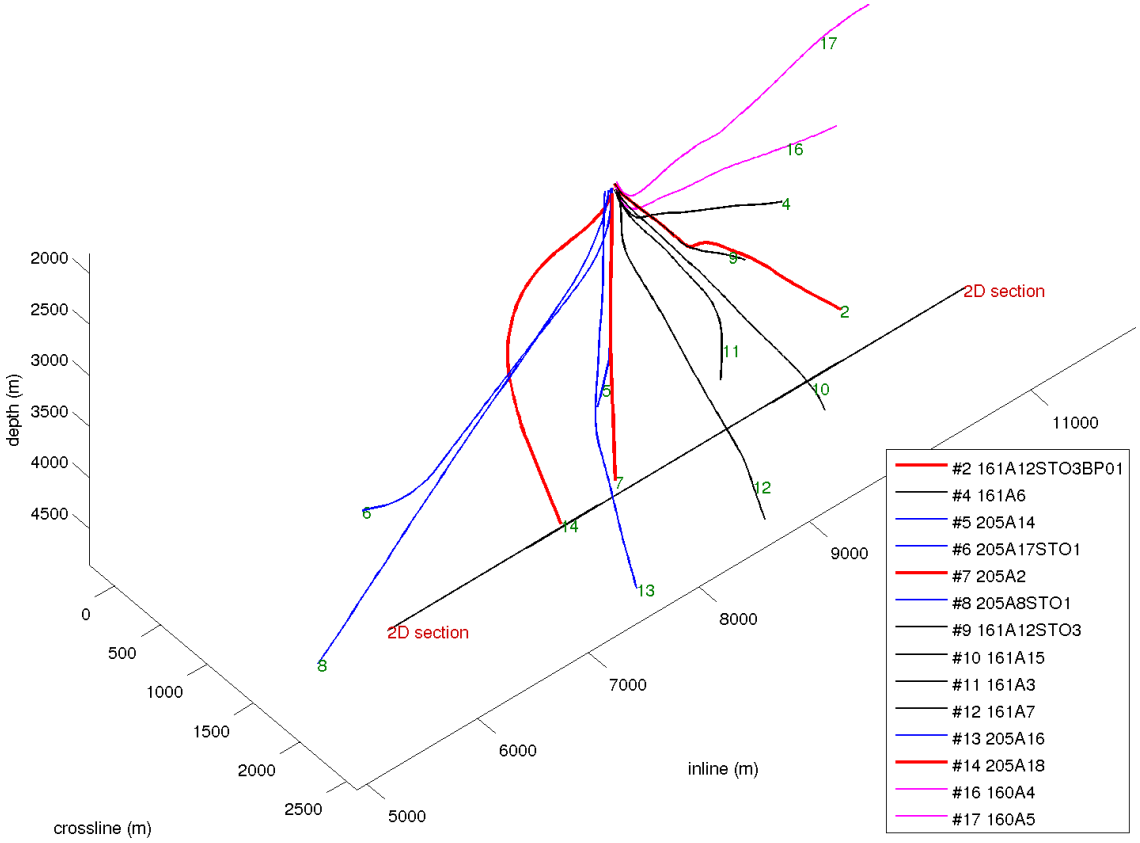


Figure 5.9: Trajectories of the Genesis production wells at the time of the monitor survey. Failed wells are shown in red, block 205 wells are plotted blue, block 160 wells are plotted magenta, and block 161 wells are plotted black. [**CR**] genesis/. cropGCwells4045

Figure 5.10 plots the local (target-oriented) inversion of velocity changes discussed in the previous section[1] against the projection of well paths. Figure 5.11 plots a *global* inversion result, i.e. with the mask $\mathbf{W} \equiv 1$ in (3.15), and using the same starting velocity as in the previous section.

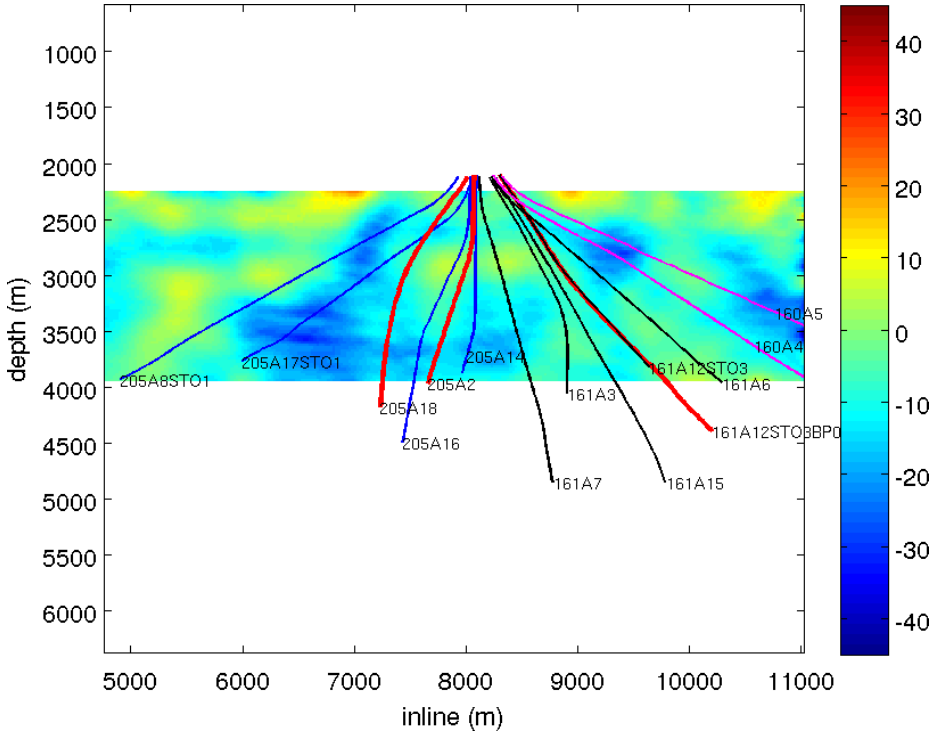


Figure 5.10: Local (target-oriented) inversion described in the previous section plotted against a projection of well paths. Wells plotted in red are known to have failed. [**CR**] genesis/. croplocal

The starting FWI model used in the targeted inversion of the previous section was constructed from the RMS velocity by inverting the Dix formula and assuming strictly vertical wave propagation. I have improved the starting model by inverting the Dix relationship along curved raypaths[2]. The resulting time-lapse inversion is shown in Figure 5.12. Note the improved match between the negative velocity anomaly and locations of the two failed wells closest to the plane of the inversion. Well 205A8STO1, which is close to the inversion plane, was previously associated with

[1] conducted in a 6×2 km window of the target reservoir

[2] A similar method is described by Cameron et al. (2007).

significant time-strains (Rickett et al., 2007), and the new inversion indicates velocity changes associated with potential deformations above the well path and higher at a depth of approximately 2.5 km—compare with Figure 4 of Rickett et al. (2007).

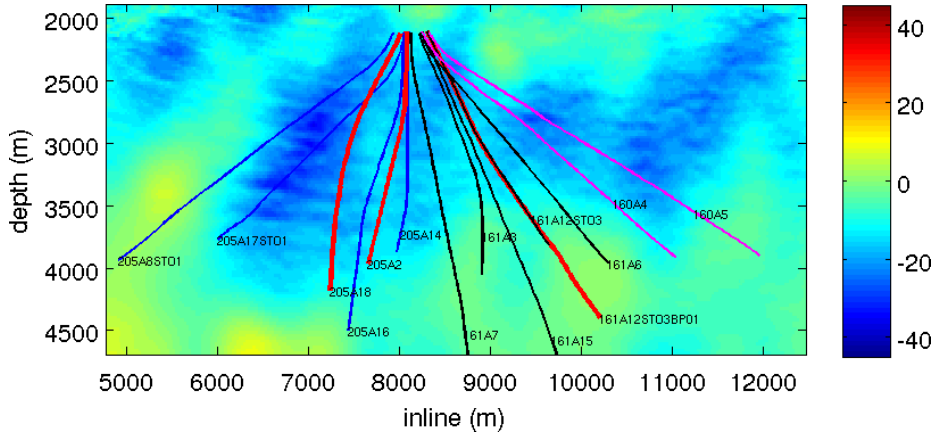


Figure 5.11: Global velocity change inversion plotted against a projection of well paths. [**CR**] genesis/. cropglobal

Wells 160A4 and 160A5 in block 160 deviate the most from the inversion plane, see Figure 5.9. However, significant volumes of production from that area indicate that the anomalies at 10000 m and 12000 m inline coordinates in Figure 5.12 may have contributions from 3D projection of overburden anomalies associated with those wells.

To truly understand the spatial correlation of overburden dilation with 3D well paths requires a 3D inversion. However, matching the quality of my 2D inversion in a 3D experiment was not computationally feasible with resources available to me. To conduct a viable 3D experiment that could still offer a valuable insight into the heterogeneity and spatial distribution of production-induced overburden changes, I conducted two experiments. First, I conducted a full 3D FWI inversion using only 60 shots. This included conducting full-waveform inversion of monitor and baseline models, followed by cross-updating and feeding the results into the simultaneous time-lapse FWI with a TV model-difference regularization (3.11,3.15). Only kinematic information was used in the inversion. The 60 shots were arranged in 12 lines of 5 shots each, with a shot spacing of 800 m, and line spacing of 300 m. The inversion

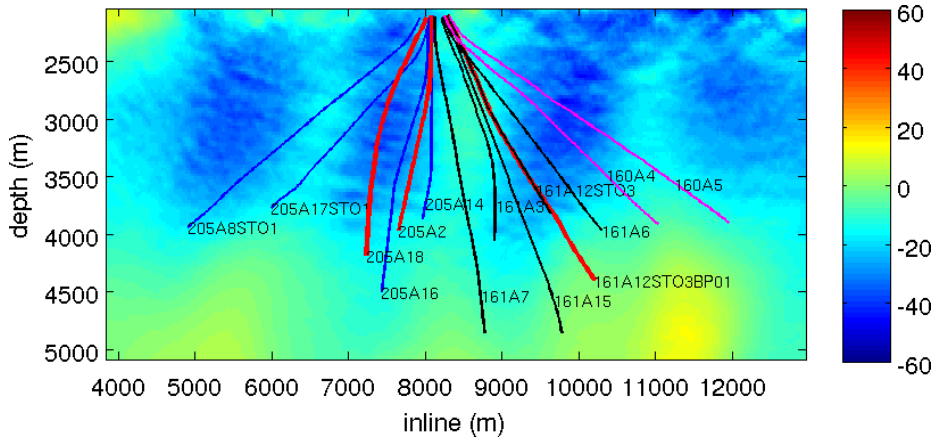


Figure 5.12: Global velocity change inversion using an improved starting model, plotted against a projection of well paths. This result appears to be in a better agreement with the trajectories of the failed wells closest to the 2D inversion plane. [**CR**] genesis/. cropalternativeglobal

was carried out as before in the frequency domain for 3–30.7 Hz, within a computational grid of 5208 m inline x 3600 m crossline x 4500 m depth. Of course, the extreme sparsity of shots along the crossline direction results in a shot aliasing and lack of continuity of inversion results along the crossline axis. Presence of reflections from around the target area[3] is a prerequisite for the time-lapse method of chapter 2. Poor resolution of reflectors around the anomaly results in the absence of a common reference for detecting phase-shifts and translating them into slowness changes using (2.24). Therefore, results of a 60-shot 3D experiment should be expected to suffer from long-wavelength oscillatory artifacts that may not be remedied by the total-variation regularization. However, partial reconstruction of model changes, even if contaminated by oscillatory artifacts, may reveal some patterns in the spatial distribution of the anomalies.

Figure 5.13 show the results of the 60-shot 3D experiment, plotting negative velocity changes in the overburden exceeding -50 m/s. While the results are contaminated by oscillatory artifacts, we can note the concentrations of negative anomalies around the well paths 205A18, 205A2, 205A14, around the curving bend of well

[3]or diving ways transmitted through it

161A12STO3BP01, and on the well path of well 160A5. However, lack of spatial continuity and spurious artifacts render these 3D results unfit for independent interpretation.

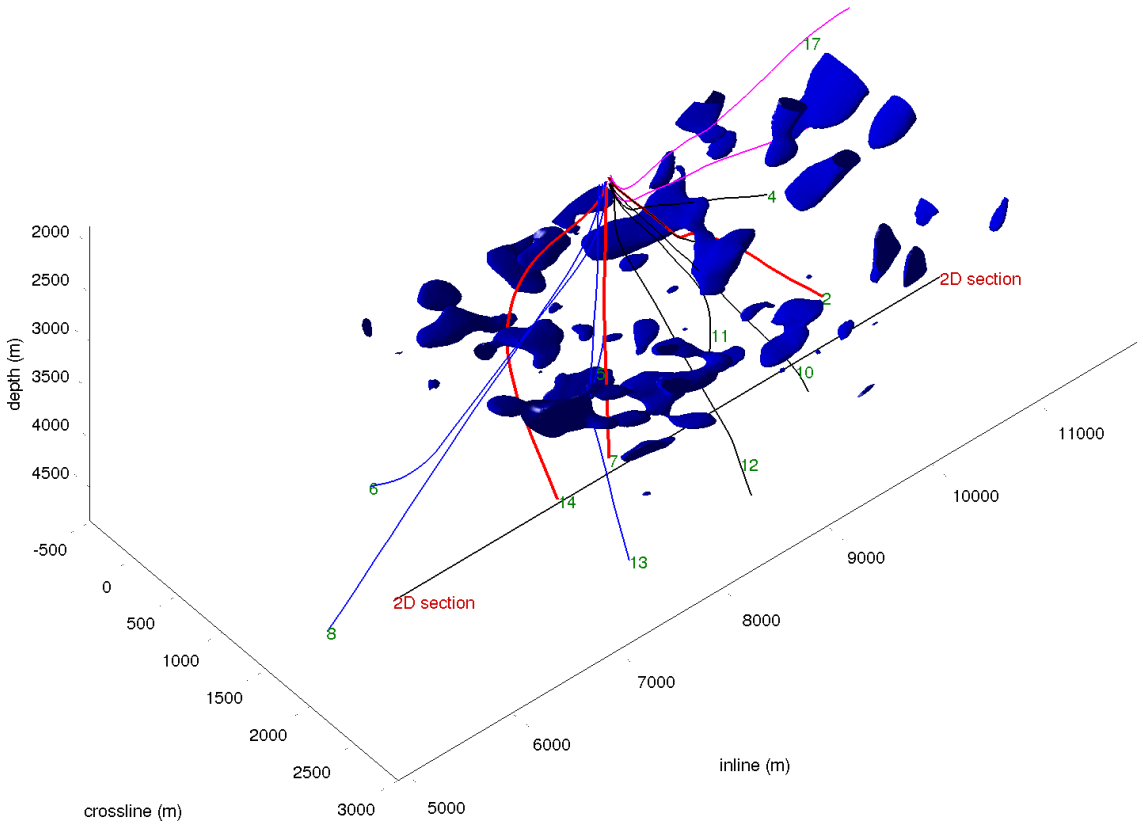
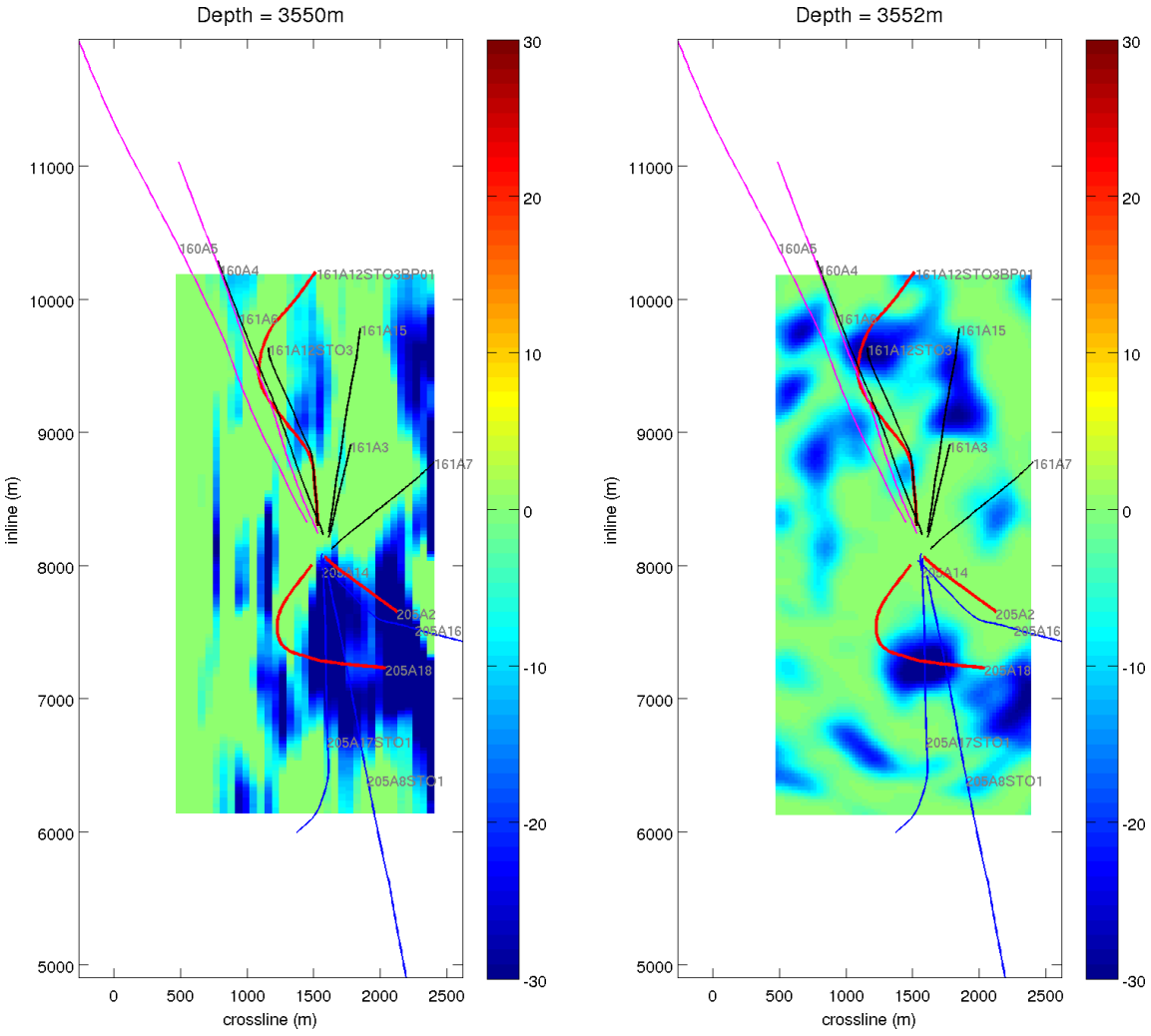


Figure 5.13: Velocity changes exceeding negative 50 m/s estimated by the 3D phase-only simultaneous FWI using only 60 shots. [**CR**] genesis/. crop3DGCwells50

To obtain more reliable spatial results without the cost of a massive 3D experiment, I conducted a "2.5D" inversion. In this experiment I have carried out 35 independent 2D inversions over parallel constant-crossline 2D sections of 4025 m inline x 4500 m depth. The crossline step is 55 m, and the entire experiment covers a crossline distance of 1870 m, see Figure 5.14(a). Each independent 2D inversion uses 120 shots spread over 3 km. The maximum offset is 4025 m as offsets in the data exceed model dimensions.

Results of the 2.5D inversion experiment are shown in Figures 5.14(a), 5.15(a),



(a) (b)

Figure 5.14: Inversion of velocity changes in excess of -30 m/s in the overburden using (a) 4200-shot 2.5D experiment (b) 60-shot 3D experiment. Approximate depth of 3550 m. [**CR**] genesis/. slice143,tslice149

5.16(a), 5.17(a). The corresponding results of the 60-shot 3D experiment are shown side-by-side in Figures 5.14(b), 5.15(b), 5.16(b), 5.17(b). One notable feature of the new results is their continuity and consistency both along the crossline axis and in depth, despite the fact that we do not impose any regularization in the crossline direction, indicating the stability of our 2.5D inversion method.

Wells 205A8STO1, 205A17STO01, 160A5 fall outside of the computational domain of the 2.5D inversion. For the remaining wells, with the exception of 161A3, the 2.5D inversion results indicate plausible and stable anomalies that can be associated with their operation—compare with Figure 4 of Rickett et al. (2007). A closer inspection of production and logging data from well 161A3 is required to understand the absence of negative velocity anomalies there.

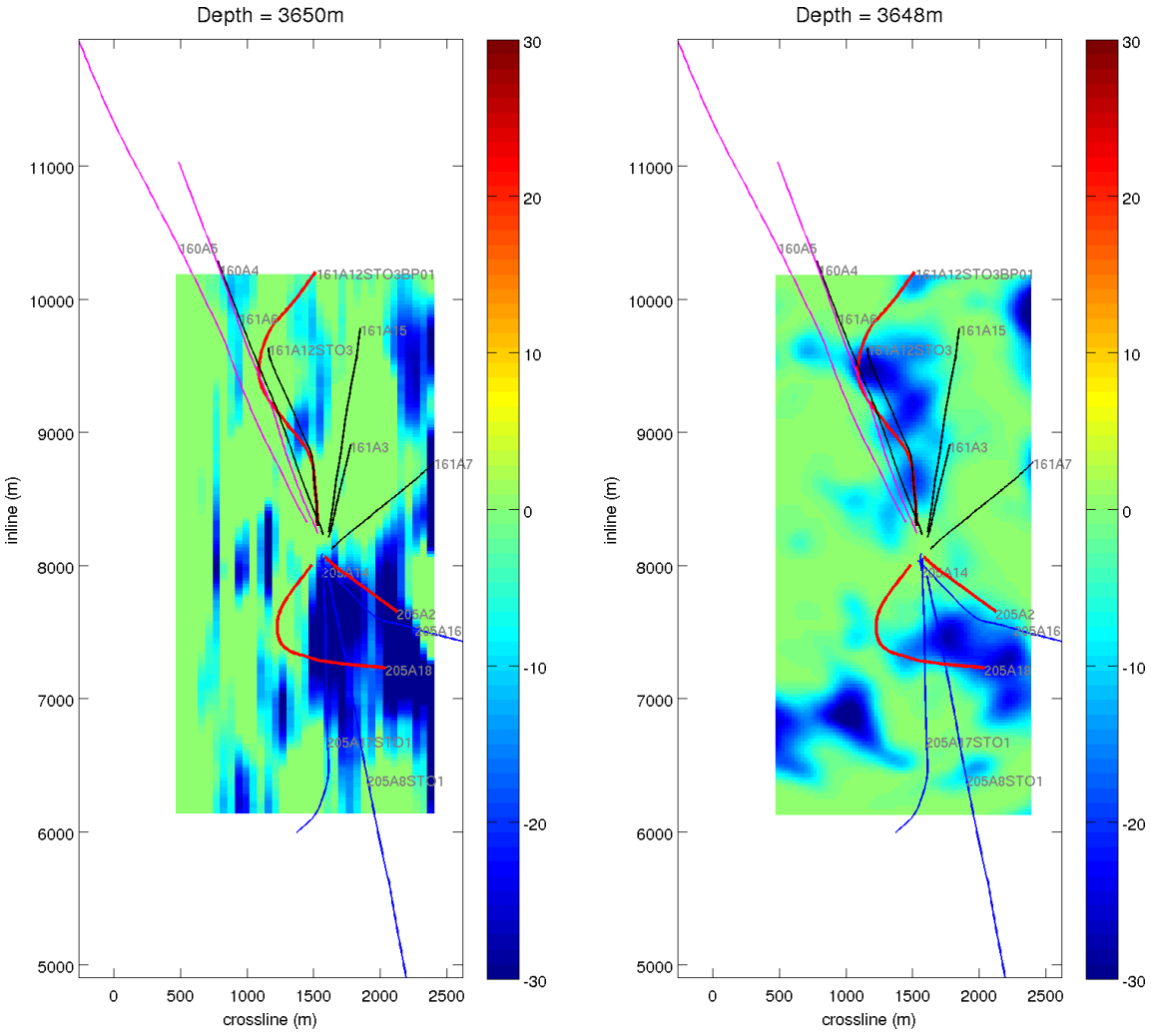
DISCUSSION AND PERSPECTIVES

Simultaneous time-lapse FWI with a total-variation difference regularization can achieve robust estimation of velocity changes in the overburden that are induced by reservoir compaction and overburden dilation. The method preserves the blocky nature of model difference while penalizing unwanted oscillations.

Application of the method to large-scale 3D time-lapse problems generally requires major computational resources. Indeed, in my experiments separate monitor and baseline full-waveform inversions were followed by two more cross-updating inversions, and solution of the simultaneous FWI (3.6,3.10) or (3.11,3.15). The latter is roughly equivalent to the cost of two FWI's, hence the total cost may reach 6 times the computational cost of a single FWI. However, in practical problems the cost can go down dramatically for a number of reasons:

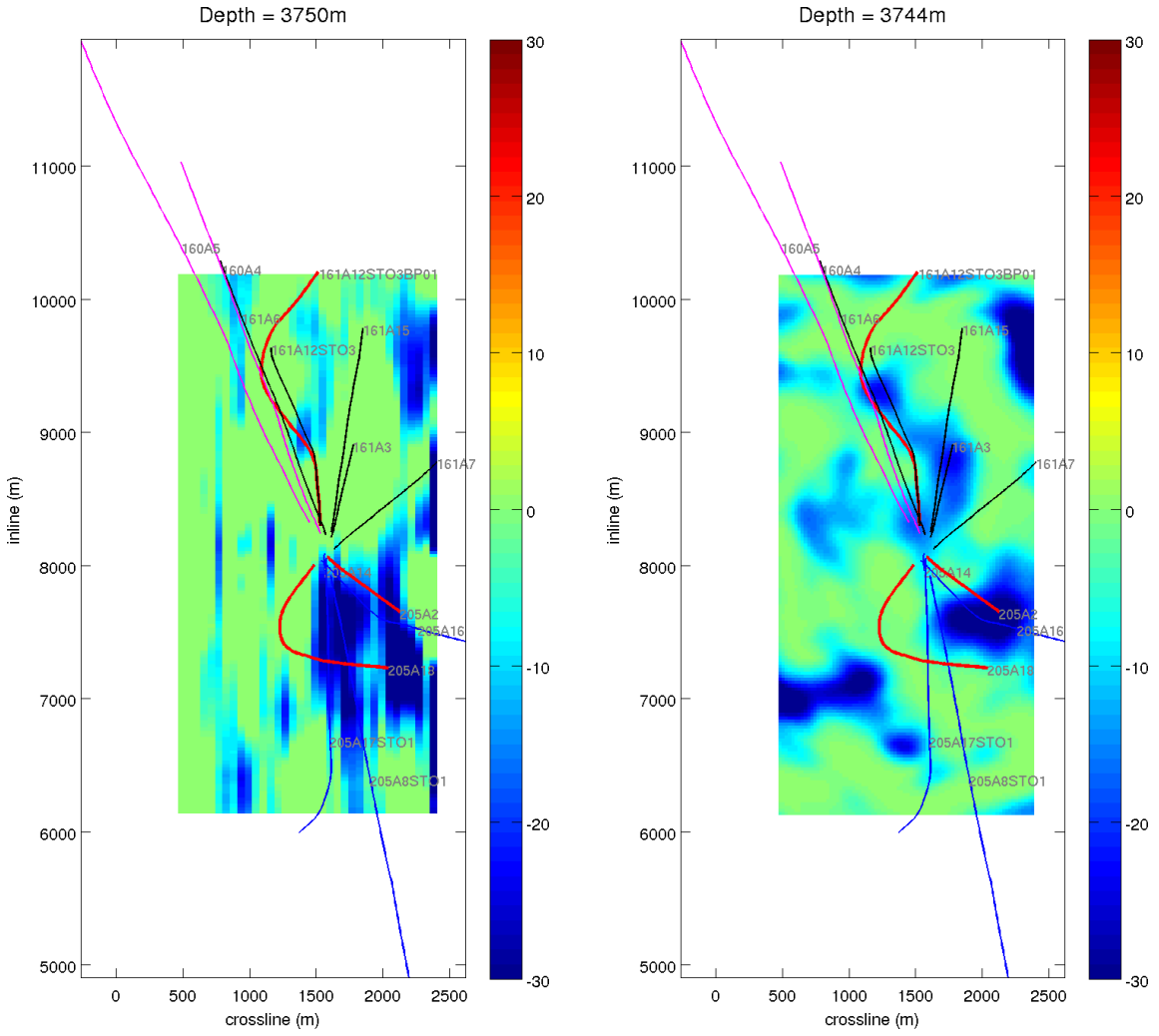
1. The results of baseline and monitor inversions may be already available, obviating two initial inversions.

2. Cross-updating may require fewer solver iterations as it starts form a good starting model.



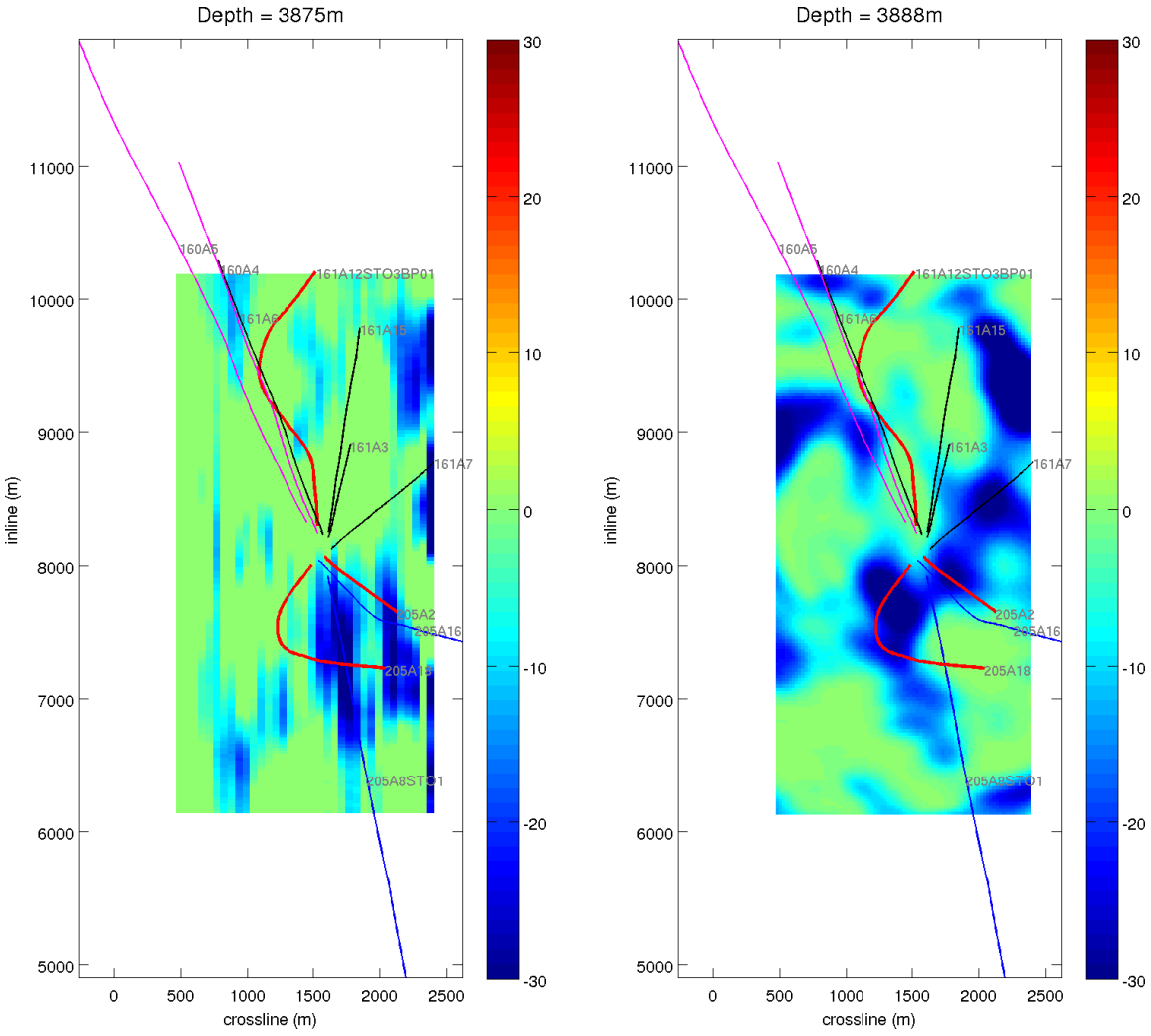
(a) (b)

Figure 5.15: Inversion of velocity changes in excess of -30 m/s in the overburden using (a) 4200-shot 2.5D experiment (b) 60-shot 3D experiment. Approximate depth of 3650 m. [**CR**] genesis/. slice147,tslice153



(a) (b)

Figure 5.16: Inversion of a velocity change in the overburden using (a) 4200-shot 2.5D experiment (b) 60-shot 3D experiment. Approximate depth of 3750 m. [**CR**] genesis/. slice151,tslice157



(a) (b)

Figure 5.17: Inversion of velocity changes in excess of -30 m/s in the overburden using (a) 4200-shot 2.5D experiment (b) 60-shot 3D experiment. Approximate depth of 3875 m. [**CR**] genesis/. slice156,tslice163

3. If baseline and monitor models are sufficiently close, we may avoid optimizing (3.6,3.10) or (3.11,3.15) by simultaneously inverting both the baseline and monitor, but instead fix one of the models and effectively use it as a prior in a regularization term (3.10) or (3.15)—see chapter 2 for a discussion of a single versus multiple model optimization.

4. 2.5D inversion can be a useful substitute for 3D time-lapse inversion, especially for simple geology (small dips and few out-of-plane reflections). In addition to providing a cost-effective alternative to the full 3D inversion, the 2.5D method provides a quality control of time-lapse data and problem parameters by expecting consistency of inversion results across multiple 2D sections.

5. In a kinematic (phase-only) inversion, we can substitute the two-way modeling of wave propagation with a one way propagator by applying one pass of downward extrapolation and a few passes of upward extrapolation from a few reference reflectors above and below the anomaly.

6. The inadequacy of a sparse shot coverage and shot aliasing are issues inherent to FWI, not the time-lapse algorithm. Time-lapse FWI requires presence of reflectors around the target anomaly, and if such reflectors have already been delineated by other methods, and reflections are generated in the synthetic data, the time-lapse method may succeed in resolving a blocky anomaly even using just a few well-positioned shots.

☞ Items 5 and 6 open up interesting possibilities for a drastic reduction of the computational cost of 3D time-lapse FWI. However, further analysis and experimentation are needed to justify and substantiate these concepts.

Part II

Geomechanical time-lapse analysis

Chapter 6

Relating surface deformation to pressure change

"We must not think of the things we could do with, but only of the things that we can't do without."

Jerome K. Jerome

Observable surface deformation associated with fluid injection and withdrawal provide a valuable insight into subsurface processes. Inversion of induced reservoir pore pressure changes from deformation measurements is a potentially powerful reservoir monitoring tool if the issues of measurement noise, uncertainty in model parametrization and numerical accuracy and stability can be resolved. In this chapter I describe a method for inverting pore pressure change from measurements of surface deformation for a poroelastostatic reservoir model. I propose numerical optimization and regularization techniques for reservoir characterization from incomplete and noisy data. Methods of this chapter are the basis of applications to synthetic and field data in the subsequent chapters of Part II.

INTRODUCTION

Surface and subsurface deformation are important and often easily measurable indicator of production-induced changes in reservoir parameters. The theory of quasistatic poroelastic deformation provides a framework for quantitative assessment of surface and subsurface deformation as a result of pore pressure changes (Rice and Cleary, 1976; Segall, 1985) and is based on Biot's theory of fluid-infiltrated porous media (Biot, 1941). Computational techniques exist (Vasco et al., 2000; Du and Olson, 2001; Hodgson et al., 2007) for numerically solving the inverse problem of inverting pore pressure from deformation based on the linear theory of poroelasticity. The main emphasis of my work is the development and analysis of regularization methods and optimization constraints for the resulting inverse problems that allow us to achieve successful characterization of the subsurface changes from noisy and limited deformation measurements.

Realistic reservoir depletion and flooding exhibit asymmetric patterns that might be indicative of a complex reservoir geometry as well as spatial heterogeneity of the reservoir permeability. In particular, differences in production-related pressure change across multiple wells can help identify reservoir compartmentalization (Zoback, 2010) and guide the drilling of subsequent production wells or affect the choice of production-enhancement techniques. In the general case of an arbitrary linear poroelastic medium, change in the pore pressure is intricately interconnected with the change in the stress field, and accurate modeling of the effects of changing pore pressure requires solving a system of governing equations in a half-space (Wang, 2000; Segall, 2010). Solving a boundary-value problem for such a system of equations governing both stress evolution and fluid flow is challenging—as much due to uncertainty of the subsurface model parameters as due to the sheer analytical complexity of these coupled equations. The importance of studying the fully-coupled poroelastic models cannot be overestimated as the coupled model is often the key to explaining counterintuitive behaviour of some real-world poroelastic models. In this work I expand on previous studies of the effects of pore pressure change on the strain and stress fields (Segall, 1992; Segall et al., 1994) by adopting an intermediate "semi-coupled" approach between the fully

coupled simulation and uncoupled analytical solutions—see also (Maharramov, 2012; Maharramov and Zoback, 2014, 2015). First, I assume that a pore pressure drop is known within the reservoir and is an arbitrary function of the reservoir coordinates—i.e., I assume partial *fluid-to-solid* coupling—and use the analytical expression for the elastostatic Green's tensor due to a concentrated center of dilatation in half-space (Segall, 2010; Mindlin, 1936) to *numerically compute* the displacement due to the pore pressure change. Next, I use the obtained numerical operator in an inversion problem, fitting a pore pressure drop to a known displacement and subsidence profile. Finally, I demonstrate an extension of the method for layered and weakly laterally heterogeneous.

Numerical modeling of deformation due to changing pore pressure is one of the key problems of reservoir geomechanics (Zoback, 2010). Reservoir depletion has been demonstrated to have appreciable effect on stress both inside and outside of the producing reservoir (Zoback, 2010),(Geertsma, 1973),(Segall, 1992),(Segall et al., 1994),(Zoback and Zinke, 2002),(Segall and Fitzgerald, 1998). While some simplifying assumptions with regard to reservoir geometry (e.g., an infinitely wide and thin horizontal layer) yield a simple law for the horizontal stress change with the pressure decline *within the reservoir* (Zoback, 2010),(Segall and Fitzgerald, 1998)), estimating induced stress-field changes around the reservoir requires more elaborate models of reservoir depletion (Geertsma, 1973),(Segall, 1992). Simple disk-shaped and radially-symmetric reservoir shapes proved adequate for many simple situations, but compartmentalization and heterogeneous permeability inside realistic reservoirs point to a departure of the pore pressure decline from simple axisymmetric patterns (Zoback, 2010). Compartmentalization with impermeable barriers still allows for the application of a radially-symmetric pore pressure change law to individual reservoir compartments. However, a fully heterogeneous pore pressure drop can account for the effects of partial permeability or complex spatial geometry. Advanced optimization and regularization techniques for inverting highly heterogeneous pore pressure changes that I develop in Part II can provide efficient non-invasive methods of reservoir characterization.

POROELASTIC DEFORMATION

I begin by formulating a closed system of four equations that describes a homogeneous quasi-static linear poroelastic medium (Segall, 2010):

$$\mu\nabla^2 u_i + \frac{\mu}{1-2\nu}\frac{\partial^2 u_j}{\partial x_i \partial x_j} = \alpha\frac{\partial p}{\partial x_i} - f_i = 0,\ i = 1,2,3 \tag{6.1}$$

and

$$S_\alpha \frac{\partial p}{\partial t} - \frac{\kappa}{\eta}\nabla^2 p = -\alpha\frac{\partial}{\partial t}\left(\nabla \cdot \mathbf{u}\right)). \tag{6.2}$$

In the above equations $u_i, i = 1,2,3$ is a spatially-distributed displacement vector field, p is the pore pressure change, f_i is a differential body-force distribution, μ, ν, α, κ, and η are the shear modulus, Poisson's ratio, Biot coefficient, permeability, and fluid viscosity, respectively. The storage coefficient S_α is a known function of medium parameters (Segall, 2010):

$$S_\alpha = \frac{3\alpha(1-2\nu)(1-\alpha B)}{2\mu B(1+\nu)}, \tag{6.3}$$

where B in equation 6.3 is Skempton's pore pressure coefficient—the ratio of the induced pore pressure change to the mean normal stress for undrained loading conditions. Note that the subsurface stress is absent from equations 6.1 and 6.2 but can be computed using constitutive laws after these equations have been solved. Also note that the displacement and pore pressure in these equations are relative to a reference state, not the total values. The equilibrium equation 6.1 and flow equation 6.2 are fully coupled and are obtained from combining the constitutive laws for a poroelastic medium with quasi-static field equations. The equations are "quasi-static" in the sense that the stress field is assumed to be in a state of static equilibrium even though changes of the pore pressure in time induce changes of the stress field. We can think of this as a "slow-change" asymptotic approximation, both in time and space.

The most mathematically rigorous way of computing the displacement field and associated pore pressure change is to solve a boundary-value problem for equations 6.1 and 6.2 with known data (e.g., known pressure evolution within existing wells,

measured earth displacements or estimated stresses) used as boundary or initial conditions. However, even in the simplest case of a homogeneous medium, analytical solutions of such boundary-value problems are very challenging. Uncoupling equations 6.1 and 6.2, where permissible, could result in more tractable problems, both analytically and numerically. For example, assuming a known pore pressure change, we can solve equation 6.1 for the displacement field u_i, using $\alpha \partial p / \partial x_i$ in the right-hand side as a "body force" distribution (Geertsma, 1973; Segall, 1992).

I use the elastostatic Green's tensor $g_i^k(x, y, z, \xi, \eta, \zeta)$ for the pure elastic equilibrium equation in the left-hand side of equation 6.1 to compute the displacement u_i as

$$\begin{aligned} u_i = &- \alpha \int_V g_i^k \frac{\partial p}{\partial x_k} = \\ &= \alpha \int_V \frac{\partial g_i^k(x, y, z, \xi, \eta, \zeta)}{\partial x_k} p(\xi, \eta, \zeta) d\xi d\eta d\zeta, \end{aligned} \tag{6.4}$$

assuming $f_i = 0$ (including body forces is trivial). The elastostatic tensor has the meaning of the displacement along axis i at point (x, y, z) due to a concentrated force along axis k at point (ξ, η, ζ) (Wang, 2000; Segall, 2010). From equation 6.4 we can see that the divergence of the elastostatic tensor has the meaning of deformation due to a concentrated dilatational force.

In order to apply equation 6.4 to practical reservoir models and computation of surface displacements, the corresponding Green's function should be constructed for a half-space with a free-surface boundary condition imposed on its bounding plane (Segall, 2010). In homogeneous medium experiments, we use the analytical expression for the Green's function obtained by Mindlin (Mindlin, 1936)—see appendix D.

FORWARD AND INVERSE PROBLEMS

We can use operator 6.4 for forward-modeling the displacement field from a specified pressure change. Note that equation 6.4 describes a non-stationary convolutional integral operator for a homogeneous medium. The convolution is non-stationary due

to the presence of $z+\zeta$ in the elastostatic Green's tensor. Integration along the horizontal axes can be accelerated by applying the operator in the wavenumber domain. However, integration along the vertical axis should still be carried out separately for different values of z, hence the integration kernel is effectively four-dimensional. Assuming the reservoir to be thin in comparison with its lateral extents, widely true in practice, we can replace the vertical integral with a mean value of the integrand times the reservoir thickness:

$$u_i(x,y,z) = \alpha \int_V h(\xi,\eta) \frac{\partial g_i^k(x,y,z,\xi,\eta,S(\xi,\eta))}{\partial x_k} \times p(\xi,\eta,S(\xi,\eta)) d\xi d\eta, \tag{6.5}$$

where $S(\xi,\eta)$ is the middle surface of the reservoir and $h(\xi,\eta)$ is the reservoir depth. For a non-flat reservoir, g_i^k effectively depends not only on differences $x-\xi$ and $y-\eta$ but on integration variables as well.

By modeling subsidence using equation 6.5, we are able to fully account for the asymmetric nature of the depletion pattern by using the most general form of Green's tensor for a homogeneous half-space. In that respect, our approach represents an advancement of the purely analytical techniques for axisymmetric reservoirs presented by Geertsma (1973) and Segall et al. (1994), and is similar to the method used by Hodgson et al. (2007).

Denoting the operator in the right-hand side of equation 6.5 as $\mathbf{A}$, the problem of recovering the pore pressure change from specified displacements can be cast as a least-squares minimization problem (Aster et al., 2011):

$$\|\mathbf{A}p - \mathbf{u}\|_{L_2}^2 \;\rightarrow\; \min. \tag{6.6}$$

Operator 6.5 uses Mindlin's analytical expressions for elastostatic Green's tensor which assumes homogeneity of the medium. However, by using a gradient-based optimization solver that only requires the application of the modeling operator $\mathbf{A}$ and its adjoint $\mathbf{A}^*$ we obviate the need to use an explicit analytical representation for

the operator, and can substitute it with a more computationally intensive deformation modeling operator for a heterogeneous medium, as described in the next section. In practical applications, problem 6.6 is often mixed-determined (Aster et al., 2011), meaning that the deformation data $\mathbf{u}$ cannot be fit exactly due to measurement noise and unaccounted physical effects, and the data are insufficient to uniquely resolve the pressure change p over the entire computational domain—i.e., problem 6.6 does not have a unique solution. I address the non-uniqueness using Tikhonov regularization (Tikhonov and Arsenin, 1977; Aster et al., 2011) by adding a term penalizing spatial pressure oscillations:

$$\|\mathbf{A}p - \mathbf{u}\|_{L_2}^2 \; + \; \epsilon\|\Delta p\|_{L_2}^2 \; \rightarrow \; \min, \tag{6.7}$$

where Δ is the Laplace operator, and ϵ is an empirically chosen regularization parameter. We assume that the pressure change is always non-negative and, to avoid hydraulic fracturing, does not exceed the minimum in-situ stress (Zoback, 2010). This results in the following inequality constraints on the pressure change:

$$0 \; \leq \; p \; \leq \; p_{\max}. \tag{6.8}$$

I solve problem 6.7,6.8 using an interior point method (Nocedal and Wright, 2006).

In Chapter 8, I study the problem of inverting pressure contrasts that are due impenetrable barriers or other heterogeneities in the permeability. For such problems, where recovering large contrasts is more important than recovering small oscillatory features, I solve the regularized problem

$$\|\mathbf{A}p - \mathbf{u}\|_{L_2}^2 \; + \; \epsilon\|\,|\nabla p|\,\|_{L_1} \; \rightarrow \; \min, \tag{6.9}$$

where the regularization term in the right-hand side of (6.9) is the L_1 norm of the gradient norm, or the *total-variation* seminorm of pressure change. Such regularization penalizes oscillations while promoting the "blockiness" of the inverted function—compare with Chapter 3, equation (3.15). Note that for pressure inversion we still need to minimize (6.9) subject to the constraint (6.8), resulting a rather challenging problem of constrained optimization with a non-smooth objective function. In

appendix C, I propose a computationally efficient numerical optimization method for solving problems of constrained total-variation minimization that I apply later in Chapter 8.

HETEROGENEOUS MODELS

For mild heterogeneity, when the medium parameters slowly change in space, asymptotic methods similar to Maslov (1990) can be used to account for the first-order effects of the heterogeneity. However, such an approach is inherently limited to moderate heterogeneity. Lateral heterogenity of medium parameters is subject to considerable uncertainty, while layered models are of particularly high importance as a dominant stratigraphy. We therefore focus on modeling displacements for a vertically heterogeneous and horizontally slowly-varying medium. Rather than trying to solve a heterogeneous analogue of system 6.1,6.2, we will assume that one or all components of the displacement at a fixed depth $z = z_{\max}$ immediately above the reservoir are known *a priori*. For example, we may use operator 6.5 to model displacements near the reservoir where the effect of the spatial heterogeneity of elastic parameters is limited. With displacements at $z = z_{\max}$ and free-surface boundary conditions at $z = 0$, the problem of modeling subsurface displacements is reduced to solving a boundary-value problem for the elastostatic system:

$$\begin{aligned} \mu\left(\frac{\partial u_i}{\partial x_j}+\frac{\partial u_j}{\partial x_i}\right)+\frac{2\mu\nu}{1-2\nu}\frac{\partial u_k}{\partial x_k}\delta_{ij} =&\sigma_{ij}, \\ \frac{\partial \sigma_{ij}}{\partial x_j} =&0, \end{aligned} \tag{6.10}$$

where indices run from 1 to 3, σ_{ij} denote the stress tensor components, summation is carried out on repeating indices and the body-force distribution is zero. The boundary conditions used with system 6.10 are the following:

$$\begin{aligned} \left(\frac{\partial u_i}{\partial x_3}+\frac{\partial u_3}{\partial x_i}\right)+\frac{2\nu}{1-2\nu}\frac{\partial u_k}{\partial x_k}\delta_{i3}|_{z=z_0} =&0, \\ u_i(x,y,z_{\mathrm{R}}) =&s_i(x,y), \end{aligned} \tag{6.11}$$

where $s_i(x, y),\ i = 1, 2, 3$ describe a known displacement field at a fixed depth. Although system 6.10 is comprised of purely elastostatic equations, it allows us to model fluid-to-solid coupling via the boundary condition at $z = z_{\mathrm{R}}$ that can be approximately computed using operator 6.5. For a laterally-homogeneous medium—or under the assumption of slow lateral varying coefficients that commute with differentiation to the first order (Maslov, 1976)—equations 6.10 can be Fourier-transformed in x_1, x_2, and the resulting system discretized in depth:

$$
\begin{aligned}
\frac{v_1(z+\Delta z) - v_1(z-\Delta z)}{2\Delta z} &= v_4(z),\\
\frac{v_2(z+\Delta z) - v_2(z-\Delta z)}{2\Delta z} &= v_5(z),\\
\frac{v_3(z+\Delta z) - v_3(z-\Delta z)}{2\Delta z} &= v_6(z),\\
\frac{\mu(z+\Delta z)v_4(z+\Delta z) - \mu(z-\Delta z)v_4(z-\Delta z)}{2\Delta z} &=\\
-(k_x^2 + k_y^2 + \frac{k_x^2}{1-2\nu(z)})\mu(z)v_1(z) &-\\
\frac{k_x k_y}{1-2\nu(z)}\mu(z)v_2(z) + \frac{ik_x}{1-2\nu(z)}\mu(z)v_6(z),&\\
\\
\frac{\mu(z+\Delta z)v_5(z+\Delta z) - \mu(z-\Delta z)v_5(z-\Delta z)}{2\Delta z} &=\\
-\left(k_x^2 + k_y^2 + \frac{k_y^2}{1-2\nu(z)}\right)\mu(z)v_2(z) &-\\
\frac{k_x k_y}{1-2\nu(z)}\mu(z)v_1(z) + \frac{ik_y}{1-2\nu(z)}\mu(z)v_6(z),&\\
\frac{\mu(z+\Delta z)v_6(z+\Delta z) - \mu(z-\Delta z)v_6(z-\Delta z)}{2\Delta z} &=\\
\frac{1}{1+\frac{1}{1-2\nu(z)}}\left[-(k_x^2+k_y^2)v_3(z) + \frac{ik_x}{1-2\nu(z)}v_4(z) + \right.&\\
\left.\frac{ik_y}{1-2\nu(z)}v_5(z)\right],&
\end{aligned}
\tag{6.12, 6.13}
$$

where k_x, k_y are the horizontal wavenumbers and Δz is a depth step, $v_{1,2,3}$ are Fourier-transforms of the three displacement components $u_{1,2,3}$ and $v_{4,5,6}$ are the z-derivatives of $v_{1,2,3}$. At the boundaries $z = 0$ and $z = z_{\mathrm{R}}$, the central differences should be

replaced with backward and forward differences (Iserles, 2008), and the boundary conditions 6.11 Fourier-transformed in a similar manner. In combination with the Fourier-transformed boundary conditions the above system is reduced to independent $6N_z \times 6N_z$ linear systems for finding $v_i(\Delta z_j), i = 1, \ldots, 6, j = 1, \ldots, N_z$ for each wavenumber pair k_x, k_y, where N_z is the number of depth steps. Solution of the above system is efficiently parallelized, with individual $6N_z \times 6N_z$ sparse systems solved independently. Furthermore, each of the systems is banded with the bandwidth of 13 elements and therefore can be solved in a linear time and memory $O(N_z)$ (Trefethen and Bau III, 1997). Although depth-varying models are common in geomechanical applications, and the diffusive nature of induced deformation favors slowly-varying models, practical applications exist where a strong lateral heterogeneity should be taken into account. The widely accepted approach to tackling such problems consists in applying the finite elements method (Iserles, 2008) to the coupled poroelastic system (Kosloff et al., 1980). While finite elements can handle arbitrary spatial heterogeneity, the main disadvantage of this approach is the computational cost associated with solving a potentially very large system of linear equations with a very sparse but generally unstructured matrix.

Chapter 7

Reservoir monitoring by inverting pore pressure changes from surface deformation

"What covers you discovers you."

Miguel de Cervantes Saavedra

Temperature and pressure changes associated with cyclic steam simulation (CSS) used in heavy oil production from sands can cause significant subsurface deformation. Inversion of induced reservoir pore pressure changes from deformation measurements, when feasible, may provides a powerful non-invasive reservoir monitoring tool. In this chapter, I apply the technique developed in Chapter 6 to estimating pore pressure change from tilt measurements at a heavy oil reservoir undergoing CSS. I achieve a stable inversion of the reservoir pore pressure change from sparse and noisy surface tilt measurements using constrained regularized optimization. The results provide an insight into the heterogeneity of reservoir stimulation and can potentially help with optimizing well locations and stimulation protocols.

CYCLIC STEAM STIMULATION

In this work, we apply the methodology and computational framework of Chapter 6, previously applied by Maharramov (2012) to pore pressure inversion in a conventional gas reservoir, to estimating pore pressure changes from surface tilt measurements over a heavy oil reservoir undergoing cyclic steam simulation described by Walters and Zoback (2013). Cyclic steam stimulation (CSS) is used to reduce the viscosity of heavy oils so that the oil will flow to production wells. This is achieved by injecting high-temperature steam into the formation during an injection period that typically lasts a few weeks or months. This is followed by a "soaking" period during which viscosity of the oil is dramatically reduced due to heat from the steam (Hinkle and Batzle, 2006). Usually the same wells are used for injection and production. Because oil recovery is dependent on effective injection, it is very important to understand and monitor the steam front. This should allow for the appropriate determination of steam paths and the effects of reservoir heterogeneity on steam injection and production. This work is part of an interdisciplinary study of a heavy oil reservoir undergoing cyclic steam stimulation (Walters and Zoback, 2013). Steam was injected in two cycles, with the first injection in Cycle 1 running from November 2007 through January 2008. Surface tilt measurements were collected from 30 surface tilt stations during Cycle 1 only. No tilt measurements were collected during the following cycle, and this work focuses only on estimating the pore pressure change in the reservoir during Cycle 1.

Cycle 1 steam injection ran in two overlapping phases: Phase 1 ran from the beginning of the injection through mid-December, and Phase 2 overlapped with Phase 1 and ran through the beginning of January. During Phase 1, steam was injected in the western part of the reservoir, followed by injection in the eastern part in Phase 2. There are 33 production/injection wells as shown in Figure 7.1 (Walters and Zoback, 2013). The positions of 25 (out of a total of 30) surface tilt stations that were deemed to provide usable data are shown in Figure 7.2.

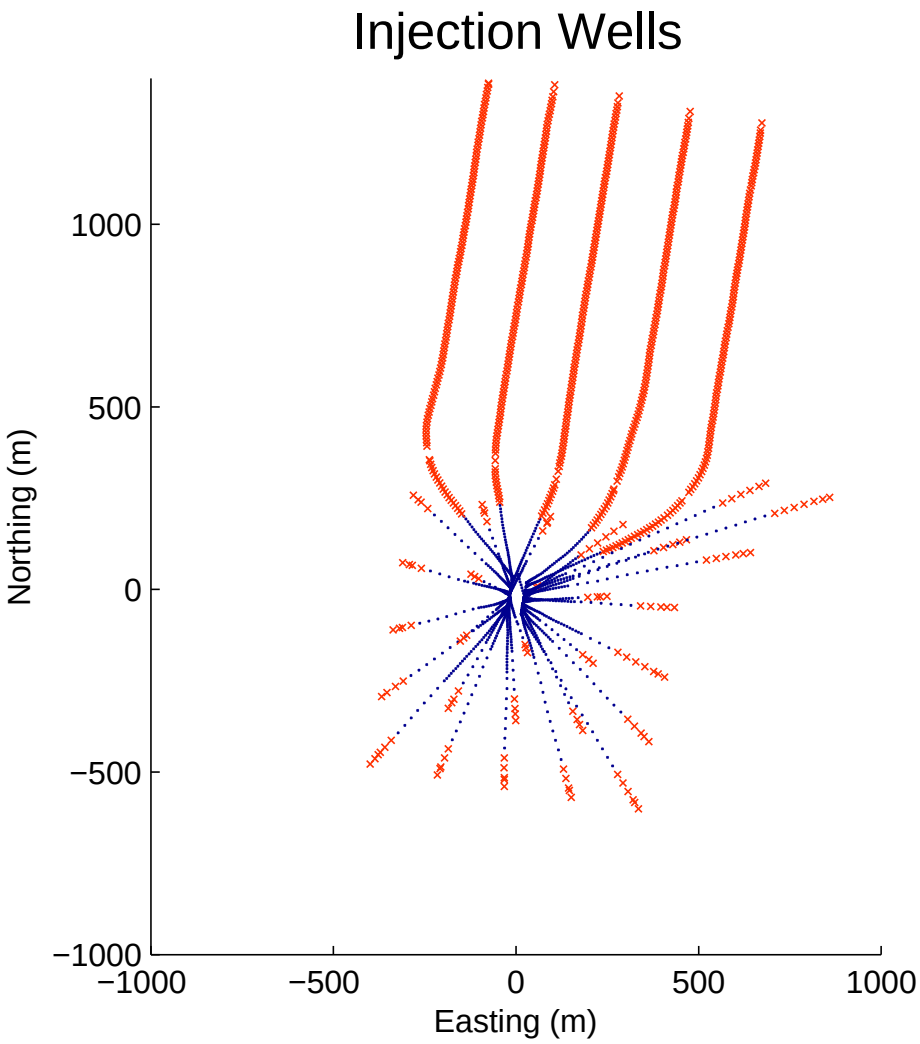


Figure 7.1: Injection well trajectories. The portions of well trajectories that are within the reservoir are shown in red. [**CR**] tilts/. welltrajectories

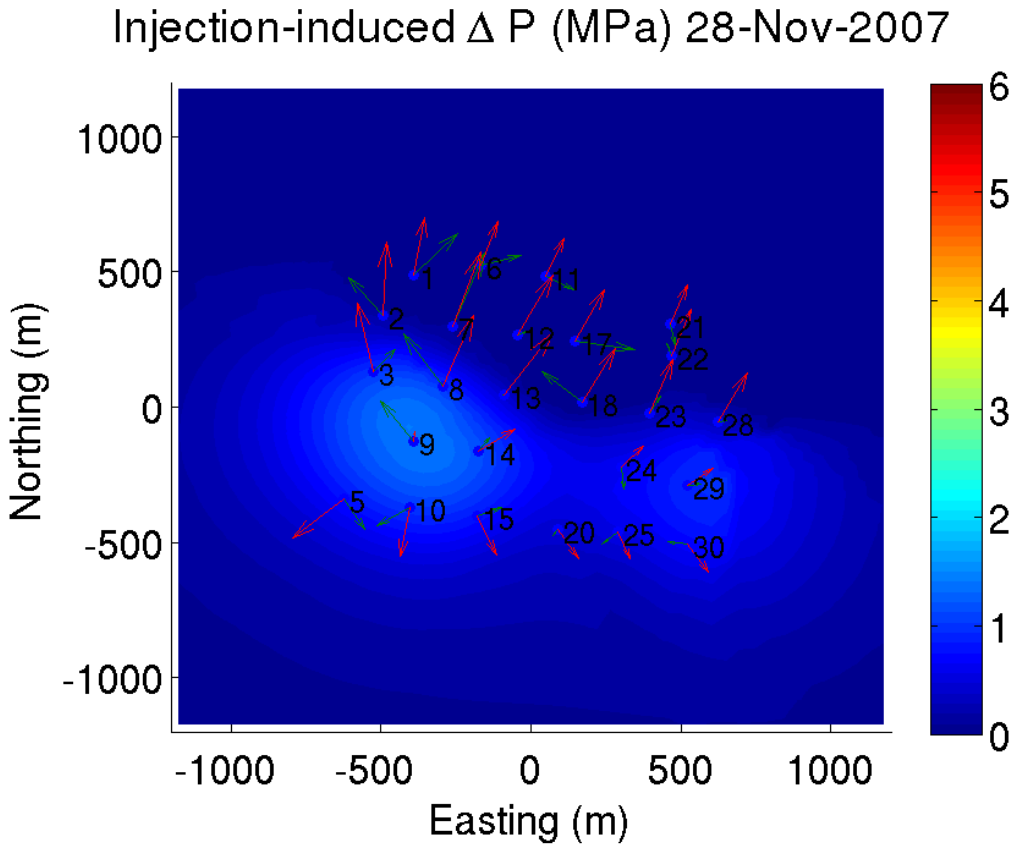


Figure 7.2: Inverted induced pore pressure change (color scale) and differential tilt measurements (arrows) for $\epsilon = 10^{-3}$ at the beginning of Phase 1 of Cycle 1, after 2 weeks of injection. The observed differential tilts are shown in green, modeled tilts are in red. [**CR**] tilts/. TdP14

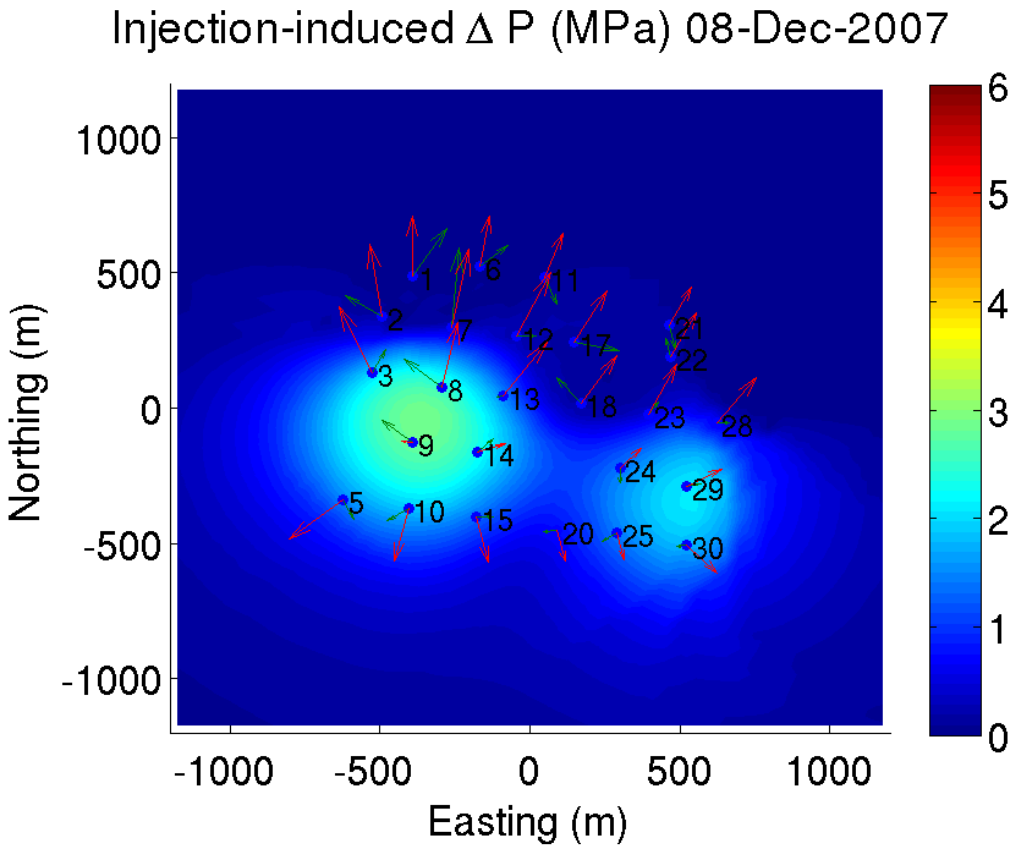


Figure 7.3: Inverted induced pore pressure change (color scale) and differential tilt measurements (arrows) for $\epsilon = 10^{-3}$ at the end of Phase 1 of Cycle 1, after 24 days of injection. [**CR**] tilts/. TdP24

INVERSION OF PRESSURE FROM TILT DATA

I used the following values for the poroelastic medium and reservoir parameters in our tests:

Name	Value	Units
Shear modulus μ	1.5	Gpa
Poisson's ratio ν	.25	–
Undrained Poisson's ratio ν_u	.45	–
Skempton's coefficient B	.9	–
Average TVD of reservoir	455	m
Average reservoir thickness	50	m

I solved the regularized optimization problem 6.7, 6.8 over a 2 km by 2 km computational domain with a 40 m spacing. I used the value of $p_{\max} = 7$ MPa in the upper constraint 6.8 based on estimated vertical stress at the reservoir depth, assuming the minimal stress to be vertical. Continuous tilt measurements from 25 functional tilt meters (see Figure 7.2) were available over a 60-day period. These measurements were decimated to 60 daily measurements from each of the 25 tilt meters and used in independent inversions of daily cumulative pressure changes.

Although there is a considerable uncertainty with regard to the magnitudes of the medium parameters (e.g., lateral heterogeneity of Skempton's coefficients), this uncertainty does not affect the qualitative evolution of the induced pressure change. Figures 7.2, 7.3, 7.4, 7.5, 7.6, and 7.7 show inversion results at various stages of Cycle 1 injection, with the regularization parameter $\epsilon = 10^{-3}$. A maximum cumulative pressure change of 6.1 MPa was achieved on January 6, 2008. Note that the migration of the pressure peak eastward with the progress of injection from Phase 1 to Phase 2 is consistent with the geometry of the injection wells in Figure 7.1. No value of pressure change was prescribed along the boundaries of the computational domain, and the inversion results indicate a significant pressure increase in the northeastern part of the modeling domain, consistent with the fact that a considerable fraction of the injection

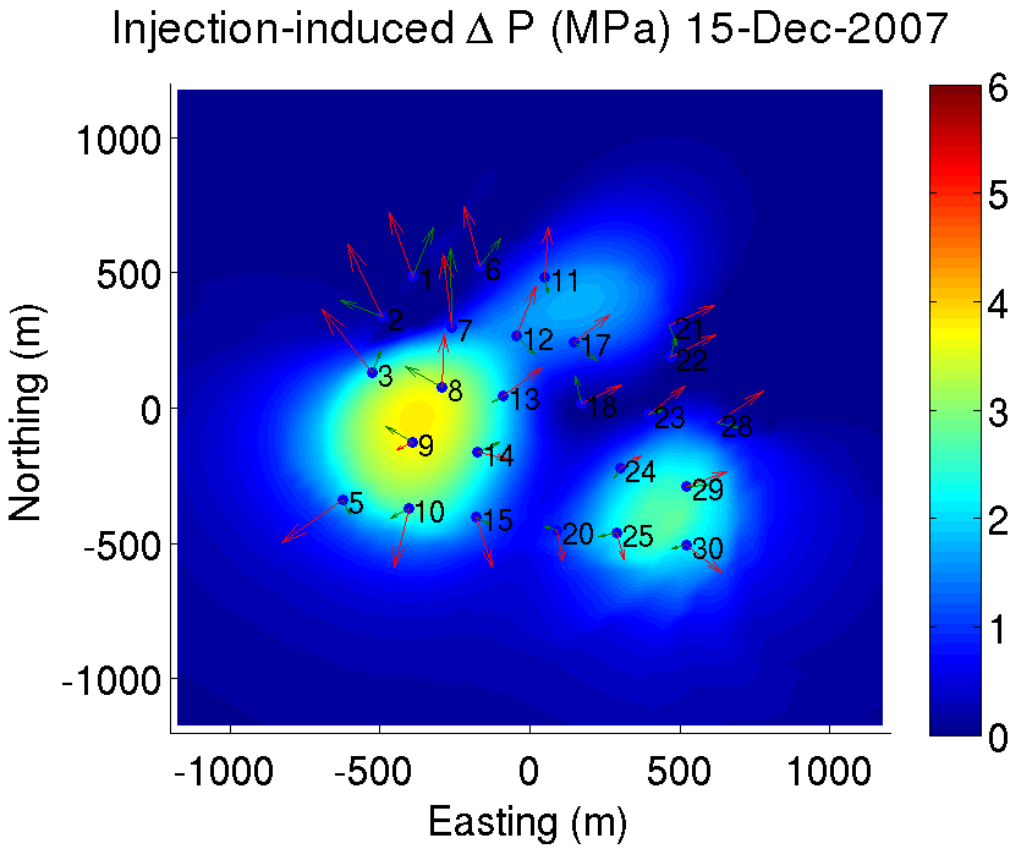


Figure 7.4: Inverted induced pore pressure change (color scale) and differential tilt measurements (arrows) for $\epsilon = 10^{-3}$ at the beginning of Phase 2 of Cycle 1, after 31 days of injection. [**CR**] tilts/. TdP31

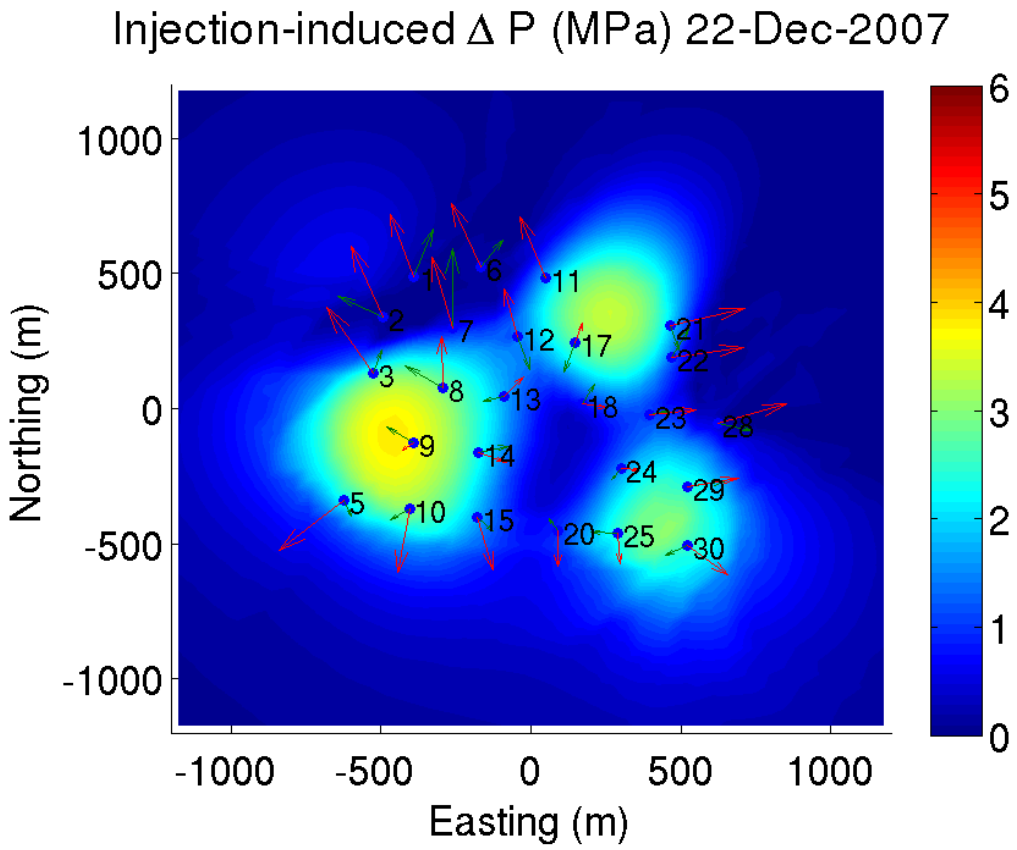


Figure 7.5: Inverted induced pore pressure change (color scale) and differential tilt measurements (arrows) for $\epsilon = 10^{-3}$ after 38 days of injection. [**CR**] tilts/. TdP38

wells are located in that area. Also note that no temporal regularization was used in the inversion, i.e., separate inversions were performed for each set of tilt observations. The fact that the resulting pore pressure estimates are continuous in time indicate consistency of our input data and stability of the inversion. I conducted optimization

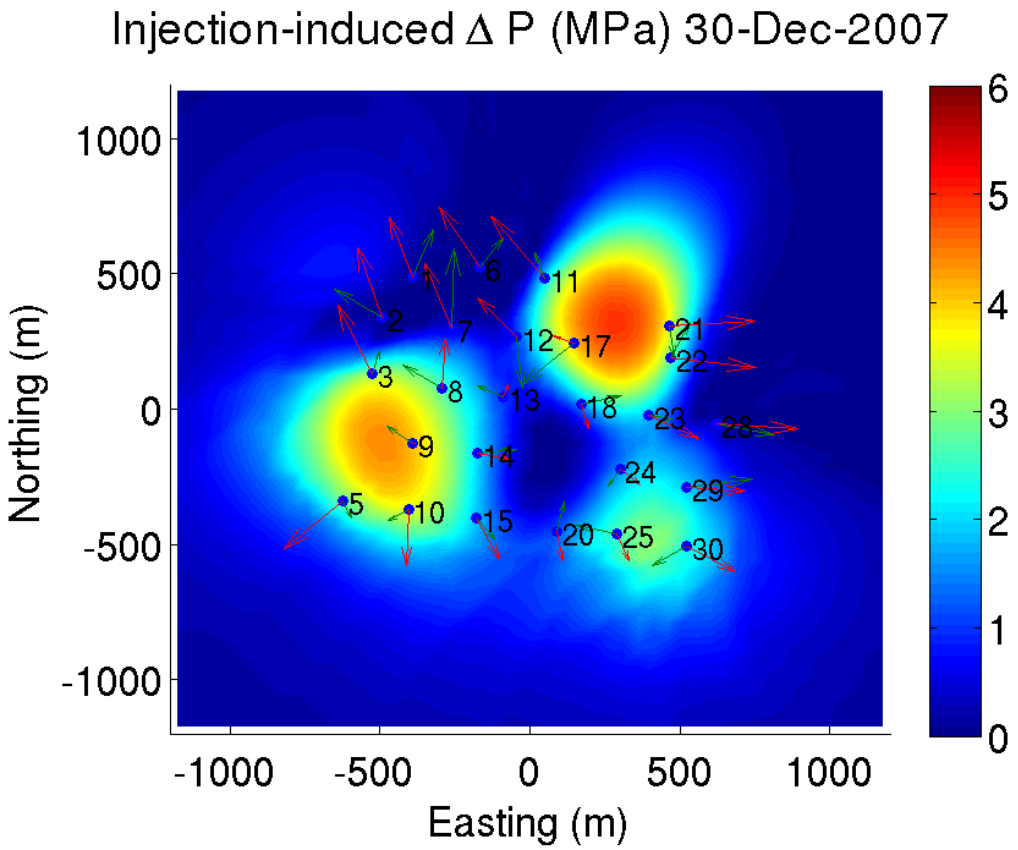


Figure 7.6: Inverted cumulative pore pressure change and differential tilt measurements for $\epsilon = 10^{-3}$ during Phase 2 of Cycle 1. Day 46 of the injection. [**CR**] tilts/. TdP46

for a wide range of regularization parameters $10^{-3} \leq \epsilon < 10^{-2}$, and show our results in Figures 7.8(a)-7.10(c). Increasing the value of the regularization parameter has, as expected, a smoothing effect (Aster et al., 2011) on the inverted pore pressure change, but does not change the qualitative picture of pressure front propagation.

CONCLUSIONS AND PERSPECTIVES

Injection-induced pore pressure changes can be stably estimated from surface tilt measurements. While quantitative estimates are affected by the uncertainty in medium and reservoir parameters, the inversion provides a useful insight into the temporal evolution of pressure profiles. Well log data could provide more accurate parameter definition and allow use of vertically heterogeneous medium models in equations 6.13. The scarcity of data and relatively low accuracy of tilt measurements result in highly ill-posed inversion problems that are, however, amenable to regularization and multi-scale solution. Note that satellite differential radar interferometry has produced maps of surface displacement with subcentimeter-level precision (Zebker et al., 1994) and many of the disadvantages of using tilt measurement data can be remedied by complementing tilt data with differential GPS or InSAR observations (Segall, 2010). I caution that the extent to which the theory of linear poroelastic deformation can be applied to heavy oil sands is not fully understood. For example, thermal effects may significantly alter heavy oil formations and their poroelastic properties. Furthermore, modeling steam injection in the presence of "wormholes" created as a result of sand production (Hinkle and Batzle, 2006), and preexisting hydraulically conducting faults may require using techniques similar to those modeling fluid-filled chambers that are ubiquitous in volcanology (Segall, 2010). Time-lapse seismic surveys and microsesimic data may provide spatial constraints on the location of activated fluid-filled faults and very high-permeability areas within the reservoir, and this information can be used to modify the underlying deformation model. However, the linear poroelastic deformation model, effectively equivalent to modeling reservoir expansion using distributed dilatational sources, appears to be a useful first approximation that can provide at least qualitative insight into the propagation of steam fronts and reservoir heterogeneity.

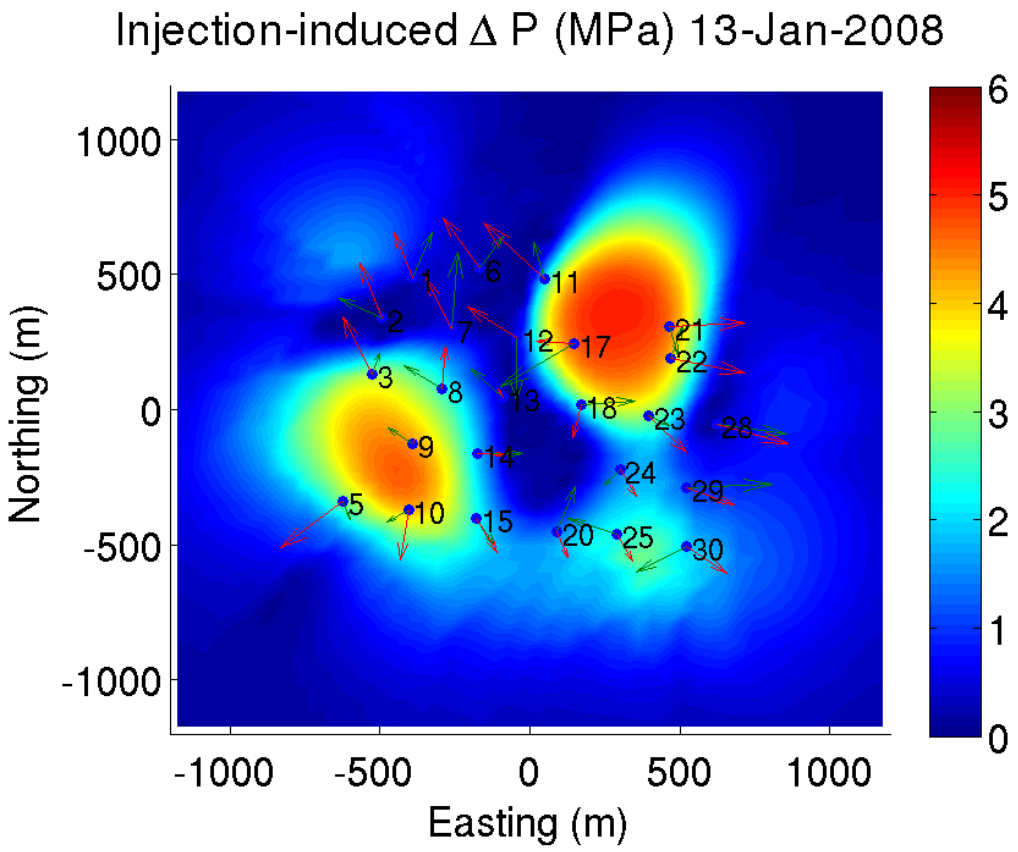
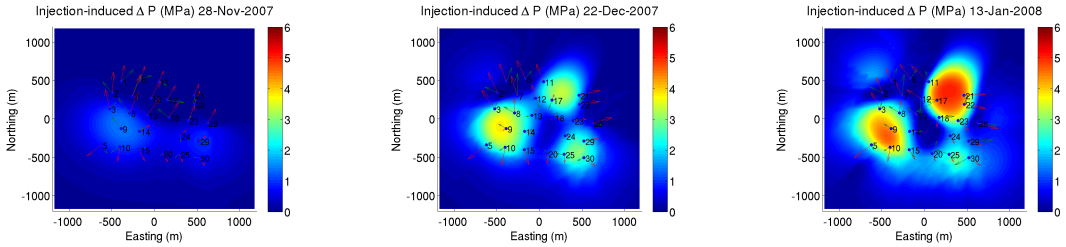
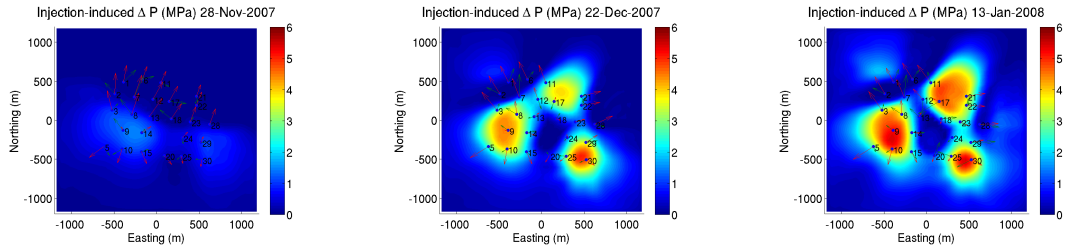


Figure 7.7: Inverted cumulative pore pressure change and differential tilt measurements for $\epsilon = 10^{-3}$ during Phase 2 of Cycle 1. Day 60 of the injection. Note the increase of pressure from Figure 7.6 to Figure 7.7 in the northeast of the computational domain. This matches the geometry of injection wells within the reservoir in Figure 7.1. [**CR**] tilts/. TdP60



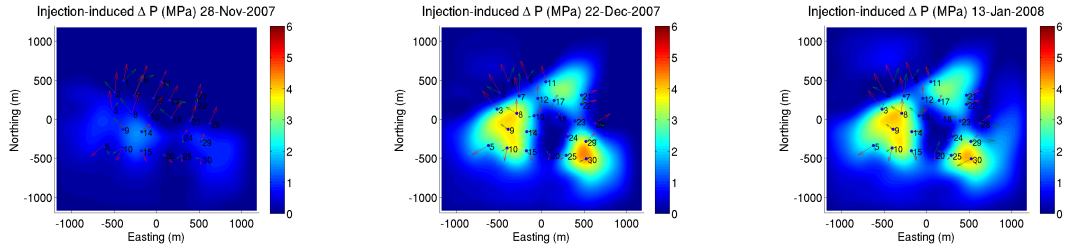
(a) (b) (c)

Figure 7.8: Cumulative pore pressure change with $\epsilon = 10^{-3}$ for (a) 14 (b) 38 (c) 60 days into the injection cycle. [**CR**] tilts/. 1TdP14,1TdP38,1TdP60



(a) (b) (c)

Figure 7.9: Cumulative pore pressure change with $\epsilon = 5 \times 10^{-3}$ for (a) 14 (b) 38 (c) 60 days into the injection cycle. [**CR**] tilts/. 5TdP14,5TdP38,5TdP60



(a) (b) (c)

Figure 7.10: Cumulative pore pressure change with $\epsilon = 10^{-3}$ for (a) 14 (b) 38 (c) 60 days into the injection cycle. [**CR**] tilts/. 2TdP14,2TdP38,2TdP60

Chapter 8

Characterization of reservoir heterogeneity

"Vision is the art of seeing things invisible"

Jonathan Swift

In the previous chapter I have demonstrated inversion of a complex evolution of pressure fronts in a heavy oil reservoir undergoing cyclic steam stimulation. My inversion method was based on solving a regularized inverse problem for inverting pore pressure change from surface displacements. In this chapter I extend the method to recover sharp contrasts in induced reservoir pressure that may be due to permeability barriers or fluid-conducting faults. I demonstrate the method by inverting pore pressure changes from uplift observations for a synthetic model of a heterogeneous reservoir undergoing fluid injection. Using my new computationally efficient technique for constrained TV-regularized optimization, I invert values and locations of sharp pressure contrasts from noisy measurements of surface deformation, and estimate the location of an impermeable boundary between reservoir compartments.

RECOVERING PRESSURE CONTRASTS

In this chapter I extend the method of chapters 6 and 7 to identifying sharp pressure contrasts in reservoirs undergoing injection, with the aim of applying it in the characterization of reservoir heterogeneities. I solve the pressure inversion problem as a bound-constrained optimization problem with a total-variation regularization, described in chapter 6, equations (6.9, 6.8) and compare the obtained solution with the result of solving the Tikhonov-regularized problem (6.7, 6.8) solved in the previous chapter.

While our numerical experiments indicate that both approaches are effective in resolving heterogeneous pressure profiles, the bound-constrained TV-regularized method enhances the "blockiness" of estimated pressure profiles, preserving boundaries between large-scale features—see appendix C and Maharramov and Levin (2015). I demonstrate both methods on a synthetic model of a reservoir featuring a low-permeability barrier between two highly permeable compartments.

COMPARISON OF TWO METHODS

I used the poroelastic medium parameters shown in Table 7, and the isotropic reservoir permeability shown in Figure 8.1, with the permeability of the barrier (shown in blue) ranging between 1 and 10 millidarcies. Water was injected at a constant well pressure of 15 MPa for up to 60 days at injection locations shown in Figure 8.2. Pressure evolution within the reservoir was simulated assuming a single phase flow, snapshots of the "true pore pressure change on day 20, 40 and 60 are shown in Figures 8.4(a), 8.5(a), 8.6(a). Random Gaussian noise with $\sigma = .05$ was added to the forward-modeled uplift measurements, see Figures 8.3(a), 8.3(b), 8.3(c). Forward-modeling and inversion were conducted on a 50 by 50 computational grid spanning a 2 km by 2 km area. A 500-sample spatial subset of the modeled surface uplift was used in the inversion to imitate realistic spatial constraints of field measurements. With only 500 spatial data point, problem (6) is ill-posed and requires regularization. I solve the constrained problem (6.9,6.8). The maximum pressure change limit of 7 MPa was set in (6.8),

estimated from the formation hydraulic fracturing pressure. Potential sparsity and

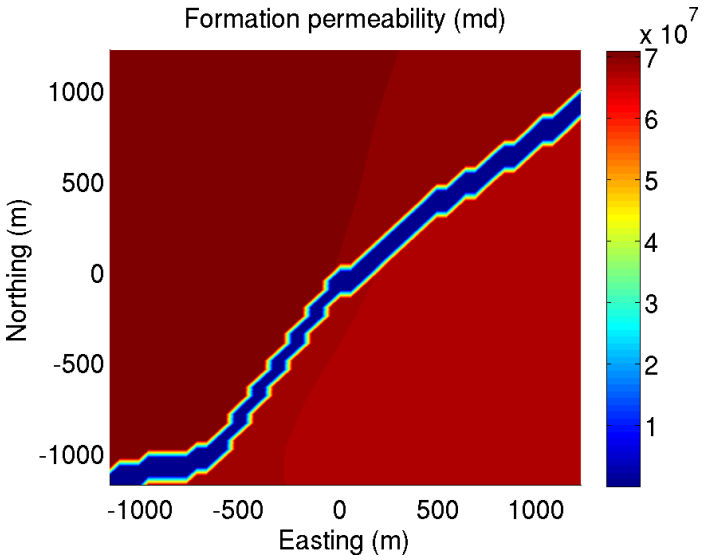


Figure 8.1: Reservoir permeability. [**CR**] char/. Perm

limited spatial distribution of uplift measurements result in non-unique fitting, or an underdetermined problem. Problem (6.7,6.8) was solved using the technique of Chapters 6,7 and the resulting snapshots of estimated pore pressure change on day 20, 40, and 60 are shown in the right panels of Figures 8.4(b),8.5(b),8.6(b) versus the true pressure change shown in the left panels. Note the very good agreement between the true and inverted pressure profiles, both quantitatively and qualitatively. Solution of (6.7,6.8) indicates the existence of a low-pressure zone roughly centered on the permeability barrier of Figure 8.1. To further improve our estimate of the boundary between the permeable compartments, we applied the technique of Maharramov and Levin (2015), based on solving the TV-regularized optimization problem with bound constraints (6.9,6.8). The result of solving this problem with $\epsilon = 1$ for day 40 is shown in Figure 8.7. As expected of a blockiness-promoting method, the solution features contrasts between large-scale pressure areas. In contrast to the corresponding solution of (6.7,6.8) shown in Figure 8.5(b), the level curves of the estimated pressure change in the low-pressure area match the geometry and location of the permeability barrier—compare the dashed line of Figure 8.7 with Figure 8.1. TV-regularized

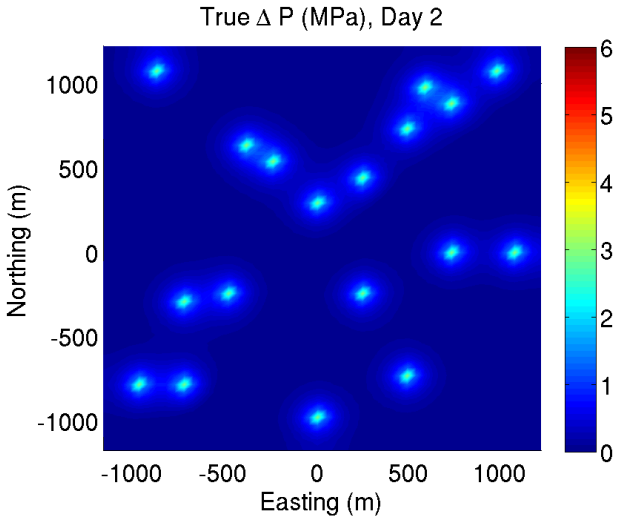
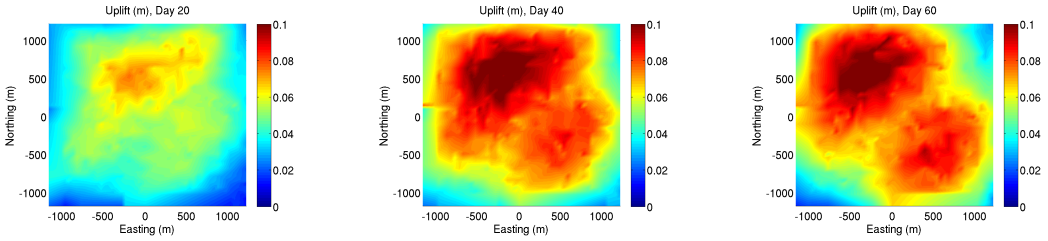


Figure 8.2: Locations of water injection wells. [**CR**] char/. tdP2



(a) (b) (c)

Figure 8.3: Noisy uplift measurements on (a) day 20, (b) day 40, (c) day 60. [**CR**] char/. uz20,uz40,uz60

inversion with bound constraints offers a mechanism for trading fitting fidelity for resolution of sharp pressure contrasts that characterize reservoir heterogeneity.

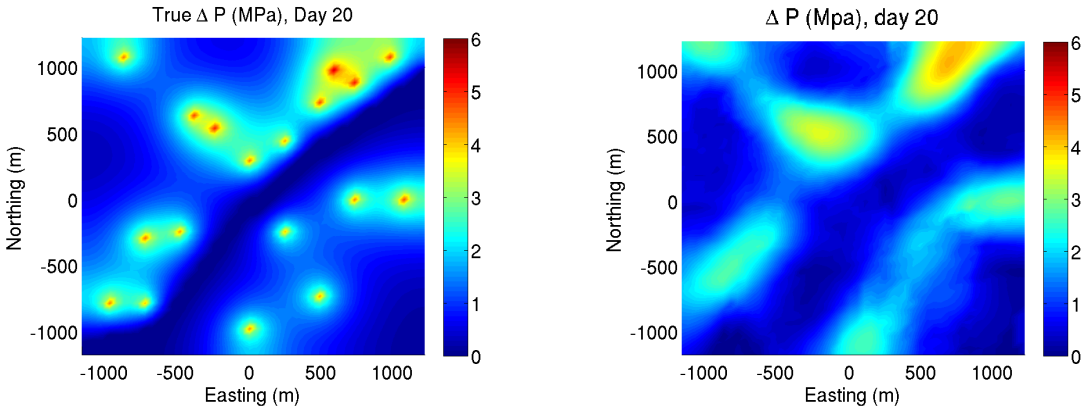
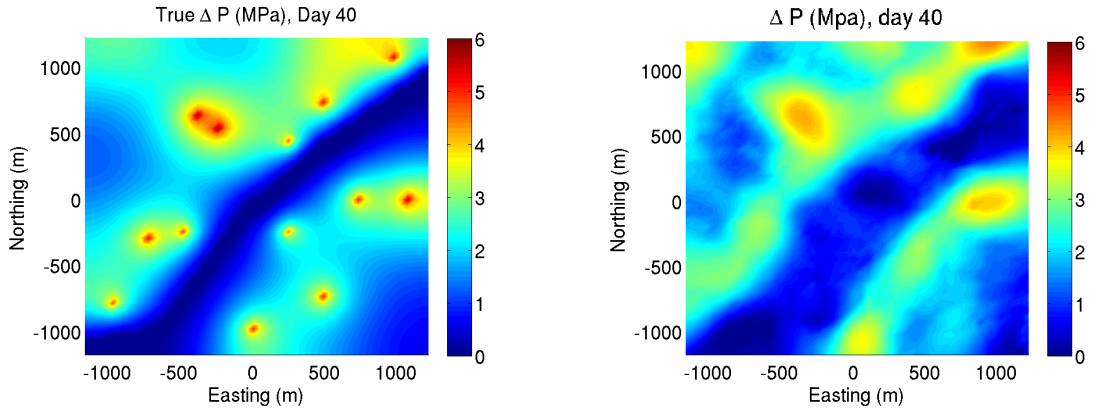
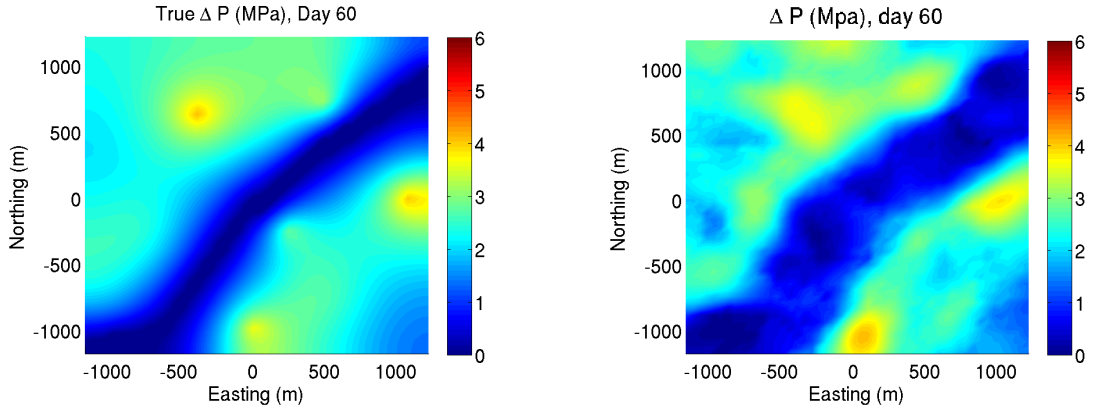


Figure 8.4: (a) True induced pore pressure change (color scale) on day 20. (b) Estimated induced pore pressure change inverted using (6.7,6.8). [**CR**] char/. tdP20,dP20



(a) (b)

Figure 8.5: (a) True induced pore pressure change (color scale) on day 60. (b) Estimated induced pore pressure change inverted using (6.7,6.8). [**CR**] char/. tdP40,dP40



(a) (b)

Figure 8.6: (a) True induced pore pressure change (color scale) on day 60. (b) Estimated induced pore pressure change inverted using (6.7,6.8). [**CR**] char/. tdP60,dP60

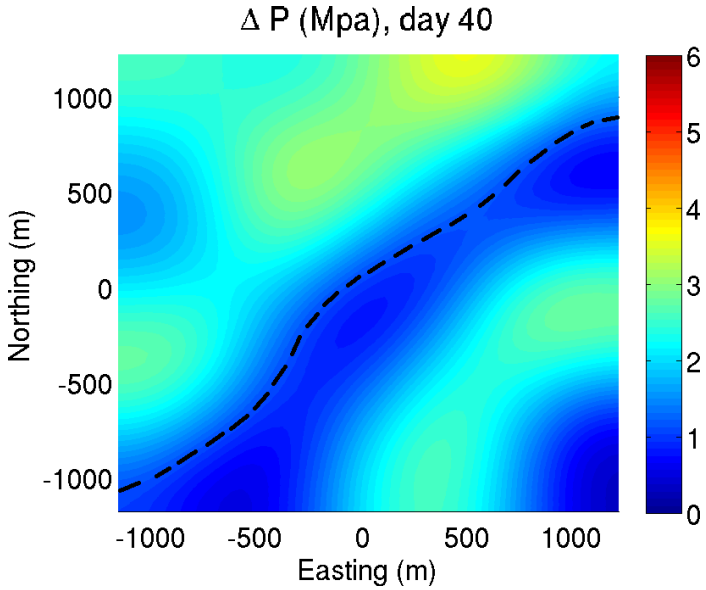


Figure 8.7: Estimated induced pore pressure change on day 40 using TV-regularized inversion (6.9,6.8) with bound constraints. Note that the permeability barrier of Figure 8.1 is well approximated by a level curve of the estimated pressure distribution. [**CR**] char/. 9tv

Part III

Appendices

Appendix A

Yet another guide to computing FWI objective functional

"There are a terrible lot of lies going about the world, and the worst of it is that half of them are true."

Winston Churchill

In this appendix I describe gradient computation by the adjoint-state method used in this thesis. With all derivations provided, this appendix presents a self-contained exposition of the adjoint-state method. Special emphasis is placed on explaining the relationship between the adjoint-state method and linearization of the forward modeling operator. The appendix concludes with three examples of applying the method in time and frequency-domain FWI. Unlike appendices B and C, this appendix does not present original research but is intended as a reference.

OBJECTIVE FUNCTIONS

Inverse problems of parameter estimation are often cast as nonlinear least squares fitting problems (Hansen, 1998; Aster et al., 2011):

$$J(\mathbf{m}) \;=\; \frac{1}{2}\|\mathbf{u}-\mathbf{d}\|_2^2 \;\rightarrow\; \min, \;\; \mathbf{u} \;=\; \mathbf{F}(\mathbf{m}), \tag{A.1}$$

where $\mathbf{m} \in \mathbb{R}^N$ denotes model, $\mathbf{u}, \mathbf{d} \in \mathbb{R}^M$ denote predicted and observed data, and $\mathbf{F} : \mathbb{R}^N \rightarrow \mathbb{R}^M$ is a (generally) nonlinear forward modeling operator. The objective of parameter estimation is to minimize the data misfit (A.1) by varying $\mathbf{m}$, often subject to additional regularization as in

$$\frac{1}{2}\|\mathbf{F}(\mathbf{m})-\mathbf{d}\|_2^2 \;+\; \Phi(\mathbf{m}) \;\rightarrow\; \min, \tag{A.2}$$

where is Φ is a regularization term. This work is concerned with solving very large problems, with the dimensions of $\mathbf{m}$ and $\mathbf{d}$ in the order of 10^5–10^9, with both computational time [1] and storage constraining our choice of numerical solvers. The nonlinear conjugate gradients method and quasi-Newton methods (Nocedal and Wright, 2006) are the most widely used derivative-based local optimization methods for solving large-scale problems as these methods only require (repetitive) evaluation of (A.2) and its gradient with respect to $\mathbf{m}$, without Hessian evaluation[2]. Therefore, efficient computation of the objective function gradient is one key to successful application of these methods.

Note that the regularization term $\Phi(\mathbf{m})$ in (A.2) is usually a known explicit function of the model $\mathbf{m}$, and evaluating derivatives with respect to $\mathbf{m}$ is computationally straightforward. However, this appendix describes a framework for efficient numerical differentiation of objective functions that are even more general than (A.2).

[1] of applying the forward-modeling operator $\mathbf{F}$

[2] Other methods such as *truncated* Newton can be adapted to solving large-scale problems so that the Hessian (either exact or approximate) is not required to be known component-wise.

Differentiating (A.1) with respect to $m^j, j = 1, \ldots, N$ we obtain

$$\frac{\partial J}{\partial m^j} \;=\; \left[\frac{\partial F^i}{\partial m^j}\right]^T (\mathbf{F}(\mathbf{m}) - \mathbf{d}) \;=\; \sum_{i=1}^{M} \frac{\partial F^i}{\partial m^j} \left(F^i(\mathbf{m}) - d^i\right) \;=\; \sum_{i=1}^{M} \frac{\partial u^i}{\partial m^j} \left(u^i - d^i\right). \tag{A.3}$$

From (A.3) we immediately see that the gradient of the objective function equals the adjoint of the Jacobi matrix $\partial\mathbf{F}/\partial\mathbf{m} = \partial\mathbf{u}/\partial\mathbf{m}$ applied to the data residual $\mathbf{F}(\mathbf{m}) - \mathbf{d}$. Formula (A.3) indicates that computation of the gradient requires a matrix-vector multiplication. It is important to note that for very large M and N component-wise evaluation of the Jacobi matrix may not be feasible.

☞ Without any additional assumptions about the modeling operator, no further computational simplification of (A.3) is possible. However, certain practically important explicit (e.g., using Green's functions) and *implicit* types of functional dependence between the modeled data $\mathbf{u}$ and modeling parameters $\mathbf{m}$ allow gradient computation without component-wise evaluation of the Jacobi matrix, as discussed in the following sections.

The adjoint-state method originated in the theory of PDE-constrained optimization (Chavent, 1971; Lions, 1971) and found its first geophysical applications in the seventies—see, e.g., Bamberger et al. (1977).

ADJOINT-STATE METHOD

In (A.1) we represented modeled data as an explicit function of modeling parameters, $\mathbf{u} = \mathbf{F}(\mathbf{m})$. However, forward modeling in typical applications often means solving a problem for operator equations[3]. After discretization, such problems are typically reduced to solving a system of M nonlinear equations with respect to M data variables

[3](pseudo-)differential, integral, integro-differential, etc.

$\mathbf{u} = [u^i], i = 1, \ldots, M$ that depend on N model parameters $\mathbf{m} = [m^j], j = 1, \ldots, N$. Before considering the most general formulation, we look into the important case of frequency-domain wave-equation modeling.

Wave-equation modeling

As an example, let us consider one-dimensional frequency-domain acoustic wave-equation modeling that requires solving

$$\left(s^2(x)\omega^2 \; + \; \Delta\right) p(x,\omega) + f(x,\omega) \; = \; 0, \; x \in \mathbb{R}^1, \tag{A.4}$$

where $s(x)$ is the slowness (reciprocal velocity) as a function of spatial coordinate x, $p(x,\omega)$ is the temporal Fourier transform of the pressure at x, and $f(x,\omega)$ is the temporal Fourier transform of a source function. Using a single frequency $\omega = \omega_0$ and spatial discretization $x = x^i, i = 1, \ldots, M$ (and appropriate boundary conditions), (A.4) can be approximated with the following *discrete* problem

$$(m^i)^2\omega^2 u^i \; + \; \sum_{k=1}^{M} \widetilde{\Delta}_k^i u^k + f(x^i,\omega) \; = \; 0, \; i = 1, \ldots, M, \tag{A.5}$$

where $m^i = s(x^i), \; u^i = p(x^i,\omega_0), \; f^i = f(x^i,\omega_0)$ and $\widetilde{\mathbf{\Delta}} = \left[\widetilde{\Delta}_k^i\right]$ is a finite-difference discretization of the Laplace operator *and* boundary conditions. In (A.5) the dimensions of the model space ($\mathbf{m}$) and data space ($\mathbf{u}$) are equal; however, we could have $N < M$ if we choose a coarser slowness discretization grid. Equation (A.5) means that forward-modeling of acoustic wave propagation requires solving a system of linear equations with respect to $\mathbf{u}$, with equation coefficients depending on $\mathbf{m}$. The system of equations (A.5) is the kind of implicit functional relationship between the model $\mathbf{m}$ and data $\mathbf{u}$ (after problem discretization) that was referred to above.

Method

In our formulation of the adjoint-state method, we assume the existence of an implicit functional dependence between model and data, applicable to *any* finite-dimensional modeling problem. Note that (A.5) can be represented in an operator notation as

$$\mathbf{S}(\mathbf{m}, \mathbf{u}) \;=\; \mathbf{0}, \tag{A.6}$$

where $\mathbf{S} : \mathbb{R}^N \times \mathbb{R}^M \to \mathbb{R}^M$ is a smooth mapping. We assume that the adjoint-state method is applied *after* discretization, and virtually any operator equation after discretization can be represented in the form of a finite-dimensional nonlinear functional dependence (A.6). Equation (A.6) establishes *implicit* dependence of $\mathbf{u}$ on $\mathbf{m}$ that is locally unique and smooth if the following condition holds

$$\det \left| \frac{\partial \mathbf{S}}{\partial \mathbf{u}} \right| \;\neq\; 0, \tag{A.7}$$

wherever equality (A.6) is satisfied. This statement is known as the *implicit function theorem* (Zorich, 2008) and is key to the existence and uniqueness of a solution to our modeling problem, and its smooth dependence on model parameters. We will now study the following problem:

Formulate a computationally efficient way of evaluating the gradient of

$$J(\mathbf{m}) \;=\; H(\mathbf{m}, \mathbf{u}), \tag{A.8}$$

☞ where $\mathbf{u}, \mathbf{m}$ satisfy (A.6) and H is an arbitrary smooth function. "Computational efficiency" means avoiding potential $O\left(\min(M, N)\right)^2$ storage and compute time requirements of component-wise $\partial\mathbf{u}/\partial\mathbf{m}$ Jacobi matrix computation.

Taking the *full* derivative of (A.6) with respect to $m^j, j = 1, \ldots, N$ and remembering that $\mathbf{u}$ is a function of $\mathbf{m}$, although implicit, we obtain

$$\frac{\partial S^i}{\partial m^j}(\mathbf{m}, \mathbf{u}(\mathbf{m})) \; + \; \sum_{k=1}^{M} \frac{\partial S^i}{\partial u^k}(\mathbf{m}, \mathbf{u}(\mathbf{m})) \frac{\partial u^k}{\partial m^j} \; = \; 0, \; i = 1, \ldots, M, \; j = 1, \ldots, N. \quad \text{(A.9)}$$

System (A.9) can be regarded as a system of equations with respect to the elements of the Jacobi matrix $\partial\mathbf{u}/\partial\mathbf{m}$, or *variational equations* (Gelfand and Fomin, 2000). Solving variational equations is the standard technique for computing Jacobi matrices for solutions of operator equations. It should be noted that even for simple linearized operators $\partial\mathbf{S}/\partial\mathbf{u}$, solving (A.9) for $M \times N$ components of the Jacobi matrix is computationally challenging unless data or model space dimension is small. However, if we are only interested in measuring predicted *response* to a specified change in model parameters,

$$\Delta\mathbf{u} \; = \; [\Delta u^i] \; \approx \; \frac{\partial \mathbf{u}}{\partial \mathbf{m}} \Delta\mathbf{m} \; = \; \left[\sum_{j=1}^{N} \frac{\partial u^i}{\partial m^j} \Delta m^j \right], \; = \; i = 1, \ldots, M, \quad \text{(A.10)}$$

—i.e., the result of applying the Jacobi matrix to a known model-space vector $\Delta\mathbf{m}$, then it can be evaluated trivially by solving a system of M linear equations with respect to $\Delta\mathbf{u}$

$$\sum_{j=1}^{N} \frac{\partial S^i}{\partial m^j}(\mathbf{m}, \mathbf{u}(\mathbf{m})) \Delta m^j \; + \; \sum_{k=1}^{M} \frac{\partial S^i}{\partial u^k}(\mathbf{m}, \mathbf{u}(\mathbf{m})) \Delta u^k \; = \; 0, \; i = 1, \ldots, M, \quad \text{(A.11)}$$

often called variational equations as well. System (A.11) is a linearization of (A.6) around solution $\mathbf{u} = \mathbf{u}(\mathbf{m})$. Note that (A.3) is the result of applying the *adjoint* of the Jacobi matrix to a known data-space vector. The rest of this section essentially is a demonstration of how the result of applying the adjoint Jacobi matrix to a known data-space vector can be obtained by solving the linearized adjoint equation.

Taking the *full* derivative of (A.8) with respect to $m^j, j = 1, \ldots, N$, we have

$$\frac{\partial J}{\partial m^j} \; = \; \frac{\partial H}{\partial m^j}(\mathbf{m}, \mathbf{u}(\mathbf{m})) + \sum_{k=1}^{M} \frac{\partial H}{\partial u^k}(\mathbf{m}, \mathbf{u}(\mathbf{m})) \frac{\partial u^k}{\partial m^j}, \; j = 1, \ldots, N. \quad \text{(A.12)}$$

In (A.12) the partial derivatives of H are trivial to compute and store. We will now eliminate the Jacobi matrix $\partial\mathbf{u}/\partial\mathbf{m}$ from (A.12) using (A.9). Indeed, from (A.9) using (A.7) we obtain

$$\frac{\partial u^k}{\partial m^j} \; = \; -\sum_{i=1}^{M} Z_i^k \frac{\partial S^i}{\partial m^j}(\mathbf{m}, \mathbf{u}(\mathbf{m})), \quad \text{(A.13)}$$

where

$$\mathbf{Z} \;=\; \left[\frac{\partial S^i}{\partial u^k}(\mathbf{m},\mathbf{u}(\mathbf{m}))\right]^{-1} \tag{A.14}$$

and the inverse exists because of (A.7). Substituting (A.13) into (A.12), we get

$$\frac{\partial J}{\partial m^j} \;=\; \frac{\partial H}{\partial m^j}(\mathbf{m},\mathbf{u}(\mathbf{m})) - \sum_{k=1}^{M} \frac{\partial H}{\partial u^k}(\mathbf{m},\mathbf{u}(\mathbf{m})) \sum_{i=1}^{M} Z_i^k \frac{\partial S^i}{\partial m^j}(\mathbf{m},\mathbf{u}(\mathbf{m})). \tag{A.15}$$

At this point we can make two important observations about (A.15): 1) for the implicit formulation (A.6), the partial derivatives of $\mathbf{S}$ with respect to $\mathbf{m}$ may be obtained trivially (for example, for (A.5) $\partial\mathbf{S}/\partial\mathbf{m}$ is a diagonal matrix) and 2) the second term in (A.15) can be represented as

$$\begin{aligned} \left[\frac{\partial H}{\partial \mathbf{u}}\right]^T \mathbf{Z}\frac{\partial \mathbf{S}}{\partial \mathbf{m}} \;&=\; \left[\mathbf{Z}^T\frac{\partial H}{\partial \mathbf{u}}\right]^T \frac{\partial \mathbf{S}}{\partial \mathbf{m}} \;= \\ &=\; \sum_{i=1}^{M}\left\{\sum_{k=1}^{M}\left[\mathbf{Z}^T\right]_k^i \frac{\partial H}{\partial u^k}(\mathbf{m},\mathbf{u}(\mathbf{m}))\right\}\frac{\partial S^i}{\partial m^j}(\mathbf{m},\mathbf{u}(\mathbf{m})), \end{aligned} \tag{A.16}$$

where the expression in the braces in (A.16) is the result of applying adjoint of the inverse of linear operator

$$\frac{\partial S^i}{\partial u^k}(\mathbf{m},\mathbf{u}(\mathbf{m}))$$

to the easily computable vector $\partial H/\partial\mathbf{u}$. Here is the source of the most important computational saving: rather than multiplying $\mathbf{Z}$ by another matrix as in (A.15), we apply its adjoint to a vector. Furthermore, from (A.14) we see that computing

$$\lambda^i \;=\; \sum_{k=1}^{M}\left[\mathbf{Z}^T\right]_k^i \frac{\partial H}{\partial u^k}(\mathbf{m},\mathbf{u}(\mathbf{m})), \tag{A.17}$$

is equivalent to solving the following system of equations

$$\begin{aligned} \sum_{i=1}^{M} \left[\mathbf{Z}^{-T}\right]_i^k \lambda^i &= \frac{\partial H}{\partial u^k}(\mathbf{m}, \mathbf{u}(\mathbf{m})) \Leftrightarrow \\ \sum_{i=1}^{M} \frac{\partial S^i}{\partial u^k}(\mathbf{m}, \mathbf{u}(\mathbf{m}))\lambda^i &= \frac{\partial H}{\partial u^k}(\mathbf{m}, \mathbf{u}(\mathbf{m})), \end{aligned} \tag{A.18}$$

i.e., to evaluate $\boldsymbol{\lambda}$ (often referred as the "adjoint field") we simply need to solve a linear system with the adjoint of the linear operator obtained by linearizing implicit equations (A.6) around the forward-modeled solution $\mathbf{u} = \mathbf{u}(\mathbf{m})$. Once (A.18) is solved, $\boldsymbol{\lambda}$ is substituted into (A.16), where the adjoint of operator

$$\frac{\partial \mathbf{S}}{\partial \mathbf{m}} = \left[\frac{\partial S^i}{\partial m^j}(\mathbf{m}, \mathbf{u}(\mathbf{m})) \right], \; i = 1, \ldots, M, \; j = 1, \ldots, N. \tag{A.19}$$

is applied to it.

☞ Utility of the adjoint state method lies in the fact that in many practical problems the computational cost of solving (A.18) does not exceed that of solving (A.6), and the linearized operators in the left-hand side of (A.18) and in (A.19) are sparse, can be derived analytically, and/or have efficient "black box" implementations.

Adjoint-state recipe

We can now describe detailed steps for evaluating the gradient of (A.8) for a discrete optimization problem. Assume that forward modeling of $\mathbf{u}$, given model parameters $\mathbf{m}$, requires solution of (A.6), and we have a method for solving (A.6). At the expense of some repetition, we provide all the equations and definitions required for an abstract adjoint-state application framework.

1. Given $\mathbf{m}$, solve (A.6) for $\mathbf{u} = \mathbf{u}(\mathbf{m})$.

2. Compute
$$\frac{\partial H}{\partial \mathbf{u}} \;=\; \left[\frac{\partial H}{\partial u^k}(\mathbf{m}, \mathbf{u}(\mathbf{m}))\right],\; k = 1, \ldots, M. \tag{A.20}$$

3. Linearize (A.6) around $\mathbf{u}(\mathbf{m})$—i.e., derive or compute the linear operator
$$\mathbf{L} \;=\; \frac{\partial \mathbf{S}}{\partial \mathbf{u}} \;=\; \left[\frac{\partial S^i}{\partial u^k}(\mathbf{m}, \mathbf{u}(\mathbf{m}))\right],\; i, k = 1, \ldots, M. \tag{A.21}$$

4. Compute $\boldsymbol{\lambda} = \lambda^i, i = 1, \ldots, M$ by solving the linearized adjoint equation
$$\mathbf{L}^T \boldsymbol{\lambda} \;=\; \frac{\partial H}{\partial \mathbf{u}} \tag{A.22}$$
with the right-hand side (A.20), or, in components,
$$\sum_{i=1}^{M} \frac{\partial S^i}{\partial u^k}(\mathbf{m}, \mathbf{u}(\mathbf{m}))\lambda^i \;=\; \frac{\partial H}{\partial u^k}(\mathbf{m}, \mathbf{u}(\mathbf{m})). \tag{A.23}$$
To solve (A.22), use the derived explicit form of $\mathbf{L}$ or a "black box" operator application routine for $\mathbf{L}$ to avoid component-wise matrix operations and storage.

5. Compute
$$\frac{\partial H}{\partial \mathbf{m}} \;=\; \left[\frac{\partial H}{\partial m^j}(\mathbf{m}, \mathbf{u}(\mathbf{m}))\right],\; j = 1, \ldots, N. \tag{A.24}$$

6. Linearize (A.6) around $\mathbf{m}$—i.e., *derive* or compute the linear operator
$$\mathbf{G} \;=\; \frac{\partial \mathbf{S}}{\partial \mathbf{m}} \;=\; \left[\frac{\partial S^i}{\partial m^j}(\mathbf{m}, \mathbf{u}(\mathbf{m}))\right],\; i = 1, \ldots, M,\; j = 1, \ldots, N. \tag{A.25}$$

7. Compute the gradient as
$$\frac{\partial J}{\partial \mathbf{m}} \;=\; \frac{\partial H}{\partial \mathbf{m}} \;-\; \mathbf{G}^T \boldsymbol{\lambda}, \tag{A.26}$$

or, in components,

$$\frac{\partial J}{\partial m^j} \;=\; \frac{\partial H}{\partial m^j}(\mathbf{m}, \mathbf{u}(\mathbf{m})) \;-\; \sum_{i=1}^{M} \frac{\partial S^i}{\partial m^j}(\mathbf{m}, \mathbf{u}(\mathbf{m}))\lambda^i. \tag{A.27}$$

To compute $\mathbf{G}^T\boldsymbol{\lambda}$ in (A.26), use the derived explicit form of $\mathbf{G}$, or a "black box" operator application routine for $\mathbf{G}$ to avoid component-wise matrix operations and storage.

EXAMPLES

In this section we discuss two applications of the adjoint-state method to PDE-constrained optimization. First, we return to the frequency-domain wave equation example discussed above. A step-by-step description is provided for applying the adjoint state method to a one-dimensional full-waveform inversion problem. Next, we consider a time-domain formulation of the same problem and again derive all the steps of the method. We demonstrate how the framework can be applied to complex-valued data.

Frequency-domain full-waveform inversion

Here we apply the adjoint-state method to computing the gradient of objective function (A.1) with respect to the slowness. Function $\mathbf{u}$ is a solution of (A.4) for $x \in [a, b]$, satisfying zero boundary conditions

$$u(z = a, \omega) \;=\; u(x = b, \omega) \;=\; 0. \tag{A.28}$$

We discretize (A.4) over a grid $x^i, i = 0, 1, \ldots, M, M+1$, with $x^0 = a, x^{M+1} = b$, using the second-order central finite-difference (Ascher, 2008) to obtain (A.5). Operator $\widetilde{\boldsymbol{\Delta}}$

(discretized Laplacian) in (A.5) is given[4] by the following $M \times M$ matrix

$$\widetilde{\mathbf{\Delta}} \;=\; \frac{1}{\Delta x^2}\begin{bmatrix} -2 & 1 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \dots & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 & \dots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 & -2 \end{bmatrix}. \tag{A.29}$$

and functions $\mathbf{S} = [S^i], i = 1, \dots, M$ in the implicit functional relationship (A.6) are given by

$$S^i(\mathbf{m}, \mathbf{u}) \;=\; (m^i)^2\omega^2 u^i \;+\; \sum_{k=1}^{M} \widetilde{\Delta}^i_k u^k + f(x^i, \omega),\; i = 1, \dots, M. \tag{A.30}$$

Forward-modeling wavefield $\mathbf{u}$ for a given slowness function $\mathbf{m}$ requires solving system (A.5) for

$$u^i \;=\; u(x^i, \omega),\; i = 1, \dots, M; \tag{A.31}$$

note that $u(x^0) \;=\; u(^{M+1}) \;=\; 0$ from (A.28).

We will now formulate steps of the adjoint-state framework (A.20–A.26) for computing $[\partial J/\partial m^j]$. After forward-modeling the wavefield, we compute (A.20):

$$\frac{\partial H}{\partial \mathbf{u}} \;=\; \mathbf{u}(\mathbf{m}) - \mathbf{d}, \tag{A.32}$$

i.e., the *residual*. Since for (A.30) we have

$$\frac{\partial S^i}{\partial u^k}(\mathbf{m}, \mathbf{u}) \;=\; (m^i)^2\omega^2\delta^i_k \;+\; \widetilde{\Delta}^i_k,\; i, k = 1, \dots, M, \tag{A.33}$$

[4]for zero boundary conditions (A.28)

the *adjoint equation* is

$$\sum_{i=1}^{M}(m^i)^2\omega^2\delta_k^i\lambda^i \;+\; \sum_{i=1}^{M}\widetilde{\Delta}_k^i\lambda^i \;=\; u^k(\mathbf{m}) - d^k \;\; k=1,\ldots,M, \tag{A.34}$$

where $\mathbf{u}(\mathbf{m}) = [u^i(\mathbf{m})]$ is the forward-modeled wavefield, i.e., solution of (A.5). Note that $\mathbf{S}$ depends on $\mathbf{u}$ linearly, and the linearized adjoint matches the adjoint of the operator $\mathbf{S}$ as a linear operator applied to $\mathbf{u}$. Specifically, (A.29) is symmetric, i.e., *self-adjoint*, trivially leading to

$$(m^k)^2\omega^2\lambda^k \;+\; \sum_{i=1}^{M}\widetilde{\Delta}_i^k\lambda^i \;=\; u^k(\mathbf{m}) - d^k \;\; k=1,\ldots,M. \tag{A.35}$$

We see that the adjoint equation matches the forward-modeling equation except for the source term, which is now given by $\mathbf{d} - \mathbf{u}(\mathbf{m})$ (note the sign). Note that (A.24) is zero for objective function (A.1), and for operator (A.25) we have

$$\frac{\partial S^i}{\partial m^j}(\mathbf{m},\mathbf{u}) \;=\; \delta_j^i 2m^i\omega^2u^i,\;\; i=1,\ldots,M,\;\; j=1,\ldots,N=M, \tag{A.36}$$

yielding for the objective function gradient

$$\frac{\partial J}{\partial m^j} \;=\; -2m^j\omega^2u^j(\mathbf{m})\lambda^j. \tag{A.37}$$

In our formulation of the adjoint-state method we assumed all functions and solutions to be real-valued. In the context of frequency-domain full waveform inversion, that means that formula (A.37) applies when $u(x,\omega)$ and $d(x,\omega)$ are either temporal sine or cosine transforms of the predicted and observed data. However, the forward modeling and adjoint equations (A.5) and (A.35) can be solved for complex Fourier transform of the data. Note that the objective function (A.1) is simply the sum of similar misfits for the cosine and sine transforms, assuming real data. Hence, if $\mathbf{u}$ and $\boldsymbol{\lambda}$ now denote complex solutions of (A.5) and (A.35) with $\mathbf{d}$ as the complex temporal Fourier transform of the observed data, gradient of the total objective function will be the sum of two contributions, leading to

$$\begin{aligned}\frac{\partial J^{\text{total}}}{\partial m^j} &= -2m^j\omega^2\text{Re}\,u^j(\mathbf{m})\text{Re}\,\lambda^j - 2m^j\omega^2\text{Im}\,u^j(\mathbf{m})\text{Im}\,\lambda^j = \\ &= -2m^j\omega^2\text{Re}\left\{u^j(\mathbf{m})\overline{\lambda^j}\right\},\end{aligned} \tag{A.38}$$

where the horizontal bar denotes complex conjugate. Formula (A.38) is applicable to frequency-domain complex full-waveform inversion and is in the form used in this thesis.

The computational cost and memory requirements of computing (A.38) are driven by the requirements of solving the forward modeling problem (A.5) and adjoint equation (A.35). Since the two equations in our case differ only by the source term, the computational cost is at most twice the cost of forward modeling. The linearized adjoint operator matches the forward-modeling operator because the latter depends linearly on $\mathbf{u}$ and is self-adjoint (Kolmogorov and Fomin, 1999; Kantorovich and Akilov, 1982). Although the above derivation was carried out for a one-dimensional

frequency-domain problem, all the equations and derivations stay valid for multi-dimensional problems, with the discretized Laplacian (A.29) replaced with its multi-dimensional analogue.

Time-domain full-waveform inversion

We again consider the one-dimensional full-waveform inversion problem (A.1), however, this time, we assume that the forward modeling is carried in time domain:

$$\left(s^2(x)\frac{\partial^2}{\partial t^2} \; - \; \Delta\right) p(x,t) - f(x,t) \; = \; 0, \; x \in \mathbb{R}^1, \; t \in [0,T]\,. \tag{A.39}$$

As before, we follow Ascher (2008) and *first discretize then optimize* (DTO). For the time-domain formulation (A.39), the zero boundary conditions

$$p(z=a,t) \; = \; p(x=b,t) \; = \; 0, \; t \in [0,T]\,, \tag{A.40}$$

must be complemented with two initial conditions. Using Duhamel's principle, we can assume that the initial conditions are zero as any source can be included in the term $f(x,t)$ in a distributional sense (Vladimirov, 1971) so that

$$p(x, t \leq 0) \; = \; 0, \; \frac{\partial p}{\partial t}(x, t \leq 0) \; = \; 0. \tag{A.41}$$

We discretize (A.39) over a spatial grid $x^i, i = 0, 1, \ldots, M, M+1$ with $x^0 = a, \; x^{M+1} = b$ and temporal grid $t^l, l = -1, 0, 1, 2, \ldots, N_t$ with $t^0 = 0, \; t^{N_t} = T$ for some large M and N_t. How I choose time indexing will be discussed shortly. After discretization we have

$$u^{ik} \; = \; p(x^i, t^k), \; m^i \; = \; s(x^i), \; i = 0, 1, \ldots, M, M+1, \; k = -1, 0, 1, \ldots, N_t. \tag{A.42}$$

For boundary and initial conditions (A.40) and (A.41) after discretization we obtain

$$\begin{aligned} u^{0\ k} &= u^{M+1\ k} = 0,\ k = -1, 0, 1, \ldots, N_t \\ u^{i\ (-1)} &= u^{i\ 0} = 0,\ i = 0, 1, \ldots, M, M+1. \end{aligned} \tag{A.43}$$

To obtain discretized functional relations (A.6), we introduce S^{jn} that is a finite-difference discretization of (A.39) for

$$x = x^j,\ t = (n-1)\Delta t,\ \text{where } \Delta t = T/N_t,\ n = 1, \ldots, N_t. \tag{A.44}$$

Note that the retarded time indexing in (A.44) is different from the time grid indexing used for discretizing $u(x,t)$, as time index $l = 1$ now corresponds to $t = 0$. Using second-order central differences for both spatial and temporal derivatives, we obtain

$$\begin{aligned} S^{jn}(\mathbf{m}, \mathbf{u}) &= (m^j)^2 \sum_{l=1}^{N_t} \widehat{\Delta}_l^n u^{j\,l} - \sum_{k=1}^{M} \widetilde{\Delta}_k^j u^{k\,n-1} - f(x^j, t^{n-1}), \\ j &= 1, \ldots, M,\ n = 1, \ldots, N_t, \end{aligned} \tag{A.45}$$

where

$$\widehat{\mathbf{\Delta}} = \frac{1}{\Delta t^2} \overbrace{\begin{bmatrix} 1 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 \\ -2 & 1 & 0 & \ldots & 0 & 0 & 0 & 0 \\ 1 & -2 & 1 & \ldots & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & \ldots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \ldots & -2 & 1 & 0 & 0 \\ 0 & 0 & 0 & \ldots & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & \ldots & 0 & 1 & -2 & 1 \end{bmatrix}}^{N_t} \Bigg\} N_t. \tag{A.46}$$

Note that (A.46) is no longer self-adjoint, and the source term in (A.45) is time-shifted because of the retarded time-indexing chosen for S^{jn} in (A.44).

Matrix (A.46) is triangular, meaning that forward modeling can be performed very efficiently by *time marching*. More specifically, values of the wavefield for $t = \Delta t$ are computed from the source injected at $t = 0$ as

$$u^{i\,1} \;=\; (m^i)^{-2}\Delta t^2 f(x^i, 0),\; i = 1, \ldots, M, \tag{A.47}$$

because of the conditions (A.43). The subsequent time sections of the wavefield are computed as

$$\begin{gathered} u^{i\,k} \;=\; 2u^{i\,k-1} - u^{i\,k} + (m^i)^{-2}\Delta t^2 \left(\sum_{j=1}^{M} \widetilde{\Delta}_j^k u^{j\,k-1} + f(x^i, (k-1)\Delta t) \right), \\ i = 1, \ldots, M,\; k = 2, 3, \ldots, N_t. \end{gathered} \tag{A.48}$$

The explicit marching scheme (A.48) makes the time-domain formulation especially attractive for multi-dimensional full-waveform inversion. Solution of (A.5) requires solving a system of linear equations. In the one dimensional case, system (A.5) has a tridiagonal matrix and can be solved in a linear time $O(M)$. In a multi-dimensional case this system has a very *sparse* but not *banded* matrix that requires application of structured or iterative solvers (Golub and Van Loan, 1996; Trefethen and Bau III, 1997; Saad, 2003). An alternative to solving (A.5) in the frequency domain is to forward-model the wavefield in the time domain and apply the discrete Fourier transform to it (Sirgue et al., 2010a).

For the objective function

$$J(\mathbf{u}) \;=\; H(\mathbf{u}) \;=\; \frac{1}{2}\sum_{i=1}^{M}\sum_{k=1}^{N_t} \left(u^{i\,k} - d^{i\,k}\right)^2 \tag{A.49}$$

(squared 2-norm of the data misfit over space and time), after differentiating with respect to the wavefield we obtain

$$\frac{\partial H}{\partial u^{i\,k}} \;= u^{i\,k} - d^{i\,k},\; i = 1, \ldots, M,\; k = 1, \ldots, N_t, \tag{A.50}$$

i.e., the residual.

To derive the linearized adjoint equation (A.22) we use (A.32) and

$$\frac{\partial S^{jn}}{\partial u^{ik}}(\mathbf{m},\mathbf{u}) \;=\; (m^j)^2\delta_i^j\widehat{\Delta}_k^n \;-\; \delta_k^{n-1}\widetilde{\Delta}_i^j \tag{A.51}$$

to obtain

$$\sum_{n=1}^{N_t}(m^i)^2\widehat{\Delta}_k^n\lambda^{i\,n} - \sum_{j=1}^{M}\widetilde{\Delta}_i^j\lambda^{j\,k+1} \;=\; u^{i\,k}(\mathbf{m})-d^{i\,k}\;i=1,\ldots,M,\;k=1,\ldots,N_t, \tag{A.52}$$

where we add the "initial" condition

$$\lambda^{i\,N_t+1} \;=\; 0,\; i=1,\ldots,M. \tag{A.53}$$

Condition (A.53) is necessary to avoid the time index in (A.52) exceeding N_t. Without (A.53) adjoint equations (A.52) would have to include an extra formula for the special case of $k = N_t$, so (A.53) are introduced solely for notational and computational convenience. Operator $\widetilde{\mathbf{\Delta}}$ is given by (A.29) and is symmetric. Equation (A.52) is equivalent to

$$\sum_{n=1}^{N_t}(m^i)^2\widehat{\Delta'}_n^k\lambda^{i\,n} \;-\; \sum_{j=1}^{M}\widetilde{\Delta}_j^i\lambda^{j\,k} \;=\; u^{i\,k}(\mathbf{m}) - d^{i\,k}\;i=1,\ldots,M,\;k=1,\ldots,N_t, \tag{A.54}$$

where

$$\left[\widehat{\Delta'}_n^k\right] \;=\; \widehat{\mathbf{\Delta}'} \;=\; \widehat{\mathbf{\Delta}}^* \tag{A.55}$$

is the transpose of operator (A.46),

$$\widehat{\mathbf{\Delta}'} \;=\; \frac{1}{\Delta t^2}\overbrace{\left[\begin{array}{ccccccccc} 1 & -2 & 1 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \dots & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 & \dots & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 1 \end{array}\right]}^{N_t} \left.\begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \\ \\ \end{array}\right\} N_t. \tag{A.56}$$

Matrix (A.56) is upper-triangular, suggesting a marching for solving (A.54) analogous to (A.47,A.48) but applied "backwards". First, we initialize $\lambda^{i\,N_t+1},\ i=1,\dots,M$ to zero per (A.53). Next, we assign

$$\lambda^{i\,N_t} \;=\; (m^i)^{-2}\Delta t^2\left(u^{i\,N_t}(\mathbf{m}) - d^{i\,N_t}\right),\ i=1,\dots,M, \tag{A.57}$$

followed by the backwards marching scheme

$$\begin{aligned} \lambda^{i\,k} \;&=\; 2\lambda^{i\,k+1} - \lambda^{i\,k+2} + (m^i)^{-2}\Delta t^2 \sum_{j=1}^{M} \widetilde{\Delta}_i^j \lambda^{j\,k+1}, \\ i \;&=\; 1,\dots,N,\ k = N_t-1,\dots,1. \end{aligned} \tag{A.58}$$

From

$$\frac{\partial S^{j\,n}}{\partial m^i}(\mathbf{m},\mathbf{u}) \;=\; \delta_i^j 2m^j \sum_{l=1}^{N_t} \widehat{\Delta}_l^n u^{j\,l} \;=\; 2m^j \sum_{l=1}^{N_t} \widehat{\Delta}_l^n u^{j\,l}, \tag{A.59}$$

constructing operator $\mathbf{G}$ (A.25) and forming (A.26), we obtain

$$\frac{\partial J}{\partial m^i} \;=\; -\sum_{j=1}^{M}\sum_{n=1}^{N_t} \frac{\partial S^{j\,n}}{\partial m^i}\lambda^{j\,n} \;=\; -2m^i \sum_{n=1}^{N_t} \lambda^{i\,n} \sum_{l=1}^{N_t} \widehat{\Delta}_l^n u^{i\,l},\ i=1,\dots,M. \tag{A.60}$$

Expression (A.60) can be interpreted as being proportional to the dot-product of the second time derivative of the forward-modeled data with the data residual "back-propagated" in time—compare with (A.38).

Note that just as in the frequency-domain example of the previous section, the cost of computing the gradient is roughly twice the cost of forward modeling.

In both frequency and time-domain examples we assumed zero boundary conditions as implemented in spatial Laplacian (A.29). In geophysical acoustic wave propagation problems zero boundary conditions are equivalent to free surface boundary conditions and are typically applied at an air or water interface. *Absorbing boundary conditions* are applied along boundaries of the *computational domain*, significantly impacting finite-difference representation of the Laplacian. Note that in our examples $\mathbf{S}$ depends *linearly* on $\mathbf{u}$, significantly simplifying the linearized adjoint equation. Interesting practical cases of modeling operators may involve nonlinear or *quasilinear* dependence on $\mathbf{u}$. Note that in any case equation (A.22) is obviously linear and may be more amenable to numerical analysis than the (potentially) nonlinear forward-modeling equation (A.6).

Phase-only full-waveform inversion

Phase-only full-waveform inversion is based on minimizing a measure of misfit between phase information in the predicted and observed data. Although phase differences can be directly measured and used in an objective function, phase is inherently discontinuous, unless unwrapped, and the resulting objective function may suffer from non-differentiability(Fichtner, 2011). However, here I demonstrate how a continuous (and periodic) function of the phase can be used in an objective function thus avoiding non-differentiability. I demonstrate a phase-only FWI based on minimizing the following objective function

$$J(\mathbf{m}) \;=\; H(\mathbf{u}(\mathbf{m}),\mathbf{m}) \;=\; H(\mathbf{u}(\mathbf{m})) \;=\; \|\exp i\arg\mathbf{u} - \exp i\arg\mathbf{d}\|_2^2, \qquad \text{(A.61)}$$

where $\mathbf{d}$ and $\mathbf{u}$ are single-frequency components of the observed and predicted data[5] and arg denotes the argument of a complex number. The gradient of (A.61) can be computed within the framework of the frequency-domain adjoint state method

[5] we conduct the inversion in the frequency domain

described in an earlier example. From (A.38) we obtain

$$\frac{\partial J}{\partial m^j} \;=\; -2m^j\omega^2 \mathrm{Re}\left\{u^j(\mathbf{m})\overline{\lambda^j}\right\}, \tag{A.62}$$

where, using (A.23) and assembling real and complex components for $\boldsymbol{\lambda}$, we have

$$(m^k)^2\omega^2\lambda^k \;+\; \sum_{i=1}^{M}\widetilde{\Delta}_i^k\lambda^i \;=\; \frac{\partial H}{\partial \mathrm{Re}\, u^k} \;+\; i\frac{\partial H}{\partial \mathrm{Im}\, u^k},\; k=1,\ldots,M. \tag{A.63}$$

From (A.63) we can see that the main difference with the phase-and-amplitude frequency-domain inversion appears in the injected sources for the adjoint wavefield propagation, in the right-hand side of (A.63). The objective function (A.61) can be expressed via the forward-propagated wavefield as follows,

$$\begin{aligned} H(\mathbf{u}) \;&=\; \sum_k \left\{\cos\arg u^k - \cos\arg d^k\right\}^2 \;+\; \left\{\sin\arg u^k - \sin\arg d^k\right\}^2 \\ &=\; \sum_k \left\{\frac{\mathrm{Re}\, u^k}{|u^k|} - \frac{\mathrm{Re}\, d^k}{|d^k|}\right\}^2 \;+\; \left\{\frac{\mathrm{Im}\, u^k}{|u^k|} - \frac{\mathrm{Im}\, d^k}{|d^k|}\right\}^2, \end{aligned} \tag{A.64}$$

where $|\cdot|$ is the absolute value of a complex number, $|z| \;=\; \sqrt{(\mathrm{Re}\, z)^2 + (\mathrm{Im}\, z)^2}$. After substituting (A.64) into the right-hand side of of (A.63), introducing the real and imaginary components of the "phase misfit"[6]

$$\begin{aligned} \rho_1^k \;&=\; \cos\arg u^k - \cos\arg d^k \;=\; \frac{\mathrm{Re}\, u^k}{|u^k|} - \frac{\mathrm{Re}\, d^k}{|d^k|}, \\ \rho_2^k \;&=\; \sin\arg u^k - \sin\arg d^k \;=\; \frac{\mathrm{Im}\, u^k}{|u^k|} - \frac{\mathrm{Im}\, d^k}{|d^k|}, \end{aligned} \tag{A.65}$$

[6] I use the term "phase misfit" loosely here; the actual misfit (A.61) involves a difference of the exponents of phase.

for the injected source we have

$$
\begin{aligned}
&\frac{\partial H}{\partial \mathrm{Re}\, u^k} \;+\; i\frac{\partial H}{\partial \mathrm{Im}\, u^k} \\
&= \frac{2}{|u^k|}\left\{\rho_1^k\left(1-\frac{\left(\mathrm{Re}\, u^k\right)^2}{|u^k|^2}\right)-\rho_2^k\frac{\mathrm{Re}\, u^k \mathrm{Im}\, u^k}{|u^k|^2}\right\} \\
&+ \frac{2i}{|u^k|}\left\{\rho_2^k\left(1-\frac{\left(\mathrm{Im}\, u^k\right)^2}{|u^k|^2}\right)-\rho_1^k\frac{\mathrm{Re}\, u^k \mathrm{Im}\, u^k}{|u^k|^2}\right\} \\
&= \frac{2\sin\arg u^k}{|u^k|}\left\{\rho_1^k \sin\arg u^k-\rho_2^k\cos\arg u^k\right\} \\
&+ \frac{2i\cos\arg u^k}{|u^k|}\left\{\rho_2^k \cos\arg u^k-\rho_1^k\sin\arg u^k\right\} \\
&= \frac{2}{|u^k|}\left\{\sin\arg u^k - i\cos\arg u^k\right\}\left\{\rho_1^k \sin\arg u^k-\rho_2^k\cos\arg u^k\right\} \\
&= \frac{2i}{|u^k|}\left\{\cos\arg u^k + i\sin\arg u^k\right\}\left\{\rho_2^k \cos\arg u^k-\rho_1^k\sin\arg u^k\right\}.
\end{aligned}
\tag{A.66}
$$

Introducing a "phase misfit residual",

$$
\mathbf{r}^k \;\equiv\; \rho_1^k + i\rho_2^k, \tag{A.67}
$$

such that

$$
H(\mathbf{u}) \;\equiv\; \sum_k \|\mathbf{r}^k\|_2^2, \tag{A.68}
$$

expression (A.66) for the source is now reduced to

$$
\frac{\partial H}{\partial \mathrm{Re}\, u^k} \;+\; i\frac{\partial H}{\partial \mathrm{Im}\, u^k} \;=\; \frac{2i}{\overline{u}^k}\mathrm{Im}\left[\frac{\overline{u^k}}{|u^k|}r^k\right] \;=\; \frac{2i}{\overline{u}^k}\mathrm{Im}\left[\frac{\overline{u^k}}{|u^k|}\left(\frac{u^k}{|u^k|}-\frac{d^k}{|d^k|}\right)\right]. \tag{A.69}
$$

Expression (A.69) expresses the adjoint source via single-frequency components of the predicted and observed wavefields. As expected, if the two wavefields match after amplitude normalization, the gradient is zero. One significant disadvantage of the phase-only inversion is that it is even more prone to cycle skipping than the phase-and-amplitude inversion: the difference of the amplitude-normalized wavefields in

(A.69) may be zero, resulting in a stationary point of the objective function (A.61) even when wavefield amplitudes do not match. To obtain qualitative conditions of convergence for the phase-only inversion, let us assume that the inversion is conducted in a 1D medium of a characteristic dimension L (e.g., average travel distance), and the average acoustic velocity in the medium is v. For a single-frequency component of the observed data we have

$$\mathbf{d}(x) \;\approx\; \exp 2\pi i \frac{\omega}{v} x, \tag{A.70}$$

(note there is obviously no need for amplitude normalization in 1D). If our initial estimate of the acoustic velocity is $v + \Delta v$, then for the predicted data we have

$$\mathbf{u}(x) \;\approx\; \exp 2\pi i \frac{\omega}{v + \Delta v} x. \tag{A.71}$$

Inversion does not cycle-skip so long as the average phase difference accumulated between (A.70) and (A.71) over the characteristic distance L does not exceed 2π—i.e.,

$$\begin{aligned} 2\pi\omega \left| \frac{1}{v + \Delta v} - \frac{1}{v} \right| L \;<\; 2\pi \;\Leftrightarrow \\ \left| \frac{\Delta v}{v} \right| \;<\; \frac{v}{\omega L}, \end{aligned} \tag{A.72}$$

i.e., relative velocity errors should not exceed the ratio of the characteristic wavelength of the observed data to the characteristic distance of wave propagation[7]. One implication of (A.72) is a temporal multi-scale application of the (phase-only) frequency-domain FWI: inversion proceeds from low to high frequencies because for low frequencies the objective function (A.61) is more tolerant to velocity errors.

I illustrate the phase-only FWI method based on (A.62) and (A.69) using the synthetic example of Figure 4.1 from chapter 4. In this experiment, a gradient computed using (A.62) was supplied to a nonlinear conjugate gradients solver (Nocedal

[7]Compare with the discussion in the section on Rytov Series in chapter 2: (A.72) means that the effect of a velocity perturbation Δv over a distance L should not exceed the characteristic wavelength.

and Wright, 2006). The true velocity model was smoothed using a 1920 m window to produce the starting model of Figure A.1. Synthetic data was generated using a towed-streamer acquisition with the maximum offset of 5 km.

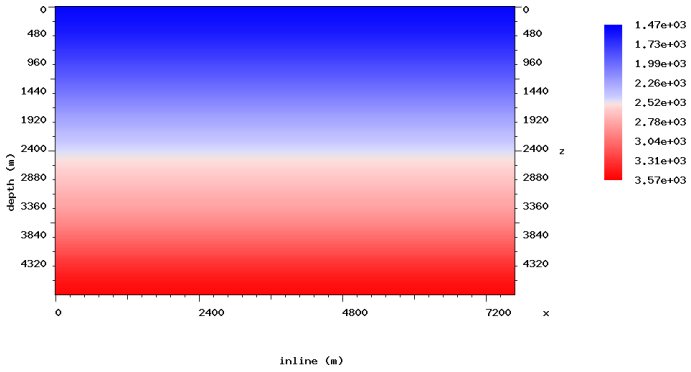


Figure A.1: Starting FWI model obtained by smoothing true model of Figure 4.1 using a 1920 m window. [**ER**] adjoint/. AAstarting

A temporal multi-scale frequency-domain phase-only inversion was conducted for the frequency range of 3–30 Hz. Ten iterations of nonlinear conjugate gradients were performed for each frequency inversion.

A gradient computed using (A.62) for $\omega = 3$ Hz is plotted in Figure A.2 after the first iteration of conjugate gradients, and in Figure A.3 after the tenth iteration. Cross-correlations (A.62) gradually emerge as long-wavelength periodic structures at depth.

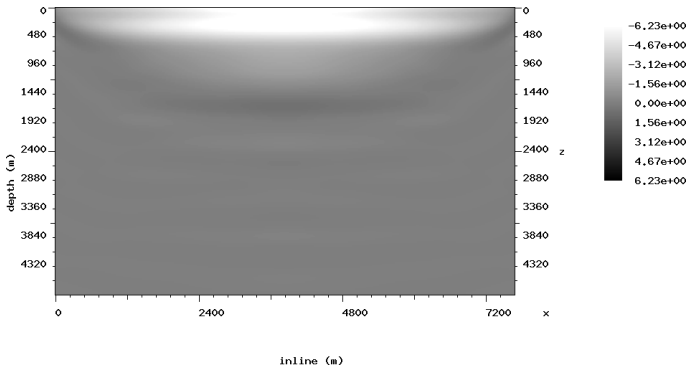


Figure A.2: A gradient computed using (A.62),(A.69) for $\omega = 3$ Hz, first iteration of conjugate gradients. Note that first arrival energy dominates the gradient, however, non-zero cross-correlations (A.62) are beginning to emerge at depth. [**CR**] adjoint/. AAgrad11

A gradient computed using (A.62) for $\omega = 7.4$ Hz is plotted in Figure A.4 after

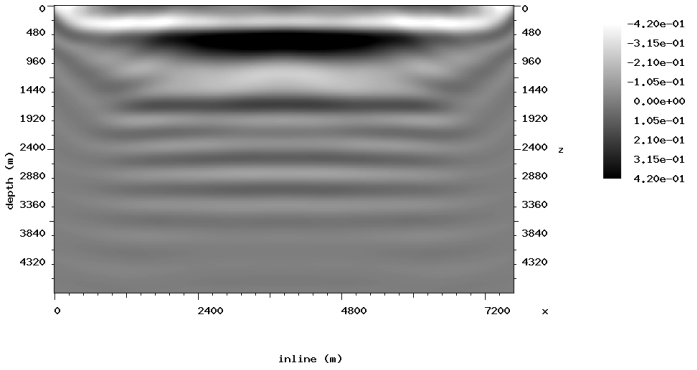


Figure A.3: A gradient of (A.61) computed using (A.62),(A.69) for $\omega = 3$ Hz, tenth iteration of conjugate gradients. Note the long-wavelength periodic cross-correlations (A.62) at depth. [**CR**] adjoint/. AAgrad110

the first iteration of conjugate gradients, and in Figure A.5 after the tenth iteration. Cross-correlations (A.62) appear as short-wavelength periodic structures.

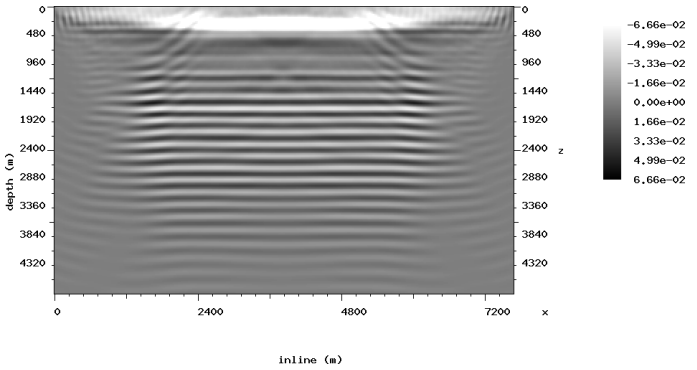


Figure A.4: A gradient of (A.61) computed using (A.62),(A.69) for $\omega = 7.4$ Hz, first iteration of conjugate gradients. Note the short-wavelength periodic cross-correlations (A.62) at depth. [**CR**] adjoint/. AAgrad101

Short and long-model updates for ω =3–30 Hz constructively interfered to produce the final result of Figure A.6. The phase-only FWI here does not cycle-skip, and this can be explained using the simple heuristic (A.72).

The average difference between the smoothed model of Figure A.1 and true model of Figure 4.1 does not exceed 5% of the VRMS velocity of approximately 2.2 km/s, hence

$$\frac{\Delta v}{v} \approx 0.05.$$

Characteristic (average) distance of wave propagation for our acquisition and 5 s maximum modeling time obviously does not exceed 11 km, hence $L \approx 11$. Substituting

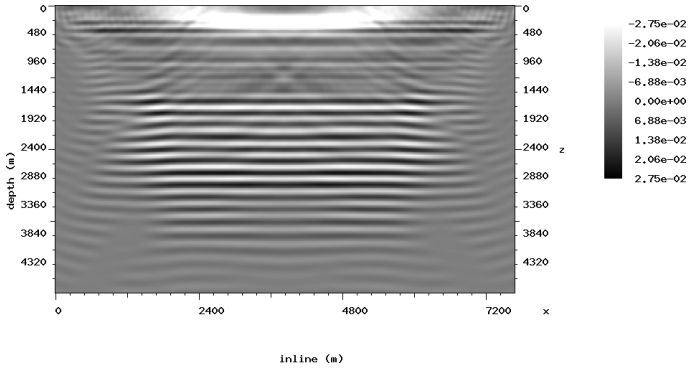


Figure A.5: A gradient of (A.61) computed using (A.62),(A.69) for $\omega = 7.4$ Hz, tenth iteration of conjugate gradients. The model is sufficiently close to the true solution for the gradient to point to the solution. [**CR**] adjoint/. AAgrad1010

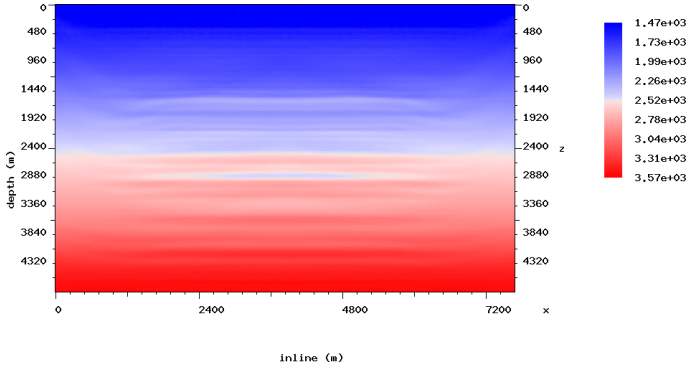


Figure A.6: Final inversion result for $\omega = 30$ Hz. Note that model updates for different frequencies constructively interfered to produce a quantitatively accurate approximation of the true model shown in Figure 4.1. [**CR**] adjoint/. AAfinal

these values and $\omega = 3$ Hz into (A.72) yields

$$0.05 \; < \; \frac{2.2}{3 \times 11} \; \approx \; .07.$$

A similar argument can be repeated for higher frequencies so long as each single-frequency inversion sufficiently reduces the velocity error Δv.

Appendix B

Compressive Conjugate Directions: Linear Theory

"Success is the sum of details."

Harvey S. Firestone

I present a powerful and easy-to-implement iterative algorithm for solving large-scale optimization problems that involve L_1/total-variation (TV) regularization. The method is based on combining the Alternating Directions Method of Multipliers (ADMM) with a Conjugate Directions technique in a way that allows reusing conjugate search directions constructed by the algorithm across multiple iterations of the ADMM. The new method achieves fast convergence by trading off multiple applications of the modeling operator for the increased memory requirement of storing previous conjugate directions. I illustrate the new method with a series of imaging and inversion applications.

INTRODUCTION

We address a class of regularized least-squares fitting problems of the form

$$\begin{aligned} &\|\mathbf{Bu}\|_1 \;+\; \frac{\alpha}{2}\|\mathbf{Au}-\mathbf{d}\|_2^2 \;\rightarrow\; \min, \\ &\mathbf{u} \;\in\; \mathbb{R}^N,\; \mathbf{d} \;\in\; \mathbb{R}^M,\; \mathbf{A}:\mathbb{R}^N \rightarrow \mathbb{R}^M,\; \mathbf{B}:\mathbb{R}^N \rightarrow \mathbb{R}^K,\; K \leq N, \end{aligned} \tag{B.1}$$

where $\mathbf{d}$ is a known vector (data), $\mathbf{u}$ a vector of unknowns[1], and $\mathbf{A}, \mathbf{B}$ are linear operators. If $\mathbf{B}$ is the identity map, then problem (B.1) is a least-squares fitting with L_1 regularization,

$$\|\mathbf{u}\|_1 \;+\; \frac{\alpha}{2}\|\mathbf{Au}-\mathbf{d}\|_2^2 \;\rightarrow\; \min\,. \tag{B.2}$$

If the unknown vector $\mathbf{u}$ is the discretization of a function, and $\mathbf{B}$ is the first-order finite difference operator

$$(\mathbf{Bu})_i \;=\; u_{i+1}-u_i,\; i=1,2,\ldots,N-1,$$

then problem (B.1) turns into a least-squares fitting with a total-variation (TV) regularization

$$\|\nabla\mathbf{u}\|_1 \;+\; \frac{\alpha}{2}\|\mathbf{Au}-\mathbf{d}\|_2^2 \;\rightarrow\; \min\,. \tag{B.3}$$

On the one hand, in (B.2) we seek a model vector $\mathbf{u}$ such that forward-modeled data $\mathbf{Au}$ match observed data $\mathbf{d}$ in the least squares sense, while imposing sparsity-promoting L_1 regularization. In (C.1), on the other hand, we impose blockiness-promoting total-variation (TV) regularization. Note that rather than using a regularization parameter as a coefficient of the regularization term, we use a data-fitting weight α. TV regularization (also known as the Rudin-Osher-Fatemi, or ROF, model Rudin et al. (1992)) acts as a form of "model styling" that helps to preserve sharp contrasts and boundaries in the model even when spectral content of input data has a limited resolution.

[1] sometimes referred to as "model"

L_1-TV regularized least-squares fitting, a key tool in imaging and de-noising applications (see, e.g. Rudin et al. (1992); Chambolle and Lions (1997); Vogel and Oman (1996); Kim et al. (2007)), is beginning to play an increasingly important role in applications where the modeling operator $\mathbf{A}$ in (B.1) is computationally challenging to apply. In particular, in seismic imaging problems of exploration geophysics such as full-waveform inversion Tarantola (1984); Fichtner (2011) modeling of seismic wave propagation in a three-dimensional medium from multiple seismic sources is by far the greatest contributor to the computational cost of inversion, and reduction of the number of applications of the operator $\mathbf{A}$ is key to success in practical applications.

L_1-regularized least-squares problems can be reduced to inequality-constrained quadratic programs and solved using interior-point methods based on, e.g., Newton Boyd and Vandenberghe (2004) or nonlinear Conjugate Gradients Kim et al. (2007) methods. Alternatively, the resulting bound-constrained quadratic programs can be solved using gradient projection Figueiredo et al. (2007) or projected Conjugate Gradients Qiu et al. (2012). A conceptually different class of techniques for solving L_1-regularized least-squares problems is based on homotopy methods Hastie et al. (2004); Efron et al. (2004); Osborne et al. (2000).

Another class of methods for solving (B.1) that merits a special mention applies splitting schemes for the sum of two operators. For example the *iterative shrinking-thresholding algorithm* (ISTA) is based on applying *forward-backward splitting* Bruck Jr. (1977); Passty (1979) to solving the L_1-regularized problem (B.2) by gradient descent Bioucas-Dias and Figueiredo (2007); Combettes and Wajs (2005); Daubechies et al. (2004):

$$\begin{aligned} \mathbf{y}_{k+1} &= \mathbf{u}_k \; - \; \gamma\alpha\mathbf{A}^T\left(\mathbf{A}\mathbf{u}_k - \mathbf{d}\right), \\ \mathbf{u}_{k+1} &= \text{shrink}\left\{\mathbf{y}_{k+1}, \gamma\right\}, \end{aligned} \tag{B.4}$$

where $\gamma > 0$ is a sufficiently small step parameter, and the *soft thresholding* or *shrinkage* operator is the Moreau resolvent (see, e.g., Bauschke and Combettes (2011))

of $\partial\gamma\|\mathbf{u}\|_1$,

$$\begin{aligned}\text{shrink}\left\{\mathbf{y},\gamma\right\} \;&=\; \left(1+\partial\gamma\|\mathbf{y}\|_1\right)^{-1} \;=\; \text{argmin}_{\mathbf{x}}\left\{\gamma\|\mathbf{x}\|_1+\frac{1}{2}\|\mathbf{y}-\mathbf{x}\|_2^2\right\} \;= \\ &\frac{\mathbf{y}}{|\mathbf{y}|}\max\left(|\mathbf{y}|-\gamma,0\right),\end{aligned} \tag{B.5}$$

and $\partial \;=\; \partial_{\mathbf{u}}$ denotes the subgradient Rockafellar (1971); Bauschke and Combettes (2011), and the absolute value of a vector is computed component-wise. The typically slow convergence of the first-order method (B.4) can be accelerated by an over-relaxation step Nesterov (1983), resulting in the *Fast ISTA* algorithm (FISTA) Beck and Teboulle (2009b):

$$\begin{aligned}\mathbf{y}_{k+1} \;&=\mathbf{u}_k \;-\; \gamma\alpha\mathbf{A}^T\left(\mathbf{A}\mathbf{u}_k-\mathbf{d}\right), \\ \mathbf{z}_{k+1} \;&=\text{shrink}\left\{\mathbf{y}_{k+1},\gamma\right\}, \\ \zeta_{k+1} \;&=\; \left(1+\sqrt{1+4\zeta_k^2}\right)/2, \\ \mathbf{u}_{k+1} \;&=\mathbf{y}_{k+1}+\frac{\zeta_k-1}{\zeta_{k+1}}\left(\mathbf{y}_{k+1}-\mathbf{y}_k\right),\end{aligned} \tag{B.6}$$

where $\zeta_1=1$ and γ is sufficiently small.

It is important to note that algorithm (B.6) is applied to the L_1-regularized problem (B.2), not the TV-regularized problem (C.1). An accelerated algorithm for solving a TV-regularized *denoising problem*[2] was proposed in Beck and Teboulle (2009a) and applied the Nesterov relaxation Nesterov (1983) to solving the dual of the TV-regularized denoising problem Chambolle (2004). However, using a similar approach to solving (C.1) with a non-trivial operator $\mathbf{A}$ results in accelerated schemes that still require inversion of $\mathbf{A}$ Beck and Teboulle (2009a); Goldstein et al. (2014) and thus lack the primary appeal of the accelerated gradient descent methods—i.e., a single application of $\mathbf{A}$ and its transpose per iteration[3].

[2]with $\mathbf{A}=\mathbf{I}$ in (C.1)

[3]In Beck and Teboulle (2009a) inversion of $\mathbf{A}$ is replaced by a single gradient descent, however, over-relaxation is applied to the dual variable.

The advantage of (B.6) compared with simple gradient descent is that Nesterov's over-relaxation step requires storing two previous solution vectors and provides improved search direction for minimization. Note, however, that the step length γ is inversely proportional to the Lipschitz constant of $\alpha \mathbf{A}^T (\mathbf{A}\mathbf{u} - \mathbf{d})$ Beck and Teboulle (2009b) and may be small in practice.

A very general approach to solving problems (B.1) involving either L_1 or TV regularization is provided by primal-dual methods. For example, in TV-regularized least-squares problem (C.1), by substituting

$$\mathbf{z} = \mathbf{B}\mathbf{u} \tag{B.7}$$

and adding (B.7) as a constraint, we obtain an equivalent equality-constrained optimization problem

$$\begin{aligned} &\|\mathbf{z}\|_1 \; + \; \frac{\alpha}{2}\|\mathbf{A}\mathbf{u} - \mathbf{d}\|_2^2 \;\; \rightarrow \;\; \min, \\ &\mathbf{z} \;\; = \;\; \mathbf{B}\mathbf{u}. \end{aligned} \tag{B.8}$$

The optimal solution of (C.5) corresponds to the saddle-point of its Lagrangian

$$L_0 (\mathbf{u}, \mathbf{z}, \boldsymbol{\mu}) \;\; = \;\; \|\mathbf{z}\|_1 + \frac{\alpha}{2}\|\mathbf{A}\mathbf{u} - \mathbf{d}\|_2^2 + \boldsymbol{\mu}^T (\mathbf{z} - \mathbf{B}\mathbf{u}), \tag{B.9}$$

that can be found by the *Uzawa method* Uzawa (1958). The Uzawa method finds the saddle point by alternating a minimization with respect to the primal variables $\mathbf{u}, \mathbf{z}$ and ascent over the dual variable $\boldsymbol{\mu}$ for the objective function equal to the standard Lagrangian (B.9), $L = L_0$,

$$\begin{aligned} (\mathbf{u}_{k+1}, \mathbf{z}_{k+1}) \;\; &= \operatorname{argmin} L (\mathbf{u}, \mathbf{z}, \boldsymbol{\mu}_k), \\ \boldsymbol{\mu}_{k+1} \;\; &= \boldsymbol{\mu}_k \; + \; \lambda [\mathbf{z}_{k+1} - \mathbf{B}\mathbf{u}_{k+1}] \end{aligned} \tag{B.10}$$

for some positive step size λ. Approach (B.10), when applied to the Augmented Lagrangian Rockafellar (1976), $L = L_+$,

$$L_+ (\mathbf{u}, \mathbf{z}, \boldsymbol{\mu}) \;\; = \;\; \|\mathbf{z}\|_1 + \frac{\alpha}{2}\|\mathbf{A}\mathbf{u} - \mathbf{d}\|_2^2 + \boldsymbol{\mu}^T (\mathbf{z} - \mathbf{B}\mathbf{u}) + \frac{\lambda}{2}\|\mathbf{z} - \mathbf{B}\mathbf{u}\|_2^2, \tag{B.11}$$

results in the *method of multipliers* Hestenes (1969). For problems (B.1) all these methods still require joint minimization with respect to $\mathbf{u}$ and $\mathbf{z}$ of some objective function that includes both $\|\mathbf{z}\|_1$ and a smooth function of $\mathbf{u}$. Splitting the joint minimization into separate steps of minimization with respect $\mathbf{u}$, followed by minimization with respect to $\mathbf{z}$, results in the *Alternating-Directions Method of Multipliers* (ADMM) Glowinski and Marroco (1975); Gabay and Mercier (1976); Glowinski and Le Tallec (1989); Eckstein and Bertsekas (1992); Boyd et al. (2011). To establish a connection to the splitting techniques applied to the sum of two operators, we note that the ADMM is equivalent to applying the Douglas-Rachford splitting Douglas and Rachford (1956) to the problem

$$\partial \left[\|\mathbf{B}\mathbf{u}\|_1 \; + \; \frac{\alpha}{2} \|\mathbf{A}\mathbf{u} - \mathbf{d}\|_2^2 \right] \ni \mathbf{0}, \tag{B.12}$$

where ∂ is the subgradient, and problem (B.12) is equivalent to (B.1). The ADMM is a particular case of a primal-dual iterative solution framework with splitting Zhang et al. (2010), where the minimization in (B.10) is split into two steps,

$$\begin{aligned} \mathbf{u}_{k+1} \; &= \operatorname{argmin} L\left(\mathbf{u}, \mathbf{z}_k, \boldsymbol{\mu}_k\right), \\ \mathbf{z}_{k+1} \; &= \operatorname{argmin} L\left(\mathbf{u}_{k+1}, \mathbf{z}, \boldsymbol{\mu}_k\right), \\ \boldsymbol{\mu}_{k+1} \; &= \boldsymbol{\mu}_k \; + \; \lambda \left[\mathbf{z}_{k+1} - \mathbf{B}\mathbf{u}_{k+1}\right] \end{aligned} \tag{B.13}$$

For the ADMM, we substitute $L = L_+$ in (B.13) but other choices of a modified Lagrange function L are possible that may produce convergent primal-dual algorithms Zhang et al. (2010). Making the substitution $L = L_+$ from (B.11) into (B.13), and introducing a scaled vector of multipliers,

$$\mathbf{b}_k \; = \; \boldsymbol{\mu}_k / \lambda, \; k = 0, 1, 2, \ldots \tag{B.14}$$

we obtain

$$\begin{aligned}
\mathbf{u}_{k+1} &= \operatorname{argmin} \frac{\alpha}{2}\|\mathbf{A}\mathbf{u}-\mathbf{d}\|_2^2 + \frac{\lambda}{2}\|\mathbf{z}_k - \mathbf{B}\mathbf{u}+\mathbf{b}_k\|_2^2, \\
\mathbf{z}_{k+1} &= \operatorname{argmin} \|\mathbf{z}\|_1 + \frac{\lambda}{2}\|\mathbf{z} - \mathbf{B}\mathbf{u}_{k+1}+\mathbf{b}_k\|_2^2, \\
\mathbf{b}_{k+1} &= \mathbf{b}_k + \mathbf{z}_{k+1} - \mathbf{B}\mathbf{u}_{k+1},\ k=0,1,2,\ldots
\end{aligned} \tag{B.15}$$

where we used the fact that adding a constant term $\lambda/2\|\mathbf{b}_k\|_2^2$ to the objective function does not alter the solution. In the iterative process (C.10), we apply splitting, minimizing

$$\|\mathbf{z}\|_1 + \frac{\alpha}{2}\|\mathbf{A}\mathbf{u}-\mathbf{d}\|_2^2 + \frac{\lambda}{2}\|\mathbf{z} - \mathbf{B}\mathbf{u}+\mathbf{b}_k\|_2^2 \tag{B.16}$$

alternately with respect to $\mathbf{u}$ and $\mathbf{z}$. Further we note that the minimization of (C.12) with respect to $\mathbf{z}$ (in a splitting step with $\mathbf{u}$ fixed) is given trivially by the shrinkage operator (C.14),

$$\mathbf{z}_{k+1} = \operatorname{shrink}\{\mathbf{B}\mathbf{u}-\mathbf{b}_k, 1/\lambda\}. \tag{B.17}$$

Combining (C.10) and (C.13) we obtain Algorithm 1.

Algorithm 1 Alternating Direction Method of Multipliers (ADMM) for (B.1)

$\mathbf{u}_0 \leftarrow \mathbf{0}^N$, $\mathbf{z}_0^K \leftarrow \mathbf{0}$
$\mathbf{b}_0 \leftarrow \mathbf{0}^K$
for $k \leftarrow 0,1,2,3,\ldots$ **do**
  $\mathbf{u}_{k+1} \leftarrow \operatorname{argmin}\left\{\frac{\lambda}{2}\|\mathbf{z}_k - \mathbf{B}\mathbf{u}+\mathbf{b}_k\|_2^2 + \frac{\alpha}{2}\|\mathbf{A}\mathbf{u}-\mathbf{d}\|_2^2\right\}$
  $\mathbf{z}_{k+1} \leftarrow \operatorname{shrink}\{\mathbf{B}\mathbf{u}_{k+1}-\mathbf{b}_k, 1/\lambda\}$
  $\mathbf{b}_{k+1} \leftarrow \mathbf{b}_k + \mathbf{z}_{k+1} - \mathbf{B}\mathbf{u}_{k+1}$
  Exit loop if $\|\mathbf{u}_{k+1}-\mathbf{u}_k\|_2/\|\mathbf{u}_k\|_2 \leq$ target accuracy
end for

Minimization on the first line of (C.10) at each step of the ADMM requires inversion of the operator $\mathbf{A}$. In the first-order gradient-descent methods like (B.6) a similar requirement is obviated by replacing the minimization with respect to variable $\mathbf{u}$ by gradient descent. However, for ill-conditioned problems the gradient may be a poor approximation to the optimal search direction. One interpretation of Nesterov's over-relaxation step in (B.6) is that it provides a better search direction by perturbing the current solution update with a fraction of the previous update on the last line of

(B.6). The intermediate least-squares problem in (C.10) can be solved approximately using, for example, a few iterations of conjugate gradients. However, repeating multiple iterations of Conjugate Gradients at each step of the ADMM may be unnecessary. Indeed, as we demonstrate in the following sections, conjugate directions constructed at earlier steps of the ADMM can be reused because the matrix of the system of normal equations associated with the minimization on the first line of (C.10) does not change between ADMM steps[4]. Therefore, we can trade the computational cost of applying the operator $\mathbf{A}$ and its transpose against the cost of storing a few solution and data-size vectors. As this approach is applied to the most general problem (B.1) with a non-trivial operator $\mathbf{B}$, in addition to the potential speed-up, this method has the advantage of working equally well for L_1 and TV-regularized problems.

We stress that our new approach does not improve the theoretical convergence properties of the classic ADMM method under the assumption of exact minimization in step 4 of Algorithm 1. The asymptotic convergence rate is still $O(1/k)$ as with exact minimization He and Yuan (2012). The new approach provides a numerically feasible way of implementing the ADMM for problems where a computationally expensive operator $\mathbf{A}$ precludes accurate minimization in step 4. However, the rate of convergence in the general method of multipliers (B.10) is sensitive to the choice of parameter λ, and an improved convergence rate for some values of λ can be accompanied with more ill-conditioned minimization problems at each step of (C.10) Glowinski and Le Tallec (1989). By employing increasingly more accurate conjugate-directions solution of the minimization problem at each iteration of (C.10) the new method offsets the deteriorating condition of the intermediate least-squares problems, and achieves a faster practical convergence at early iterations.

Practical utility of the ADMM in applications that involve sparsity-promoting (B.2) or edge-preserving (C.1) inversion is often determined by how quickly we can resolve sparse or blocky model components. These features can often be *qualitatively* resolved within relatively few initial iterations of the ADMM (see discussion in the appendix of Goldstein and Osher (2009)). In our Section 5, fast recovery of such *local*

[4]Only the right-hand sides of the system are updated as a result of thresholding.

features will be one of the key indicators for judging the efficiency of the proposed method.

In the next section we describe two new algorithms, *Steered* and *Compressive Conjugate Gradients* based on the principle of reusing conjugate directions for multiple right-hand sides. In Section 3 we prove convergence and demonstrate that the new algorithm coincides with the exact ADMM in a finite number of iterations. Section 4 contains a practical implementation of the Compressive Conjugate Gradients method. We test the method on a series of problems from imaging and mechanics, and compare its performance against FISTA and ADMM with gradient descent and restarted conjugate gradients.

STEERED AND COMPRESSIVE CONJUGATE DIRECTIONS

Step 4 of Algorithm 1 is itself a least-squares optimization problem of the form

$$\|\mathbf{F}\mathbf{u} \;-\; \mathbf{v}_k\|_2^2 \;\rightarrow\; \min, \tag{B.18}$$

where

$$\mathbf{F} \;=\; \begin{bmatrix} \sqrt{\alpha}\mathbf{A} \\ \sqrt{\lambda}\mathbf{B} \end{bmatrix} \tag{B.19}$$

and

$$\mathbf{v}_k \;=\; \begin{bmatrix} \sqrt{\alpha}\mathbf{d} \\ \sqrt{\lambda}\,(\mathbf{z}_k + \mathbf{b}_k) \end{bmatrix} \tag{B.20}$$

Solving optimization problem (B.18) is mathematically equivalent to solving the following system of normal equations Trefethen and Bau III (1997),

$$\mathbf{F}^T\mathbf{F}\mathbf{u} \;=\; \mathbf{F}^T\mathbf{v}_k, \tag{B.21}$$

as operator (B.19) has maximum rank. Solving (B.21) has the disadvantage of squaring the condition number of operator (B.19) Trefethen and Bau III (1997). When the operator $\mathbf{A}$ is available in a matrix form, and a factorization of operator $\mathbf{F}$ is numerically feasible, solving the normal equations (B.21) should be avoided and a technique based on a matrix factorization should be applied directly to solving (B.18) Björk (1996); Saad (2003). However, when matrix $\mathbf{A}$ is not known explicitly or its size exceeds practical limitations of direct methods, as is the case in applications of greatest interest for us, an iterative algorithm, such as the Conjugate Gradients for Normal Equations (CGNE) Björk (1996); Saad (2003), can be used to solve (B.21). Solving (B.18) exactly may be unnecessary and we can expect that for large-scale problems only a few steps of an iterative method need be carried out. However, every iteration typically requires the application of operator $\mathbf{A}$ and its adjoint, and in large-scale optimization problems we are interested in minimizing the number of applications of these operations. For large-scale optimization problems we need an alternative to re-starting an iterative solver for each intermediate problem (B.18). We propose to minimize restarting iterations[5] by devising a conjugate-directions technique for solving (B.18) with a non-stationary right-hand side. At each iteration of the proposed algorithm we find a search direction that is conjugate to previous directions with respect to the operator $\mathbf{F}^T\mathbf{F}$. In the existing conjugate direction techniques, iteratively constructed conjugate directions span the Krylov subspaces Trefethen and Bau III (1997),

$$\mathcal{K}_k \;=\; \text{span}\left\{\mathbf{F}^T\mathbf{v}_0, \left(\mathbf{F}^T\mathbf{F}\right)\mathbf{F}^T\mathbf{v}_0, \ldots, \left(\mathbf{F}^T\mathbf{F}\right)^k\mathbf{F}^T\mathbf{v}_0\right\},\; k=0,1,\ldots\;. \tag{B.22}$$

However, in our approach we construct a sequence of vectors (search directions) that are conjugate with respect to operator $\mathbf{F}^T\mathbf{F}$ at the kth step but may not span the Krylov subspace $\mathcal{K}_k$. This complicates convergence analysis of our technique, but allows "steering" search directions by iteration-dependent right-hand sides. Since the right-hand side in (B.18) is the result of the shrinkage (C.13) at previous iterations that steer or compress the solution, we call our approach "steered" or "compressive"

[5] avoiding restarting altogether in the theoretical limit of infinite computer storage

conjugate directions.

For the least-squares problem (B.18), we construct two sets of vectors for $k = 0, 1, 2, \ldots$

$$\begin{aligned} &\{\mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_k\}, \ \{\mathbf{q}_0, \mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_k\}, \\ &\mathbf{q}_i \ = \ \mathbf{F}\mathbf{p}_i, \ i = 0, 1, 2, \ldots, k, \end{aligned} \tag{B.23}$$

such that

$$\mathbf{q}_i^T \mathbf{q}_j \ = \ \mathbf{p}_i^T \mathbf{F}^T \mathbf{F} \mathbf{p}_j \ = \ 0 \text{ if } i \ \neq \ j. \tag{B.24}$$

Equations (B.23) and (B.24) mean that the vectors $\mathbf{p}_i$ form *conjugate directions* Trefethen and Bau III (1997); Saad (2003). At each iteration we find an approximation $\mathbf{u}_k$ to the solution of (B.18) as a linear combination of vectors $\mathbf{p}_i, i = 0, 1, \ldots, k$, for which the residual

$$\mathbf{r}_{k+1} \ = \ \mathbf{v}_{k+1} \ - \ \mathbf{F}\mathbf{u}_{k+1}, \tag{B.25}$$

is orthogonal to vectors $\mathbf{q}_i$,

$$\mathbf{q}_i^T \mathbf{r}_{k+1} \ = \ \mathbf{q}_i^T \left(\mathbf{v}_{k+1} \ - \ \mathbf{F}\mathbf{u}_{k+1}\right) \ = \ 0, \ i = 0, 1, \ldots, k. \tag{B.26}$$

Vector $\mathbf{p}_k$ is constructed as a linear combination of *all* previous vectors $\mathbf{p}_i, i = 0, 1, \ldots, k$ and $\mathbf{F}^T \mathbf{r}_k$ so that the conjugacy condition in (B.23) is satisfied. The resulting algorithm for *arbitrary* $\mathbf{v}_k$ depending on k is given by Algorithm 2.

Note that the above algorithm is not specific to a particular sequence of right-hand-side vectors $\mathbf{v}_k$ and its applicability goes beyond solving the constrained optimization problems (C.5). The algorithm requires storing $2k + 2$ vectors (B.23), as well as one vector each for the current solution iterate $\mathbf{u}_k$, variable right-hand side $\mathbf{v}_k$, intermediate vectors $\mathbf{w}_k$ and $\mathbf{s}_k$. The requirement of storing a growing number of vectors makes the algorithm resemble the GMRES method Saad (2003) for solving linear systems with non-self-adjoint operators. However, in our case, this is a consequence of having a variable right-hand side, requiring re-computation of solution iterates as linear combinations of all of the previous search directions (B.23). This requirement can be relaxed in applications where vector $\mathbf{v}_k$ is updated, for example, by the modified

Algorithm 2 Steered Conjugate Directions for solving (B.18)

$\mathbf{u}_0 \leftarrow \mathbf{0}^N$
$\mathbf{p}_0 \leftarrow \mathbf{F}^T\mathbf{v}_0,\ \mathbf{q}_0 \leftarrow \mathbf{F}\mathbf{p}_0,\ \delta_0 \leftarrow \mathbf{q}_0^T\mathbf{q}_0$
for $k = 0, 1, 2, 3, \ldots$ **do**
    for $i = 0, 1, \ldots, k$ **do**
        $\tau_i \leftarrow \mathbf{q}_i^T\mathbf{v}_k/\delta_i$
    end for
    $\mathbf{u}_{k+1} \leftarrow \sum_{i=0}^{k} \tau_i\mathbf{p}_i$
    $\mathbf{r}_{k+1} \leftarrow \mathbf{v}_{k+1} - \sum_{i=0}^{k} \tau_i\mathbf{q}_i$
    $\mathbf{w}_{k+1} \leftarrow \mathbf{F}^T\mathbf{r}_{k+1}$
    $\mathbf{s}_{k+1} \leftarrow \mathbf{F}\mathbf{w}_{k+1}$
    for $i = 0, 1, \ldots, k$ **do**
        $\beta_i \leftarrow -\mathbf{q}_i^T\mathbf{s}_{k+1}/\delta_i$
    end for
    $\mathbf{p}_{k+1} \leftarrow \sum_{i=0}^{k} \beta_i\mathbf{p}_i + \mathbf{w}_{k+1}$
    $\mathbf{q}_{k+1} \leftarrow \sum_{i=0}^{k} \beta_i\mathbf{q}_i + \mathbf{s}_{k+1}$
    $\delta_{k+1} \leftarrow \mathbf{q}_{k+1}^T\mathbf{q}_{k+1}$
    if $\delta_{k+1} = 0$ **then** ▷ Use condition "$\delta_{k+1} <$ tolerance" in practice
        $\delta_{k+1} \leftarrow 1,\ \mathbf{p}_{k+1} \leftarrow \mathbf{0}^N,\ \mathbf{q}_{k+1} \leftarrow \mathbf{0}^{M+K}$
    end if
    Exit loop if $\|\mathbf{u}_{k+1} - \mathbf{u}_k\|_2/\|\mathbf{u}_k\|_2 \leq$ target accuracy
end for

Lagrangian technique for solving a constrained optimization problem, and converges to a limit. In Section 4 we describe practical applications of the algorithm achieving fast convergence while storing only a subset of vectors (B.23). The algorithm requires one application of $\mathbf{F}$ and its transpose at each iteration and $2k + 3$ dot-products of large vectors.

Combining Algorithms 1 and 2 we obtain the *Compressive Conjugate Directions* Algorithm 3.

Algorithm 3 Compressive Conjugate Directions for (B.1)

$\mathbf{u}_0 \leftarrow \mathbf{0}^N,\ \mathbf{z}_0 \leftarrow \mathbf{0}^K;\ \mathbf{b}_0 \leftarrow \mathbf{0}^K,\ \mathbf{v}_0 \leftarrow \begin{bmatrix} \sqrt{\alpha}\mathbf{d} \\ \sqrt{\lambda}\,(\mathbf{z}_0 + \mathbf{b}_0) \end{bmatrix}$
$\mathbf{p}_0 \leftarrow \mathbf{F}^T\mathbf{v}_0,\ \mathbf{q}_0 \leftarrow \mathbf{F}\mathbf{p}_0,\ \delta_0 \leftarrow \mathbf{q}_0^T\mathbf{q}_0$
for $k = 0, 1, 2, 3, \ldots$ **do**
  for $i = 0, 1, \ldots, k$ **do**
    $\tau_i \leftarrow \mathbf{q}_i^T\mathbf{v}_k/\delta_i$
  end for
  $\mathbf{u}_{k+1} \leftarrow \sum_{i=0}^{k} \tau_i\mathbf{p}_i$
  $\mathbf{z}_{k+1} \leftarrow \text{shrink}\,\{\mathbf{B}\mathbf{u}_{k+1} - \mathbf{b}_k, 1/\lambda\}$
  $\mathbf{b}_{k+1} \leftarrow \mathbf{b}_k + \mathbf{z}_{k+1} - \mathbf{B}\mathbf{u}_{k+1}$
  $\mathbf{v}_{k+1} \leftarrow \begin{bmatrix} \sqrt{\alpha}\mathbf{d} \\ \sqrt{\lambda}\,(\mathbf{z}_{k+1} + \mathbf{b}_{k+1}) \end{bmatrix}$
  $\mathbf{r}_{k+1} \leftarrow \mathbf{v}_{k+1} - \sum_{i=0}^{k} \tau_i\mathbf{q}_i$
  $\mathbf{w}_{k+1} \leftarrow \mathbf{F}^T\mathbf{r}_{k+1}$
  $\mathbf{s}_{k+1} \leftarrow \mathbf{F}\mathbf{w}_{k+1}$
  for $i = 0, 1, \ldots, k$ **do**
    $\beta_i \leftarrow -\mathbf{q}_i^T\mathbf{s}_{k+1}/\delta_i$
  end for
  $\mathbf{p}_{k+1} \leftarrow \sum_{i=0}^{k} \beta_i\mathbf{p}_i + \mathbf{w}_{k+1}$
  $\mathbf{q}_{k+1} \leftarrow \sum_{i=0}^{k} \beta_i\mathbf{q}_i + \mathbf{s}_{k+1}$
  $\delta_{k+1} \leftarrow \mathbf{q}_{k+1}^T\mathbf{q}_{k+1}$
  if $\delta_{k+1} = 0$ **then** ▷ Use condition "$\delta_{k+1} <$ tolerance" in practice
    $\delta_{k+1} \leftarrow 1,\ \mathbf{p}_{k+1} \leftarrow \mathbf{0}^N,\ \mathbf{q}_{k+1} \leftarrow \mathbf{0}^{M+K}$
  end if
  Exit loop if $\|\mathbf{u}_{k+1} - \mathbf{u}_k\|_2/\|\mathbf{u}_k\|_2 \leq$ target accuracy
end for

CONVERGENCE ANALYSIS

Convergence properties of the ADMM were studied in many publications and are well known. However, here we provide a self-contained proof of convergence for Algorithm 1 that mostly follows the presentation of Boyd et al. (2011). Later, we use this result to study the convergence of Algorithm 3.

Theorem 1. *Assume that $M \geq N$, operators $\mathbf{A}$, $\mathbf{B}$ are maximum rank, and*

$$\begin{aligned} \mathbf{u} &= \mathbf{u}^*, \\ \mathbf{z} &= \mathbf{z}^* = \mathbf{B}\mathbf{u}^*, \end{aligned} \tag{B.27}$$

is the unique solution of problem (C.5). Assume that a vector $\mathbf{b}^*$ is defined as

$$\mathbf{b}^* = \boldsymbol{\mu}^*/\lambda, \tag{B.28}$$

where $\boldsymbol{\mu}^*$ is the vector of Lagrange multipliers for the equality constraint in (C.5). Algorithm 1 then converges to this solution if $\lambda > 0$, that is,

$$\mathbf{u}_k \to \mathbf{u}^*,\ \mathbf{z}_k \to \mathbf{z}^*,\ \mathbf{b}_k \to \mathbf{b}^*,\ k \to \infty. \tag{B.29}$$

Proof. Problem (C.5) has a convex objective function and equality constraints, hence (B.27,B.28) is a saddle point of its Lagrangian (B.9) Boyd and Vandenberghe (2004). Substituting $\mathbf{z}_{k+1}, \mathbf{u}_{k+1}$ from Algorithm 1, we have

$$\begin{aligned} & L_0\left(\mathbf{z}^*, \mathbf{u}^*, \boldsymbol{\mu}^*\right) \leq L_0\left(\mathbf{z}_{k+1}, \mathbf{u}_{k+1}, \boldsymbol{\mu}^*\right) \Longleftrightarrow \\ p^* = & \|\mathbf{B}\mathbf{u}^*\|_1 + \frac{\alpha}{2}\|\mathbf{A}\mathbf{u}^* - \mathbf{d}\|_2^2 = \|\mathbf{z}^*\|_1 + \frac{\alpha}{2}\|\mathbf{A}\mathbf{u}^* - \mathbf{d}\|_2^2 \leq \\ & \|\mathbf{z}_{k+1}\|_1 + \frac{\alpha}{2}\|\mathbf{A}\mathbf{u}_{k+1} - \mathbf{d}\|_2^2 + \boldsymbol{\mu}^{*T}\left(\mathbf{z}_{k+1} - \mathbf{B}\mathbf{u}_{k+1}\right) = \\ & p_{k+1} + \boldsymbol{\mu}^{*T}\left(\mathbf{z}_{k+1} - \mathbf{B}\mathbf{u}_{k+1}\right) = p_{k+1} + \lambda \mathbf{b}^{*T}\left(\mathbf{z}_{k+1} - \mathbf{B}\mathbf{u}_{k+1}\right), \end{aligned} \tag{B.30}$$

where p^* is the optimal value of the objective function and p_{k+1} is its approximation at iteration k of the algorithm. Inequality (B.30) provides a lower bound for the objective function estimate p_{k+1}. Step 4 of the algorithm is equivalent to

$$\alpha \mathbf{A}^T\mathbf{A}\mathbf{u}_{k+1} + \lambda \mathbf{B}^T\mathbf{B}\mathbf{u}_{k+1} = \alpha \mathbf{A}^T\mathbf{d} + \lambda \mathbf{B}^T\left(\mathbf{z}_k + \mathbf{b}_k\right). \tag{B.31}$$

Substituting the expression for $\mathbf{b}_k$ from steps 6 into (B.31), we obtain

$$\alpha \mathbf{A}^T\mathbf{A}\mathbf{u}_{k+1} = \alpha \mathbf{A}^T\mathbf{d} + \lambda \mathbf{B}^T\left(\mathbf{z}_k - \mathbf{z}_{k+1} + \mathbf{b}_{k+1}\right). \tag{B.32}$$

Equality (B.32) is equivalent to

$$\mathbf{u}_{k+1} \;=\; \operatorname{argmin} \frac{\alpha}{2}\|\mathbf{A}\mathbf{u}-\mathbf{d}\|_2^2 \;-\; \lambda\left(\mathbf{z}_k-\mathbf{z}_{k+1}+\mathbf{b}_{k+1}\right)^T\mathbf{B}\mathbf{u}. \tag{B.33}$$

Substituting $\mathbf{u}_{k+1}$ and $\mathbf{u}^*$ into the right-hand side of (B.33), we obtain

$$\begin{aligned}\frac{\alpha}{2}\|\mathbf{A}\mathbf{u}_{k+1}-\mathbf{d}\|_2^2 \;\le\; &\frac{\alpha}{2}\|\mathbf{A}\mathbf{u}^*-\mathbf{d}\|_2^2 + \\ &\lambda\left(\mathbf{z}_k-\mathbf{z}_{k+1}+\mathbf{b}_{k+1}\right)^T\mathbf{B}\left(\mathbf{u}_{k+1}-\mathbf{u}^*\right).\end{aligned} \tag{B.34}$$

Step 5 is equivalent to

$$\begin{aligned}\mathbf{0} \;\in\; \partial_{\mathbf{z}}\|\mathbf{z}\|_1 \;+\; \lambda\left(\mathbf{z}_{k+1}-\mathbf{B}\mathbf{u}_{k+1}+\mathbf{b}_k\right) \;=\; \partial_{\mathbf{z}}\|\mathbf{z}\|_1 \;+\; \lambda b_{k+1}, \\ \mathbf{z}_{k+1} \;=\operatorname{argmin}\left\{\|\mathbf{z}\|_1 \;+\; \lambda\mathbf{b}_{k+1}^T\mathbf{z}\right\},\end{aligned} \tag{B.35}$$

where we used the expression for $\mathbf{b}_k$ from step 6. Substituting $\mathbf{z}=\mathbf{z}_{k+1}$ and $\mathbf{z}=\mathbf{z}^*$ into the right-hand side of the second line of (B.35), we obtain

$$\|\mathbf{z}_{k+1}\|_1 \;\le\; \|\mathbf{z}^*\|_1 \;+\; \lambda\mathbf{b}_{k+1}^T\left(\mathbf{z}^*-\mathbf{z}_{k+1}\right). \tag{B.36}$$

Adding (B.34) and (B.36), we get

$$\begin{aligned}p_{k+1} \;\le p^* \;+\; &\lambda\mathbf{b}_{k+1}^T\left(\mathbf{z}^*-\mathbf{z}_{k+1}\right)+ \\ &\lambda\left(\mathbf{z}_k-\mathbf{z}_{k+1}+\mathbf{b}_{k+1}\right)^T\mathbf{B}\left(\mathbf{u}_{k+1}-\mathbf{u}^*\right),\end{aligned} \tag{B.37}$$

an upper bound for p_{k+1}. Adding (B.30) and (B.37), we get

$$\begin{aligned}0 \;\le \lambda\mathbf{b}^{*T}\left(\mathbf{z}_{k+1}-\mathbf{B}\mathbf{u}_{k+1}\right) \;+\; &\lambda\mathbf{b}_{k+1}^T\left(\mathbf{z}^*-\mathbf{z}_{k+1}\right)+ \\ &\lambda\left(\mathbf{z}_k-\mathbf{z}_{k+1}+\mathbf{b}_{k+1}\right)^T\mathbf{B}\left(\mathbf{u}_{k+1}-\mathbf{u}^*\right),\end{aligned} \tag{B.38}$$

or after rearranging,

$$\begin{aligned}0 \;\le \lambda\left(\mathbf{b}^*-\mathbf{b}_{k+1}\right)^T\left(\mathbf{z}_{k+1}-\mathbf{B}\mathbf{u}_{k+1}\right) \;-\; &\lambda\left(\mathbf{z}_k-\mathbf{z}_{k+1}\right)^T\left(\mathbf{z}_{k+1}-\mathbf{B}\mathbf{u}_{k+1}\right) \;+ \\ &\lambda\left(\mathbf{z}_k-\mathbf{z}_{k+1}\right)^T\left(\mathbf{z}_{k+1}-\mathbf{z}^*\right).\end{aligned} \tag{B.39}$$

We will now use (B.39) to derive an upper estimate for

$$\|\mathbf{b}_k - \mathbf{b}^*\|_2^2 \; + \; \|\mathbf{z}_k - \mathbf{z}^*\|_2^2.$$

Using step 6 of Algorithm 1 for the first term in (B.39) and introducing $\boldsymbol{\rho}_{k+1} = \mathbf{z}_{k+1} - \mathbf{B}\mathbf{u}_{k+1}$, we get

$$\begin{aligned}
&\lambda\left(\mathbf{b}^* - \mathbf{b}_{k+1}\right)^T \boldsymbol{\rho}_{k+1} \; = \\
&\lambda\left(\mathbf{b}^* - \mathbf{b}_k - \boldsymbol{\rho}_{k+1}\right)^T \boldsymbol{\rho}_{k+1} \; = \; \lambda\left(\mathbf{b}^* - \mathbf{b}_k\right)^T \boldsymbol{\rho}_{k+1} - \lambda\|\boldsymbol{\rho}_{k+1}\|_2^2 \; = \\
&\lambda\left(\mathbf{b}^* - \mathbf{b}_k\right)^T \left(\mathbf{b}_{k+1} - \mathbf{b}_k\right) - \frac{\lambda}{2}\|\boldsymbol{\rho}_{k+1}\|_2^2 - \frac{\lambda}{2}\|\boldsymbol{\rho}_{k+1}\|_2^2 \; = \\
&\lambda\left(\mathbf{b}^* - \mathbf{b}_k\right)^T \left(\mathbf{b}_{k+1} - \mathbf{b}_k\right) - \frac{\lambda}{2}\|\boldsymbol{\rho}_{k+1}\|_2^2 - \frac{\lambda}{2}\left(\mathbf{b}_{k+1} - \mathbf{b}_k\right)^T \left(\mathbf{b}_{k+1} - \mathbf{b}_k\right) \; = \\
&-\lambda\left(\mathbf{b}_k - \mathbf{b}^*\right)^T \left[\left(\mathbf{b}_{k+1} - \mathbf{b}^*\right) - \left(\mathbf{b}_k - \mathbf{b}^*\right)\right] - \frac{\lambda}{2}\|\boldsymbol{\rho}_{k+1}\|_2^2 - \\
&\frac{\lambda}{2}\left[\left(\mathbf{b}_{k+1} - \mathbf{b}^*\right) - \left(\mathbf{b}_k - \mathbf{b}^*\right)\right]^T \left[\left(\mathbf{b}_{k+1} - \mathbf{b}^*\right) - \left(\mathbf{b}_k - \mathbf{b}^*\right)\right] \; = \\
&\frac{\lambda}{2}\|\mathbf{b}_k - \mathbf{b}^*\|_2^2 - \frac{\lambda}{2}\|\mathbf{b}_{k+1} - \mathbf{b}^*\|_2^2 - \frac{\lambda}{2}\|\boldsymbol{\rho}_{k+1}\|_2^2.
\end{aligned} \tag{B.40}$$

Substituting (B.40) into (B.39), we obtain

$$\begin{aligned}
0 \; \leq \; &\frac{\lambda}{2}\|\mathbf{b}_k - \mathbf{b}^*\|_2^2 - \frac{\lambda}{2}\|\mathbf{b}_{k+1} - \mathbf{b}^*\|_2^2 - \frac{\lambda}{2}\|\boldsymbol{\rho}_{k+1}\|_2^2 \; - \; \lambda\left(\mathbf{z}_k - \mathbf{z}_{k+1}\right)^T \boldsymbol{\rho}_{k+1} + \\
&\lambda\left(\mathbf{z}_k - \mathbf{z}_{k+1}\right)^T \left(\mathbf{z}_{k+1} - \mathbf{z}^*\right) \; = \\
&\frac{\lambda}{2}\|\mathbf{b}_k - \mathbf{b}^*\|_2^2 - \frac{\lambda}{2}\|\mathbf{b}_{k+1} - \mathbf{b}^*\|_2^2 - \frac{\lambda}{2}\|\boldsymbol{\rho}_{k+1}\|_2^2 \; - \; \lambda\left(\mathbf{z}_k - \mathbf{z}_{k+1}\right)^T \boldsymbol{\rho}_{k+1} + \\
&\lambda\left(\mathbf{z}_k - \mathbf{z}_{k+1}\right)^T \left[\left(\mathbf{z}_{k+1} - \mathbf{z}_k\right) + \left(\mathbf{z}_k - \mathbf{z}^*\right)\right] \; = \\
&\frac{\lambda}{2}\|\mathbf{b}_k - \mathbf{b}^*\|_2^2 - \frac{\lambda}{2}\|\mathbf{b}_{k+1} - \mathbf{b}^*\|_2^2 - \frac{\lambda}{2}\|\boldsymbol{\rho}_{k+1}\|_2^2 \; - \; \lambda\left(\mathbf{z}_k - \mathbf{z}_{k+1}\right)^T \boldsymbol{\rho}_{k+1} - \\
&\lambda\left(\mathbf{z}_k - \mathbf{z}_{k+1}\right)^T \left(\mathbf{z}_k - \mathbf{z}_{k+1}\right) + \lambda\left(\mathbf{z}_k - \mathbf{z}_{k+1}\right)^T \left(\mathbf{z}_k - \mathbf{z}^*\right) \; = \\
&\frac{\lambda}{2}\|\mathbf{b}_k - \mathbf{b}^*\|_2^2 - \frac{\lambda}{2}\|\mathbf{b}_{k+1} - \mathbf{b}^*\|_2^2 - \frac{\lambda}{2}\left(\mathbf{z}_k - \mathbf{z}_{k+1} + \boldsymbol{\rho}_{k+1}\right)^T \left(\mathbf{z}_k - \mathbf{z}_{k+1} + \boldsymbol{\rho}_{k+1}\right) \; - \\
&\frac{\lambda}{2}\|\mathbf{z}_k - \mathbf{z}_{k+1}\|_2^2 + \lambda\left(\mathbf{z}_k - \mathbf{z}_{k+1}\right)^T \left(\mathbf{z}_k - \mathbf{z}^*\right) \; =
\end{aligned}$$

$$
\begin{aligned}
&\frac{\lambda}{2}\|\mathbf{b}_k-\mathbf{b}^*\|_2^2-\frac{\lambda}{2}\|\mathbf{b}_{k+1}-\mathbf{b}^*\|_2^2-\frac{\lambda}{2}\|\mathbf{z}_k-\mathbf{z}_{k+1}+\boldsymbol{\rho}_{k+1}\|_2^2-\frac{\lambda}{2}\|\mathbf{z}_k-\mathbf{z}_{k+1}\|_2^2+\\
&\lambda\left[(\mathbf{z}_k-\mathbf{z}^*)-(\mathbf{z}_{k+1}-\mathbf{z}^*)\right]^T(\mathbf{z}_k-\mathbf{z}^*)\;=\\
&\frac{\lambda}{2}\|\mathbf{b}_k-\mathbf{b}^*\|_2^2-\frac{\lambda}{2}\|\mathbf{b}_{k+1}-\mathbf{b}^*\|_2^2-\frac{\lambda}{2}\|\mathbf{z}_k-\mathbf{z}_{k+1}+\boldsymbol{\rho}_{k+1}\|_2^2-\\
&\frac{\lambda}{2}\|\left(\mathbf{z}_k-\mathbf{z}^*\right)-\left(\mathbf{z}_{k+1}-\mathbf{z}^*\right)\|_2^2+\lambda\left[(\mathbf{z}_k-\mathbf{z}^*)-(\mathbf{z}_{k+1}-\mathbf{z}^*)\right]^T(\mathbf{z}_k-\mathbf{z}^*)\;=\\
&\frac{\lambda}{2}\|\mathbf{b}_k-\mathbf{b}^*\|_2^2-\frac{\lambda}{2}\|\mathbf{b}_{k+1}-\mathbf{b}^*\|_2^2-\frac{\lambda}{2}\|\mathbf{z}_k-\mathbf{z}_{k+1}+\boldsymbol{\rho}_{k+1}\|_2^2-\\
&\frac{\lambda}{2}\|\mathbf{z}_{k+1}-\mathbf{z}^*\|_2^2+\frac{\lambda}{2}\|\mathbf{z}_k-\mathbf{z}^*\|_2^2,
\end{aligned}
\tag{B.41}
$$

yielding

$$
\begin{aligned}
&\frac{\lambda}{2}\|\mathbf{z}_k-\mathbf{z}_{k+1}+\boldsymbol{\rho}_{k+1}\|_2^2\;\leq\\
&\frac{\lambda}{2}\left(\|\mathbf{z}_k-\mathbf{z}^*\|_2^2+\|\mathbf{b}_k-\mathbf{b}^*\|_2^2\right)-\frac{\lambda}{2}\left(\|\mathbf{z}_{k+1}-\mathbf{z}^*\|_2^2+\|\mathbf{b}_{k+1}-\mathbf{b}^*\|_2^2\right).
\end{aligned}
\tag{B.42}
$$

Expanding the left-hand side of (B.42), we obtain

$$
\begin{aligned}
&\frac{\lambda}{2}\left(\|\mathbf{z}_k-\mathbf{z}_{k+1}\|_2^2+2\left(\mathbf{z}_k-\mathbf{z}_{k+1}\right)^T\boldsymbol{\rho}_{k+1}+\|\boldsymbol{\rho}_{k+1}\|_2^2\right)\;\leq\\
&\frac{\lambda}{2}\left(\|\mathbf{z}_k-\mathbf{z}^*\|_2^2+\|\mathbf{b}_k-\mathbf{b}^*\|_2^2\right)-\frac{\lambda}{2}\left(\|\mathbf{z}_{k+1}-\mathbf{z}^*\|_2^2+\|\mathbf{b}_{k+1}-\mathbf{b}^*\|_2^2\right).
\end{aligned}
\tag{B.43}
$$

Let us prove that the middle term in the left-hand side of (B.43) is non-negatve,

$$
0\;\leq\;\left(\mathbf{z}_k-\mathbf{z}_{k+1}\right)^T\boldsymbol{\rho}_{k+1}\;=\;\left(\mathbf{z}_k-\mathbf{z}_{k+1}\right)^T\left(\mathbf{b}_{k+1}-\mathbf{b}_k\right)
$$

where we used step 6 of Algorithm 1. Indeed, since $\mathbf{z}_{k+1}$ minimizes (C.12) with $\mathbf{u}=\mathbf{u}_{k+1}$, using the convexity of L_1 norm, we have for $\mathbf{z}=\mathbf{z}_{k+1}$,

$$
\begin{aligned}
&\partial_z\frac{\lambda}{2}\|\mathbf{z}-\mathbf{B}\mathbf{u}_{k+1}+\mathbf{b}_k\|_2^2=\lambda\left(\mathbf{z}-\mathbf{B}\mathbf{u}_{k+1}+\mathbf{b}_k\right)\in-\partial\|\mathbf{z}\|_1\;\Rightarrow\\
&\|\mathbf{z}_{k+1}\|_1-\|\mathbf{z}_k\|_1\;\leq\;\left(\mathbf{z}_k-\mathbf{z}_{k+1}\right)^T\left(\mathbf{z}_{k+1}-\mathbf{B}\mathbf{u}_{k+1}+\mathbf{b}_k\right)=\left(\mathbf{z}_k-\mathbf{z}_{k+1}\right)^T\mathbf{b}_{k+1}.
\end{aligned}
\tag{B.44}
$$

Similarly, since $\mathbf{z}_k$ minimizes (C.12) for $\mathbf{u} = \mathbf{u}_k$ and $\mathbf{b} = \mathbf{b}_{k-1}$, for $\mathbf{z} = \mathbf{z}_k$ we have

$$\begin{aligned} &\partial_z \frac{\lambda}{2}\|\mathbf{z} - \mathbf{B}\mathbf{u}_k + \mathbf{b}_{k-1}\|_2^2 = \lambda\left(\mathbf{z} - \mathbf{B}\mathbf{u}_k + \mathbf{b}_{k-1}\right) \in -\partial\|\mathbf{z}\|_1 \;\Rightarrow \\ &\|\mathbf{z}_k\|_1 - \|\mathbf{z}_{k+1}\|_1 \;\leq\; \left(\mathbf{z}_{k+1} - \mathbf{z}_k\right)^T\left(\mathbf{z}_k - \mathbf{B}\mathbf{u}_k + \mathbf{b}_{k-1}\right) = \left(\mathbf{z}_{k+1} - \mathbf{z}_k\right)^T \mathbf{b}_k. \end{aligned} \tag{B.45}$$

In both (B.44) and (B.45) we used step 6 of Algorithm 1 and the fact that for any convex function $f(\mathbf{x})$

$$f(\mathbf{x}_0) + \boldsymbol{\xi}^T\left(\mathbf{x} - \mathbf{x}_0\right) \;\leq\; f(\mathbf{x}) \;\Leftrightarrow\; f(\mathbf{x}_0) - f(\mathbf{x}) \;\leq\; -\boldsymbol{\xi}^T\left(\mathbf{x} - \mathbf{x}_0\right), \;\text{ if } \boldsymbol{\xi} \in \partial f(\mathbf{x}_0),$$

where ∂ is subgradient Rockafellar (1971). Summing (B.44) and (B.45) we get

$$0 \;\leq\; \left(\mathbf{z}_k - \mathbf{z}_{k+1}\right)^T\left(\mathbf{b}_{k+1} - \mathbf{b}_k\right). \tag{B.46}$$

From (B.46) and (B.43), we have

$$\begin{aligned} &\|\mathbf{z}_k - \mathbf{z}_{k+1}\|_2^2 + \|\boldsymbol{\rho}_{k+1}\|_2^2 \;\leq \\ &\left(\|\mathbf{z}_k - \mathbf{z}^*\|_2^2 + \|\mathbf{b}_k - \mathbf{b}^*\|_2^2\right) - \left(\|\mathbf{z}_{k+1} - \mathbf{z}^*\|_2^2 + \|\mathbf{b}_{k+1} - \mathbf{b}^*\|_2^2\right), \end{aligned} \tag{B.47}$$

or

$$\begin{aligned} &\|\mathbf{z}_{k+1} - \mathbf{z}^*\|_2^2 + \|\mathbf{b}_{k+1} - \mathbf{b}^*\|_2^2 \;\leq \\ &\|\mathbf{z}_k - \mathbf{z}^*\|_2^2 + \|\mathbf{b}_k - \mathbf{b}^*\|_2^2 - \|\mathbf{z}_{k+1} - \mathbf{z}_k\|_2^2 - \|\boldsymbol{\rho}_{k+1}\|_2^2. \end{aligned} \tag{B.48}$$

From (B.48) we can see that the sequence $\|\mathbf{z}_k - \mathbf{z}^*\|_2^2 + \|\mathbf{b}_k - \mathbf{b}^*\|_2^2$ and consequently $\mathbf{z}_k$ and $\mathbf{b}_k$ are bounded. Summing (B.47) for $k = 0, 1, \ldots, \infty$, we obtain convergence of the series

$$\sum_{k=0}^{\infty}\left\{\|\mathbf{z}_k - \mathbf{z}_{k+1}\|_2^2 + \|\boldsymbol{\rho}_{k+1}\|_2^2\right\} \;\leq\; \|\mathbf{z}_0 - \mathbf{z}^*\|_2^2 + \|\mathbf{b}_0 - \mathbf{b}^*\|_2^2. \tag{B.49}$$

From (B.49) follows

$$\mathbf{z}_k - \mathbf{z}_{k+1} \to 0, \;\; \mathbf{z}_k - \mathbf{B}\mathbf{u}_k \to 0, \; k \to \infty. \tag{B.50}$$

Now using (B.37) we obtain

$$\begin{aligned} p_{k+1} \;-\; p^* \;\leq\; & \lambda \mathbf{b}_{k+1}^T \left(\mathbf{z}^* - \mathbf{z}_{k+1}\right) + \lambda \left(\mathbf{z}_k - \mathbf{z}_{k+1} + \mathbf{b}_{k+1}\right)^T \mathbf{B} \left(\mathbf{u}_{k+1} - \mathbf{u}^*\right) \;= \\ & \lambda \mathbf{b}_{k+1}^T \left(\mathbf{z}_k - \mathbf{z}_{k+1}\right) + \lambda \mathbf{b}_{k+1}^T \left(\mathbf{z}^* - \mathbf{z}_k\right) \;+ \\ & \lambda \left(\mathbf{z}_k - \mathbf{z}_{k+1}\right)^T \mathbf{B} \left(\mathbf{u}_{k+1} - \mathbf{u}^*\right) + \lambda \mathbf{b}_{k+1}^T \mathbf{B} \left(\mathbf{u}_{k+1} - \mathbf{u}^*\right) \;= \\ & \lambda \mathbf{b}_{k+1}^T \left(\mathbf{z}_k - \mathbf{z}_{k+1}\right) \;+\; \lambda \left(\mathbf{z}_k - \mathbf{z}_{k+1}\right)^T \mathbf{B} \left(\mathbf{u}_{k+1} - \mathbf{u}^*\right) \;+ \\ & \lambda \mathbf{b}_{k+1}^T \left(\mathbf{z}^* - \mathbf{z}_k\right) \;+\; \lambda \mathbf{b}_{k+1}^T \mathbf{B} \left(\mathbf{u}_{k+1} - \mathbf{u}^*\right) \;= \\ & \lambda \mathbf{b}_{k+1}^T \left(\mathbf{z}_k - \mathbf{z}_{k+1}\right) \;+\; \lambda \left(\mathbf{z}_k - \mathbf{z}_{k+1}\right)^T \mathbf{B} \left(\mathbf{u}_{k+1} - \mathbf{u}^*\right) \;+ \\ & \lambda \mathbf{b}_{k+1}^T \left(\mathbf{B}\mathbf{u}_{k+1} - \mathbf{z}_{k+1} + \mathbf{z}_{k+1} - \mathbf{z}_k + \mathbf{z}^* - \mathbf{B}\mathbf{u}^*\right) \;\to\; 0,\; k \to \infty, \end{aligned} \tag{B.51}$$

where the right-hand side of (B.51) converges to zero because of (B.50), boundedness of $\mathbf{z}_k$ and $\mathbf{b}_k$ and $\mathbf{z}^* = \mathbf{B}\mathbf{u}^*$. Likewise, from (B.30) we have

$$p^* - p_{k+1} \;\leq\; \lambda \mathbf{b}^{*T} \left(\mathbf{z}_{k+1} - \mathbf{B}\mathbf{u}_{k+1}\right) \;\to\; 0,\; k \to \infty. \tag{B.52}$$

Combining (B.51) and (B.52) we obtain $p_k \to p^*$—i.e., value of the objective function estimate at iteration k converges to the true minimum as $k \to \infty$. From the bounded sequence $\mathbf{u}_k \in \mathbb{R}^N$ we can extract a convergent subsequence

$$\mathbf{u}_{k_i} \;\to\; \mathbf{u}^{**}. \tag{B.53}$$

Because our objective function is continuous, $\mathbf{u}^{**}$ is a solution of (B.1) and (C.5). However, if $\mathbf{A}$ is maximum rank the objective function of (B.1) is strictly convex, hence $\mathbf{u}^* = \mathbf{u}^{**}$. The sequence $\mathbf{u}_k$ must converge to $\mathbf{u}^*$ because otherwise we would be able to extract a subsequence convergent to a different limit and repeat the above analysis.

And finally, to prove that $\mathbf{b}_k \to \mathbf{b}^*$, we see that from the Karush-Kuhn-Tucker (KKT) conditions Boyd and Vandenberghe (2004) for (C.5) we have

$$\alpha \mathbf{A}\mathbf{A}^T \mathbf{u}^* \;=\; \mathbf{A}^T \mathbf{d} \;+\; \lambda \mathbf{B}^T \mathbf{b}^*. \tag{B.54}$$

Passing (B.32) to limit as $k \to \infty$, using (B.50) and replacing $\mathbf{b}_{k+1}$ with a convergent subsequence as necessary, we get

$$\alpha \mathbf{A}\mathbf{A}^T\mathbf{u}^* \;=\; \mathbf{A}^T\mathbf{d} \;+\; \lambda \mathbf{B}^T \lim \mathbf{b}_k. \tag{B.55}$$

Since $\mathbf{B}$ is maximum rank, $\operatorname{rank}\mathbf{B} = K \leq N$, (B.55) means that $\lim \mathbf{b}_k = \mathbf{b}^*$. □

Note that our our proof does not depend on the selection of starting values for $\mathbf{u}_0$, $\mathbf{z}_0$ and $\mathbf{b}_0$, and this fact will be used later on in proving the convergence of Algorithm 3. Before we study convergence properties of Algorithm 3, we prove one auxiliary result.

Theorem 2. *Algorithm 3 constructs a sequence of subspaces of* $\mathbb{R}^N$ *spanning expanding sets of conjugate directions,*

$$\begin{aligned} &S_k \;=\; \text{span}\left\{\mathbf{p}_0, \mathbf{p}_1, \ldots, \mathbf{p}_k\right\},\; k = 0, 1, 2, \ldots \\ &S_0 \subseteq S_1 \subseteq S_2 \subseteq \ldots \subseteq S_k \subseteq \ldots \end{aligned} \tag{B.56}$$

such that

$$\lim_{k\to\infty} S_k \;=\; S \;\subseteq\; \mathbb{R}^N. \tag{B.57}$$

Under the assumptions of Theorem 1, solution of the constrained optimization problem

$$\begin{aligned} &\|\mathbf{z}\|_1 \;+\; \frac{\alpha}{2}\|\mathbf{A}\mathbf{u} - \mathbf{d}\|_2^2 \;\to\; \min, \\ &\mathbf{z} \;=\; \mathbf{B}\mathbf{u}, \\ &\mathbf{u} \;\in\; S. \end{aligned} \tag{B.58}$$

matches the solution of (C.5).

Proof. If $S = \mathbb{R}^N$ statement of the theorem is trivial, so we assume that $\dim S < N$. Since our problem is finite-dimensional, the limit (B.57) is achieved at a finite iteration,

$$\exists k_1\; \forall k \geq k_1 :\; S_k \;\equiv\; S. \tag{B.59}$$

steps 4-7 of Algorithm 3 are equivalent to projecting the solution of the system of normal equations (B.21) onto the space S_k. If $p_{k+1} = 0$ in steps 20-22, then the right-hand side of (B.21) for any $k \geq k_1$ can be represented as a linear combination of vectors from $S_{k_1} \equiv S$. Steps 8 and 9 of Algorithm 3 are equivalent to steps 5 and 6 of Algorithm 1. Step 10 prepares the right-hand side of (B.21) for the minimization in step 4 of Algorithm 1 for iteration $k + 1$. However, since the right-hand side of (B.21) is a linear combination of vectors $\mathbf{p}_0, \mathbf{p}_1, \ldots, \mathbf{p}_k$ that span $S_k \equiv S$, steps 4-7 of Algorithm 3 are equivalent to the exact solution of the unconstrained minimization problem in step 4 of Algorithm 1. Hence, starting from iteration k_1 the two algorithms become equivalent. From Theorem 1 and

$$\forall k \geq k_1 : \ \mathbf{u}_{k+1} \ \in \ S$$

follows that the solution of (C.11) coincides with that of (C.5). □

Convergence of Algorithm 3 now becomes a trivial corollary of theorems 1 and 2.

Theorem 3. *Under the assumptions of Theorem 1, Algorithm 3 converges to the unique solution (B.27) of problem (C.5), and (B.29) holds.*

Proof. In the proof of Theorem 2 we have demonstrated that starting from $k = k_1$ defined in (B.59) Algorithm 3 is mathematically equivalent to Algorithm 1 starting from an initial approximation $\mathbf{u}_{k_1-1}$, $\mathbf{z}_{k_1-1}$ and $\mathbf{b}_{k_1-1}$. Convergence of Algorithm 1 does not depend on these starting values, hence Algorithm 3 converges to the same unique solution as Algorithm 1 and (B.29) holds. □

The result of Theorem 3 indicates that our Compressive Conjugate Directions method matches the ADMM in exact arithmetic after a finite number of iterations, while avoiding direct inversion of operator $\mathbf{A}$. This obvously means that the (worst-case) asymptotic convergence rate of Algorithm 3 matches that of the ADMM and is $O(1/k)$ He and Yuan (2012).

LIMITED-MEMORY COMPRESSIVE CONJUGATE DIRECTIONS METHOD

Algorithm 3 (that we call "unlimited-memory" Compressive Conjugate Directions Method) requires storing all of the previous conjugate directions (B.23) because in step 7 the algorithm computes the expansion

$$\mathbf{u}_{k+1} \;=\; \sum_{i=0}^{k} \tau_i \mathbf{p}_i, \tag{B.60}$$

of these solution approximations with respect to all conjugate direction vectors (B.23) at each iteration. It is a consequence of changing right-hand sides of the normal equations system (B.18) that *all* of the coefficients of expansion (B.60) may require updating. However, in a practical implementation we may expect that only the last $m+1$ expansion coefficients (B.60) significantly change, and freeze the coefficients

$$\tau_i, \; i \;<\; k-m$$

at and after iteration k. This approach requires storing up to $2m+2$ latest vectors

$$\mathbf{p}_k, \mathbf{p}_{k-1}, \ldots, \mathbf{p}_{k-m}, \quad \mathbf{q}_k, \mathbf{q}_{k-1}, \ldots, \mathbf{q}_{k-m}. \tag{B.61}$$

A "limited-memory" variant of the method is implemented in Algorithm 4 that stores vectors (B.61) in a circular first-in-first-out buffer. An index variable j points to the latest updated element within the buffer. Once j exceed the buffer size for the first time and is reset to point to the head of the buffer, a flag variable *cycle* is set, indicating that a search direction is overwritten at each subsequent iteration of the algorithm. The projection of the current solution iterate onto the old vector $\tau_j \mathbf{p}_j$ (now to be overwritten in the buffer) is then accumulated in a vector $\tilde{\mathbf{u}}$; the corresponding contribution to the predicted data equals $\tau_j \mathbf{q}_j$ and is accumulated in a vector $\tilde{\mathbf{v}}$,

$$\tilde{\mathbf{u}} \;=\; \sum_{i=0}^{k-m-1} \tau_i \mathbf{p}_i, \quad \tilde{\mathbf{v}} \;=\; \sum_{i=0}^{k-m-1} \tau_i \mathbf{q}_i. \tag{B.62}$$

Contributions (B.62) to the solution and predicted data from the discarded vectors (B.23) are then added back to the approximate solution and residual in steps 8 and 12 of Algorithm 4.

Algorithm 4 Limited-Memory Compressive Conjugate Directions Method for (B.1)

1: $m \leftarrow$ memory size, $\tilde{\mathbf{u}} \leftarrow \mathbf{0}^N$, $\tilde{\mathbf{v}} \leftarrow \mathbf{0}^{N+K}$, $j \leftarrow 0$, $cycle \leftarrow .false.$
2: $\mathbf{u}_0 \leftarrow \mathbf{0}$, $\mathbf{z}_0 \leftarrow \mathbf{0}^K$; $\mathbf{b}_0 \leftarrow \mathbf{0}^K$, $\mathbf{v}_0 \leftarrow \begin{bmatrix} \sqrt{\alpha}\mathbf{d} \\ \sqrt{\lambda}\,(\mathbf{z}_0 + \mathbf{b}_0) \end{bmatrix}$
3: $\mathbf{p}_0 \leftarrow \mathbf{F}^T\mathbf{v}_0$, $\mathbf{q}_0 \leftarrow \mathbf{F}\mathbf{p}_0$, $\delta_0 \leftarrow \mathbf{q}_0^T\mathbf{q}_0$
4: **for** $k = 0, 1, 2, 3, \ldots$ **do**
5: **for** $i = 0, 1, \ldots, \min(k, m)$ **do**
6: $\tau_i \leftarrow \mathbf{q}_i^T\,(\mathbf{v}_k - \tilde{\mathbf{v}})/\delta_i$
7: **end for**
8: $\mathbf{u}_{k+1} \leftarrow \tilde{\mathbf{u}} + \sum_{i=0}^{\min(k,m)} \tau_i \mathbf{p}_i$
9: $\mathbf{z}_{k+1} \leftarrow \text{shrink}\,\{\mathbf{B}\mathbf{u}_{k+1} - \mathbf{b}_k, 1/\lambda\}$
10: $\mathbf{b}_{k+1} \leftarrow \mathbf{b}_k + \mathbf{z}_{k+1} - \mathbf{B}\mathbf{u}_{k+1}$
11: $\mathbf{v}_{k+1} \leftarrow \begin{bmatrix} \sqrt{\alpha}\mathbf{d} \\ \sqrt{\lambda}\,(\mathbf{z}_{k+1} + \mathbf{b}_{k+1}) \end{bmatrix}$
12: $\mathbf{r}_{k+1} \leftarrow \mathbf{v}_{k+1} - \sum_{i=0}^{\min(k,m)} \tau_i \mathbf{q}_i - \tilde{\mathbf{v}}$
13: $\mathbf{w}_{k+1} \leftarrow \mathbf{F}^T\mathbf{r}_{k+1}$
14: $\mathbf{s}_{k+1} \leftarrow \mathbf{F}\mathbf{w}_{k+1}$
15: **for** $i = 0, 1, \ldots, \min(k, m)$ **do**
16: $\beta_i \leftarrow -\mathbf{q}_i^T\mathbf{s}_{k+1}/\delta_i$
17: **end for**
18: $j \leftarrow j + 1$
19: **if** $j = m + 1$ **then**
20: $j \leftarrow 0$, $cycle \leftarrow .true.$
21: **end if**
22: **if** $cycle$ **then**
23: $\tilde{\mathbf{u}} \leftarrow \tilde{\mathbf{u}} + \tau_j\mathbf{p}_j$
24: $\tilde{\mathbf{v}} \leftarrow \tilde{\mathbf{v}} + \tau_j\mathbf{q}_j$
25: **end if**
26: $\mathbf{p}_j \leftarrow \sum_{i=0}^{\min(k,m)} \beta_i\mathbf{p}_i + \mathbf{w}_{k+1}$
27: $\mathbf{q}_j \leftarrow \sum_{i=0}^{\min(k,m)} \beta_i\mathbf{q}_i + \mathbf{s}_{k+1}$

Algorithm 4 Limited-Memory Compressive Conjugate Directions Method (continued)

28: $\delta_j \leftarrow \mathbf{q}_j^T \mathbf{q}_j$
29: **if** $\delta_j = 0$ **then** ▷ Use condition "$\delta_j <$ tolerance" in practice
30: $\delta_j \leftarrow 1$, $\mathbf{p}_j \leftarrow \mathbf{0}^N$, $\mathbf{q}_j \leftarrow \mathbf{0}^{M+K}$
31: **end if**
32: Exit loop if $\|\mathbf{u}_{k+1} - \mathbf{u}_k\|_2 / \|\mathbf{u}_k\|_2 \leq$ target accuracy
33: **end for**

Trade-off between the number of iterations and problem condition number

In practical implementations of the ADMM when the operator $\mathbf{A}$ does not lend itself to direct solution methods, an iterative method can be used to solve the minimization problem in step 4 of Algorithm 1 Goldstein and Osher (2009). Algorithm 5, representing such an approach, uses a fixed number of iterations N_c of CGNE in step 4. At each iteration of the ADMM conjugate gradients are hot-restarted from the previous solution approximation $\mathbf{u}_k$. For comparison purposes we will refer to this method as *restarted Conjugate Gradients* or *RCG*. Note that Algorithm 5 with $N_c = 1$ performs

Algorithm 5 ADMM and hot-restarted CG (*RCG*)

1: $\mathbf{u}_0 \leftarrow \mathbf{0}^N$, $\mathbf{z}_0 \leftarrow \mathbf{0}^K$, $\mathbf{b}_0 \leftarrow \mathbf{0}^K$, $N_c \leftarrow$ prescribed number of CG iterations
2: $\mathbf{p}_0 \leftarrow \mathbf{F}^T \mathbf{v}_0$, $\mathbf{q}_0 \leftarrow \mathbf{F}\mathbf{p}_0$
3: **for** $k = 0, 1, 2, 3, \ldots$ **do**
4: Solve

$$\mathbf{u}_{k+1} \leftarrow \operatorname{argmin} \left\{ \frac{\lambda}{2} \|\mathbf{z}_k - \mathbf{B}\mathbf{u} + \mathbf{b}_k\|_2^2 + \frac{\alpha}{2} \|\mathbf{A}\mathbf{u} - \mathbf{d}\|_2^2 \right\},$$

starting from $\mathbf{u}_k$ and using N_c iterations of CGNE.
5: $\mathbf{z}_{k+1} \leftarrow \text{shrink}\{\mathbf{B}\mathbf{u}_{k+1} - \mathbf{b}_k, 1/\lambda\}$
6: $\mathbf{b}_{k+1} \leftarrow \mathbf{b}_k + \mathbf{z}_{k+1} - \mathbf{B}\mathbf{u}_{k+1}$
7: Exit loop if $\|\mathbf{u}_{k+1} - \mathbf{u}_k\|_2 / \|\mathbf{u}_k\|_2 \leq$ target accuracy
8: **end for**

a single step of gradient descent when solving the following intermediate least-squares

minimization problem in step 4,

$$\mathbf{u}_{k+1} \;=\; \text{argmin}\, \frac{\alpha}{2}\|\mathbf{A}\mathbf{u}-\mathbf{d}\|_2^2 \;+\; \frac{\lambda}{2}\|\mathbf{z}_k \;-\; \mathbf{B}\mathbf{u}+\mathbf{b}_k\|_2^2. \tag{B.63}$$

The performance of Algorithm 5 depends on the condition number of the leasts-squares problem (B.63) Trefethen and Bau III (1997): for well-conditioned problems only a small number of conjugate gradients iterations N_c may achieve a sufficiently accurate approximation to $\mathbf{u}_{k+1}$. The condition number of (B.63) depends on properties of operators $\mathbf{A}$ and $\mathbf{B}$, as well as the value of λ. In applications with a simple modeling operator $\mathbf{A}$, such as is the case in denoising with $\mathbf{A}=\mathbf{I}$, a value of λ may be experimentally selected so as to reduce the condition number of (B.63). However, a trade-off may exist between the condition number of (B.63) and the number of ADMM iterations in the outer loop (Step 3) of Algorithm 1: well-conditioned interim least-squares problems may result in a significantly higher number of ADMM iterations. Such a trade-off is a well-known phenomenon in applications of the Augmented Lagrangian Method of Multipliers for smooth objective functions, see, e.g., Glowinski and Le Tallec (1989). For example, large values of λ in (C.10) more strongly penalize violations of the equality constraint, as in the Quadratic Penalty Function Method Nocedal and Wright (2006) with a larger penalty and a more ill-conditioned quadratic minimization. Of course, in the case of ADMM applied to (B.1), a non-smooth objective function, arbitrary and potentially ill-conditioned operator $\mathbf{A}$, and (most importantly) alternating splitting minimization of the modified Augmented Lagrangian (C.10)[6] complicate the picture. In fact, for an arbitrary $\mathbf{A}$, the condition number of (B.63) is not always an increasing function of λ. Some of the numerical examples described in the following subsections exhibit this trade-off between the condition number of the intermediate least-squares problem (B.63) and the number of ADMM iterations: the better the condition-number of (B.63), the more ADMM iterations are typically required. The main advantage of our Compressive Conjugate Directions approach implemented in Algorithms 3 and 4 is that information on the geometry of the objective function (B.63) accumulates through *external* ADMM iterations thus

[6] "modified" because of the added constant term $\lambda/2\|\mathbf{b}_k\|_2^2$

potentially reducing the amount of effort required to perform minimization of (B.63) at each step. Since our objective is a practical implementation of the ADMM for (B.1) with computationally expensive operators $\mathbf{A}$, the overall number of operator $\mathbf{A}$ and $\mathbf{A}^T$ applications required to achieve given accuracy will be the principal benchmark for measuring the performance of various algorithms.

APPLICATIONS

In this section we apply the method of Compressive Conjugate Directions to solving L_1 and TV-regularized inversion problems for several practical examples.

Image Denoising

A popular image denoising technique for removing short-wavelength random Gaussian noise from an image is based on solving (C.1) with $\mathbf{A} = \mathbf{I}$. Vector $\mathbf{d}$ is populated with a noisy image, a denoised image is returned in $\mathbf{u}$,

$$\mathbf{u} = u_{i,j},\ i = 1, \ldots, N_y,\ j = 1, \ldots, N_x,$$

with an *anisotropic TV seminorm* in (C.1) defined by the linear gradient operator

$$\nabla \mathbf{u} \ = \ \begin{bmatrix} \nabla_x \mathbf{u} \\ \nabla_y \mathbf{u} \end{bmatrix} \ = \ \begin{bmatrix} u_{i,2} - u_{i,1} \\ \cdots \\ u_{i,N_x} - u_{i,N_x-1} \\ \cdots \\ u_{2,j} - u_{1,j} \\ \cdots \\ u_{N_y,j} - u_{N_y-1,j} \end{bmatrix},\ i = 1, \ldots, N_y,\ j = 1, \ldots, N_x. \tag{B.64}$$

Here, the dimension of the model space is $N = N_x \times N_y$ with $M = N$ and $K = N - N_x - N_y$. Since operator $\mathbf{A} = \mathbf{I}$ is trivial, minimization of the number of operator

applications in this problem carries no practical advantage. The only reason for providing this example is to demonstrate the stability of the proposed Compressive Conjugate Directions method with respect to choosing a value of λ.

Figure B.1(a) shows the true, noise-free 382×382 image used in this experiment. Random Gaussian noise with a standard deviation σ of 15% of maximum signal amplitude was added to the true image to produce the noisy image of Figure B.1(b). All low-wavenumber or "blocky" components of the noise below a quarter of the Nyquist wavenumber were filtered out, leaving only high-wavenumber "salt-and-pepper" noise. Parameter $\alpha = 10$ was chosen experimentally based on the desired trade-off of fidelity and "blockiness" of the resulting denoised image. The result of solving (C.1) using Algorithm 5 with $\lambda = 1$, one hundred combined applications of $\mathbf{A}$ and $\mathbf{A}^T$, and $N_c = 1$ is shown in Figure B.1(d). The result of applying our limited-memory Conjugate Directions Algorithm 4 for $m = 50$ is shown in Figure B.1(c)[7]. Note that $N_c = 1$ means that only a single step of Conjugate Gradients, or a single gradient descent, is made in step 4 of Algorithm 5. For this choice of λ, problem (B.63) is very well conditioned, with a condition number of $\kappa = 1.8$. A single iteration of gradient descent achieves sufficient accuracy of minimization (B.63) and for $\lambda = 1$ there is no practical advantage in using our method as both methods perform equally well, see Figure B.2(a). In fact, the overhead of storing and using conjugate directions from previous iterations may exceed the cost of operator $\mathbf{A}$ and its adjoint applications if the latter are computationally cheap. The approximation errors of applying the limited-memory Compressive Conjugate Directions Algorithm 4 with $m = 50$ versus Algorithm 5 with $N_c = 1, 5, 10$ for $\lambda = 10^2, 10^3, 10^4$ are shown in Figures B.2(a),B.2(b),B.2(c),B.2(d). Note that larger values of λ result in increasingly larger condition numbers of (B.63) shown on top of the plots. The performance of Algorithm 5 here depends on a choice of N_c: increasing N_c as required to achieve a sufficiently accurate approximate solution of (B.63) results in fewer available ADMM iterations for a fixed "budget" of operator $\mathbf{A}$ and adjoint applications. However, Algorithm 4 accumulates conjugate directions (B.23) computed at earlier iterations and requires only one application of

[7]Here, this matches the results for *any* memory size $m > 0$ due to a well-conditioned problem (B.63).

Figure B.1: (a) Clean image [**NR**]; (b) Noisy image contaminated with Gaussian noise with $\sigma = 15\%$ of maximum amplitude [**CR**]; (c) Image denoised using Algorithm 4 with $\alpha = 10$, $\lambda = 1$ and memory size $m = 50$ [**CR**]; (d) Image denoised using Algorithm 5 with $\alpha = 10$, $\lambda = 1$, $N_c = 1$ [**CR**]. ccd/. X1trueimg,X1noisyimg,X1ccd,X1rcg

the operator and its adjoint per ADMM iteration. Note that at iteration steps less than N_c, Algorithm 5 may still outperform Algorithm 4 as it conducts more Conjugate Gradient iterations per solution of each problem (B.63). However, once the ADMM iteration count exceeds the largest N_c, and sufficient information is accumulated by Algorithm 4 about the geometry of the objective function, the Compressive Conjugate Directions outperforms Algorithm 5. Note that this example does not demonstrate the trade-off between the condition number of (B.63) and the number of ADMM iterations. The reason for this is that for large λ convergence is achieved relatively quickly within a number of iterations comparable to a number of Conjugate Gradients steps required to solve (B.63). However, this example demonstrate another feature of the proposed Compressive Conjugate Directions Method: compared with a technique based on a restarted iterative solution of (B.63), the method may be less sensitive to a suboptimal choice of λ.

Inversion of Dilatational Point Pseudo-sources

In our second example, we demonstrate our method on a geomechanical inversion problem with a non-trivial forward-modeling operator $\mathbf{A}$. Here, we are interested in inverting subsurface sources of deformation from noisy measurements of surface displacements, such as GPS, tilt-meter and InSAR observations.

The forward modeling operator simulates vertical surface displacements in response to distributed dilatational (e.g. pressure change) sources Segall (2010). Our modeling operator is defined as

$$\mathbf{Au} \;=\; d(z),\; d(z) \;=\; c\int_0^A \frac{Du(\xi)d\xi}{\left(D^2+(z-\xi)^2\right)^{3/2}}, \tag{B.65}$$

where we assume that $\mathbf{u} = u(\xi), \xi \in [0, A]$ is a relative pore pressure change along a horizontal segment $[0, A]$ of a reservoir at a constant depth D, $\mathbf{d} = d(x), x \in [0, A]$ is the induced vertical displacement on the surface, and a factor c is determined by the

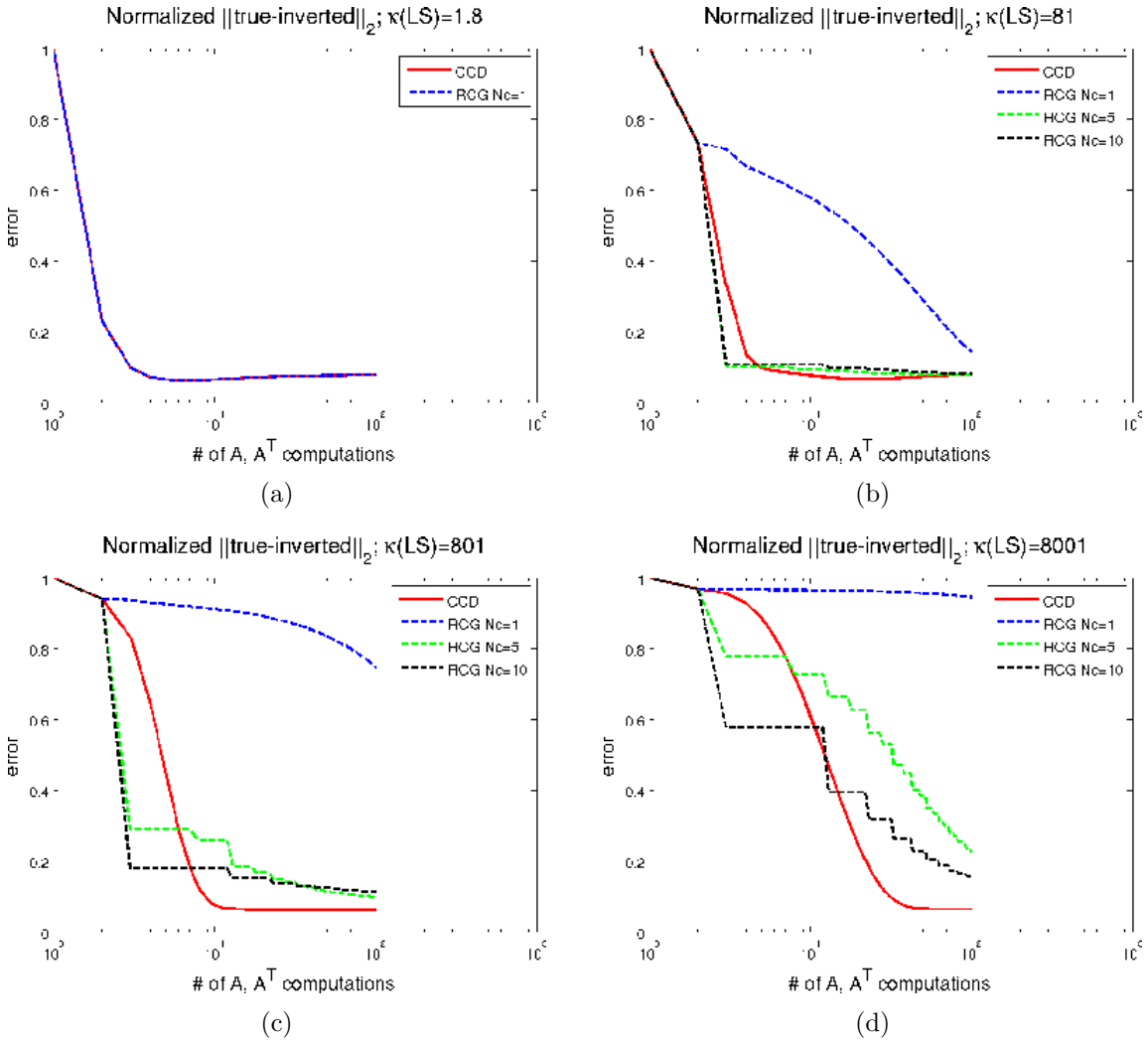
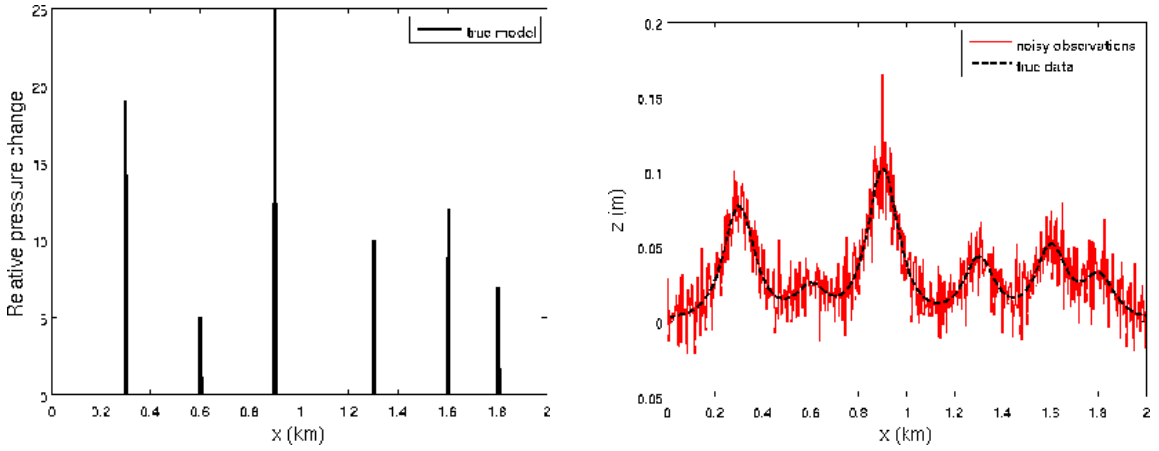


Figure B.2: Performance of Algorithm 4 with $m = 20$ versus Algorithm 5 with varying N_c for (a) $\lambda = 1$ [**CR**]; (b) $\lambda = 100$ [**CR**]; (c) $\lambda = 1000$ [**CR**]; (d) $\lambda = 10000$ [**CR**]. ccd/. X1lam1,X1lam100,X1lam1000,X1lam10000

poroelastic medium properties, and reservoir dimensions. In this example, for demonstration purposes we consider a two-dimensional model, but a three-dimensional model is studied in the next subsection. Operator (C.19) is a smoothing integral operator that, after discretization and application of a simple quadrature, can be represented by a dense matrix. Analytical representation of the surface displacement modeling operator (C.19) is possible for simple homogeneous media; however, modeling surface displacements in highly heterogeneous media will involve computationally expensive numerical methods such as Finite Elements Kosloff et al. (1980).

In this experiment we seek to recover a spiky model of subsurface sources shown in Figure B.3(a) from noisy observations of the induced surface displacements shown in Figure B.3(b). Such sparse dilatational pseudo-sources are mathematically equiv-



(a) (b)

Figure B.3: (a) A spiky true pseudosources [**CR**]; (b) the resulting true (black) and noisy (red) surface displacements [**CR**]. ccd/. X2true,X2data

alent to concentrated reservoir pressure changes in hydrogeology and exploration geophysics, as well as expanding spherical lava chambers (the "Mogi model") in volcanology Segall (2010). We forward-modeled surface displacements due to the sources of Figure B.3(a) using operator (C.19), and, as in our denoising tests, added random Gaussian noise with $\sigma = 15\%$ of the maximum data amplitude. Prior to adding the noise, all low-wavenumber noise components below a fifth of the Nyquist wavenumber

were muted, leaving only the high-wavenumber noise shown in Figure B.3(b).

We set $D = .1$ km, $A = 2$ km, $c = 10^{-2}$ in (C.19), and discretized both the model and data space using a 500-point uniform grid, $N = M = 500$. We solve problem (B.2) with $\alpha = 10000$, and our objective is to accurately identify locations of the spikes in Figure B.3(a) and their relative magnitudes, carrying out as few applications of operator (C.19) as possible. Inversion results of using the limited-memory Compressive Conjugate Directions Algorithm 4 with $m = 100$, ADMM with restarted Conjugate Gradients Algorithm 5 and FISTA of (B.6) are shown in Figures B.4(a),B.4(b),B.4(c),B.4(d) for $\lambda = 0.05, 0.1, 1, 100$. In each case one hundred combined products of operators $\mathbf{A}$ and $\mathbf{A}^T$ with vectors were computed. We used the maximum FISTA step size of $\tau = 10^{-4}$ in (B.6) computed for operator (C.19). These results indicate that the Compressive Conjugate Directions method achieves qualitative recovery of the spiky model at early iterations. Superiority of the new method is especially pronounced when the intermediate least-squares minimization problem (B.63) is ill-conditioned (see plot tops). The method retains its advantage after 1000 operator and adjoint applications, as shown in Figures B.5(a),B.5(b),B.5(c),B.5(d). Note that the error plots of the CCD in Figures B.6(a),B.6(b),B.6(c),B.6(d) exhibit a trade-off between the convergence rate and condition number of problem (B.63) discussed earlier: a more ill-conditioned (B.63) is associated with a faster convergence rate of the new method.

Figures B.7(a),B.7(b),B.7(c),B.7(d) show error plots for the CCD, ADMM with *exact* minimization of (B.63), and FISTA. The said trade-off between the convergence rate and condition number of (B.63) is exhibited by the ADMM. The CCD curves approach the convergence rates of the ADMM once Algorithm 4 has accumulated enough information about the geometry of the objective function in vectors (B.61). Note that the advantage of a faster asymptotic convergence rate of FISTA kicks in only when the ADMM-based methods use values of λ that are not optimal for their convergence—see Figures B.6(d) and B.7(d). In this case (B.63) is very well conditioned, and its adequate solution requires only a single step of gradient descent at each iteration of the ADMM, depriving conjugate-gradients-based methods

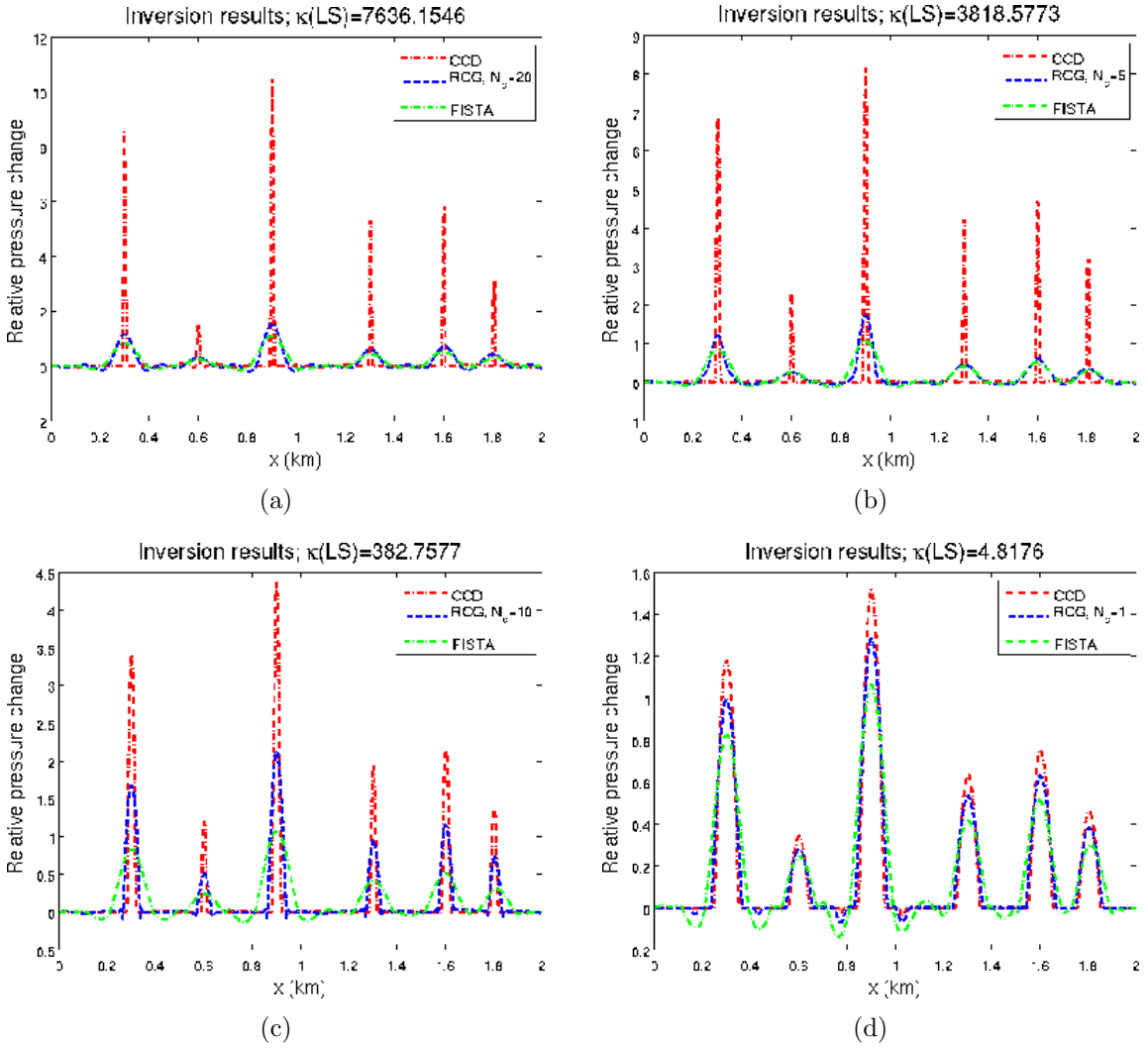


Figure B.4: Inversion results for CCD (red), RCG (blue), FISTA (green) after 100 operator and adjoint applications for (a) $\lambda = .05$ [**CR**]; (b) $\lambda = 0.1$ [**CR**]; (c) $\lambda = 1$ [**CR**]; (d) $\lambda = 100$ [**CR**]. Note that FISTA does not use λ and the same FISTA results are shown in all plots but using different vertical scales. Improving condition number of (B.63) is accompanied by slower convergence. Compressive Conjugate Directions method most accurately resolves the spiky model at early iterations, and performs well when (B.63) is ill-conditioned. ccd/. X2lam005invonly,X2lam01invonly,X2lam1invonly,X2lam100invonly

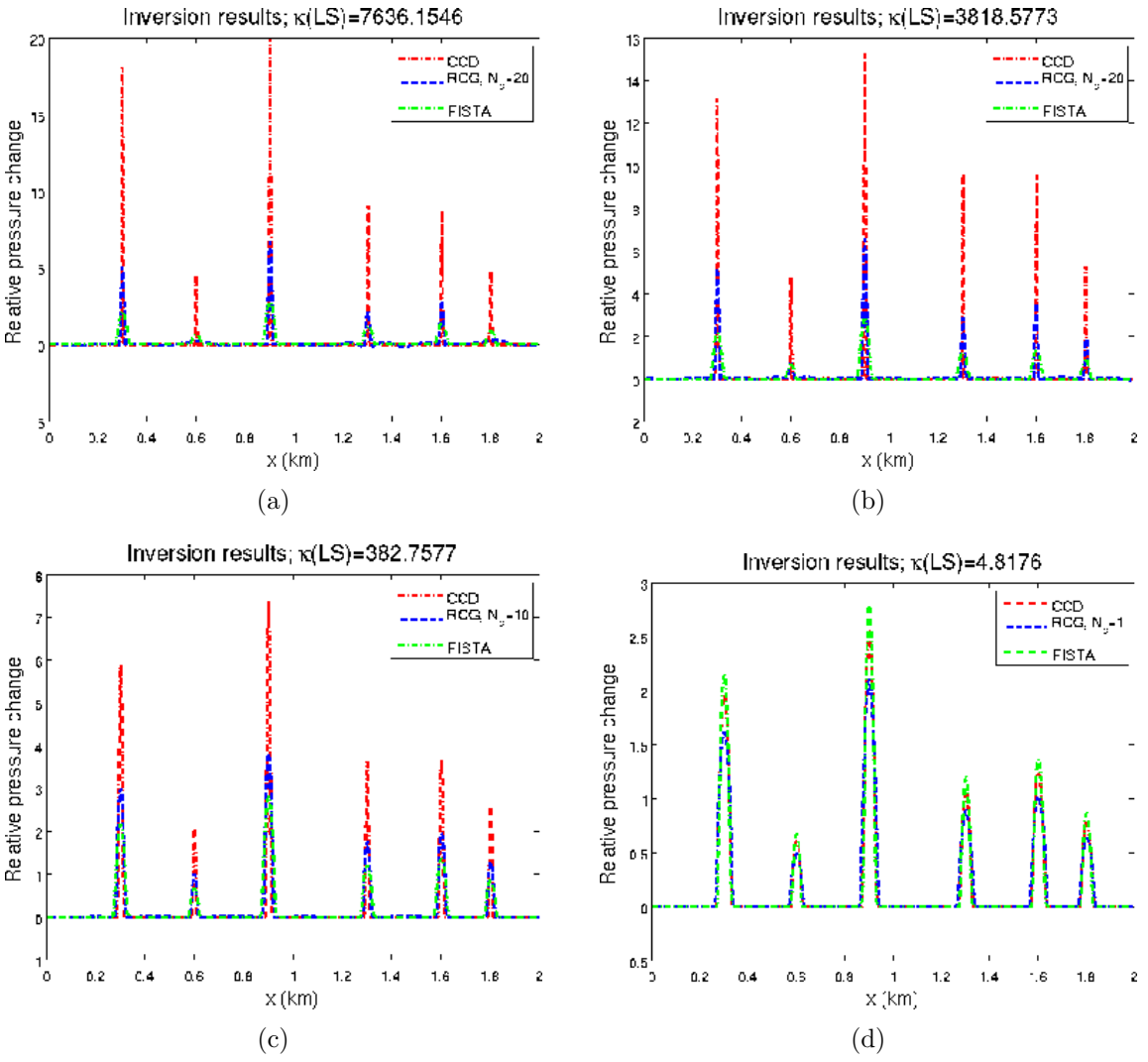


Figure B.5: Inversion results for CCD (red), RCG (blue), FISTA (green) after 1000 operator and adjoint applications for (a) $\lambda = .05$ [**CR**]; (b) $\lambda = 0.1$ [**CR**]; (c) $\lambda = 1$ [**CR**]; (d) $\lambda = 100$ [**CR**]. Note that FISTA does not use λ and the same FISTA results are shown in all plots but using different vertical scales. Compressive Conjugate Directions method still retains its advantage in resolving the spiky model at earlier iterations. *Asymptotically* faster convergence of FISTA kicks in when $\lambda = 100$ with a well-conditioned (B.63), when the ADMM convergence is slowed—compare with Figure B.7(d).

ccd/. X21000itlam005invonly,X21000itlam01invonly,X21000itlam1invonly,X21000itlam100invonly

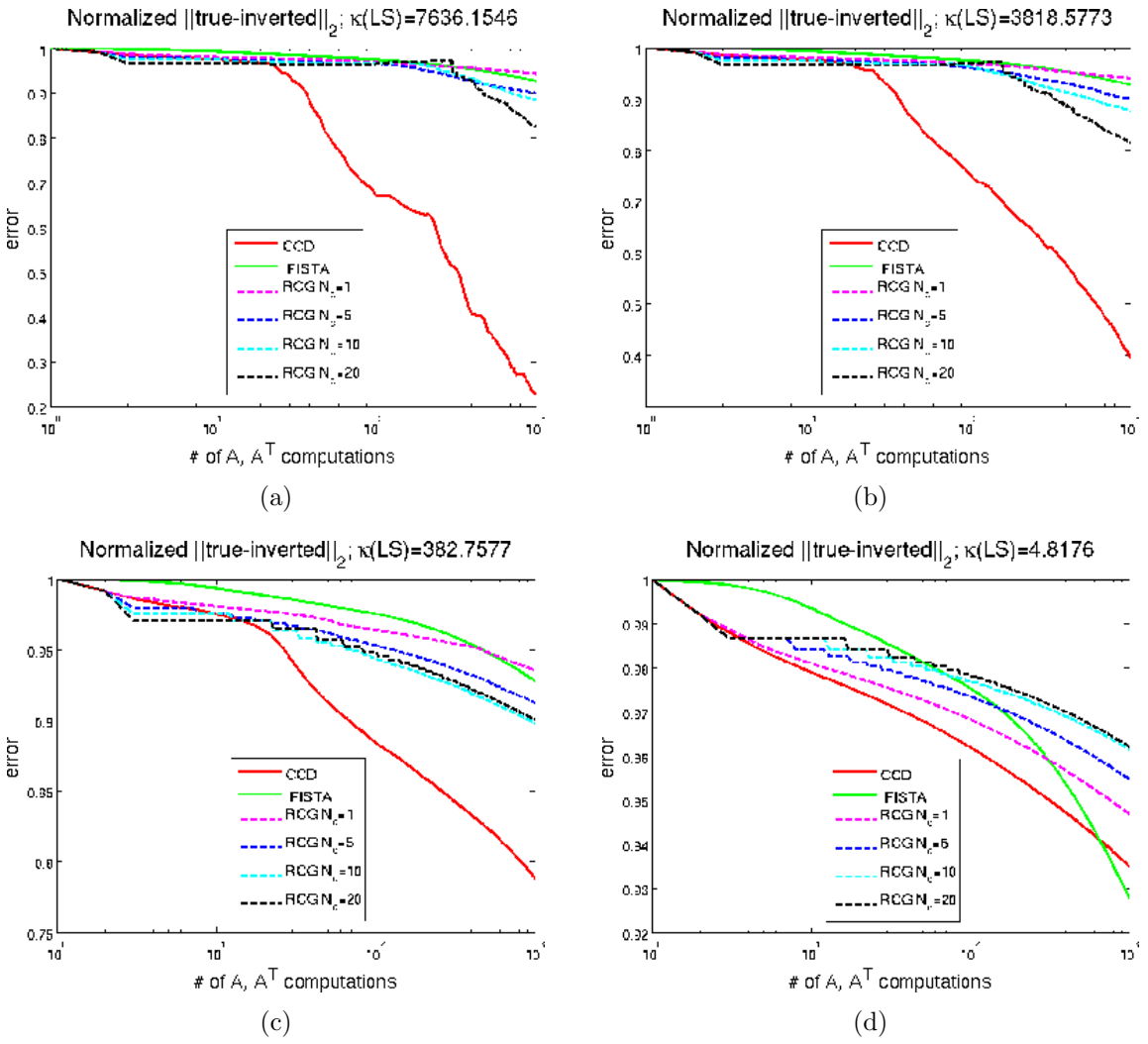


Figure B.6: Convergence curves for CCD (solid red), RCG (dashed), FISTA (solid green) for (a) $\lambda = .05$ [**CR**]; (b) $\lambda = 0.1$ [**CR**]; (c) $\lambda = 1$ [**CR**]; (d) $\lambda = 100$—compare with Figures B.5(a),B.5(b),B.5(c),B.5(d) [**CR**]. ccd/. X2lam005conv,X2lam01conv,X2lam1conv,X2lam100conv

of their advantage. FISTA, being based on accelerating a gradient-descent method, now *asymptotically* beats the convergence rates of the other techniques but this happens too late through the iterations to be of practical significance. In other words, in this particular example FISTA can beat the ADMM (and CCD) only if the latter use badly selected values of λ. Generalizing this observation about FISTA and ADMM for problem (B.2) with a general operator $\mathbf{A}$ goes beyond the scope of our work.

Inversion of Pressure Contrasts

In this section we apply the Compressive Conjugate Gradients method to identify sharp subsurface pressure contrasts in a reservoir from observations of induced surface displacements. We use a 3-dimensional geomechanical poroelastostatic model of pressure-induced deformation based on Biot's theory Segall (2010).

We solve a TV-regularized inversion problem (C.1) with operator $\mathbf{B}$ given by (B.64), and operator $\mathbf{A}$ given by extension of (C.19)

$$\mathbf{Au} \;=\; d(x,y),\; d(x,y) \;=\; c\int_0^A \int_0^A \frac{Du(\xi,\eta)d\xi d\eta}{(D^2+(x-\xi)^2+(y-\eta)^2)^{3/2}}, \tag{B.66}$$

where we assume that $\mathbf{u} = u(\xi,\eta), (\xi,\eta) \in [-A,A]\times[-A,A]$ is a relative pore pressure change at a point (ξ,η) of the reservoir at a constant depth D, $2A$ is the reservoir length and breadth, $\mathbf{d} = d(x,y), (x,y) \in [-A,A]\times[-A,A]$ is the induced vertical displacement at a point (x,y) on the surface, and a constant factor c is determined by the poroelastic medium properties and reservoir thickness.

In this experiment, we discretize the pressure and displacement using a 50×50 grid, with $A = 1.2$ km, $D = .455$ km and $c = 5.8515 \times 10^3$, based on a poroelastic model of a real-world unconventional hydrocarbon reservoir Maharramov and Zoback (2014). We use a least-squares fitting weight $\alpha = .1$ in (C.1) to achieve a desirable trade-off between fitting fidelity and blockiness of the inverted pressure change. The blocky model shown in Figure B.8(a) was used to forward-model surface displacements using operator (B.66). Random Gaussian noise with $\sigma = 0.15\%$ of maximum data

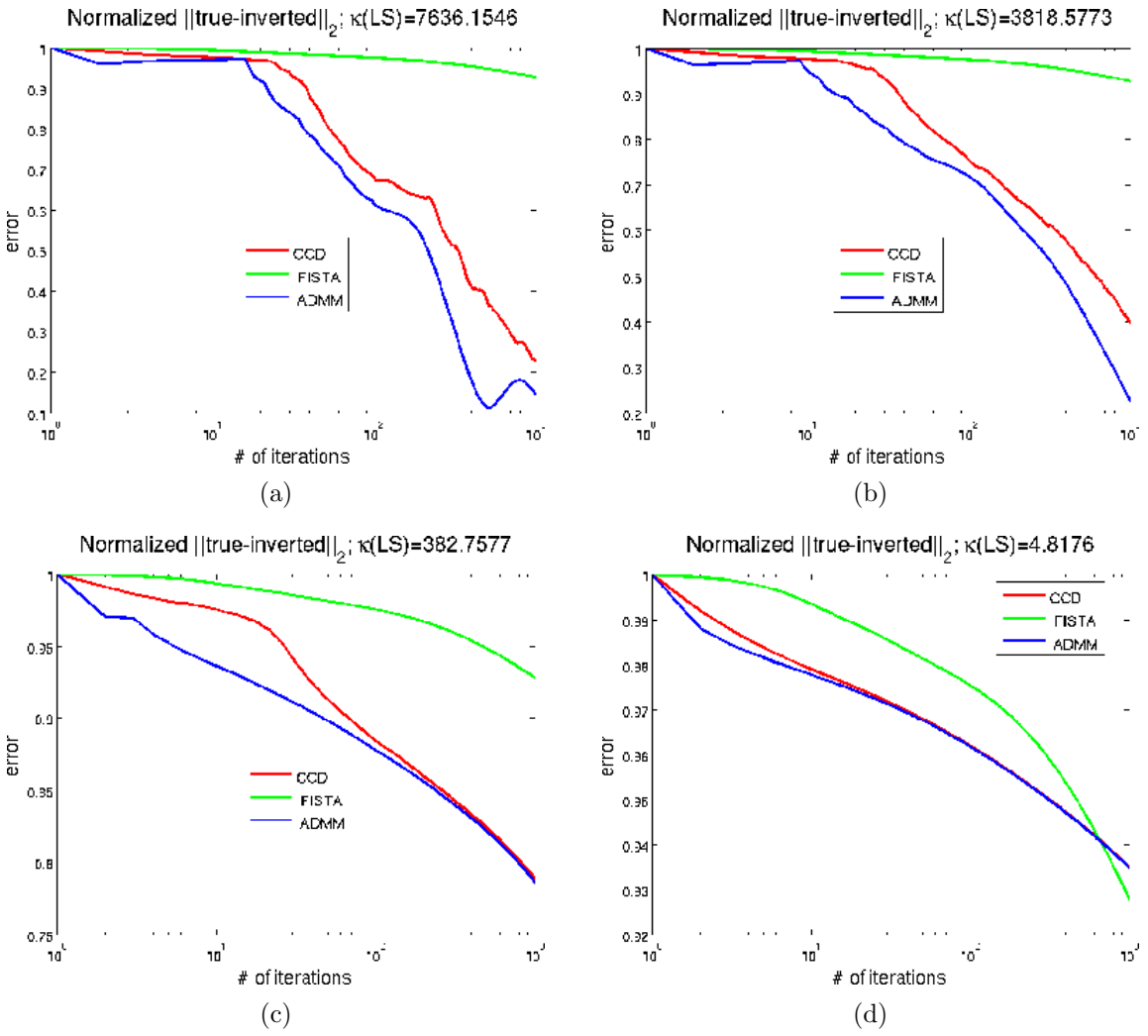
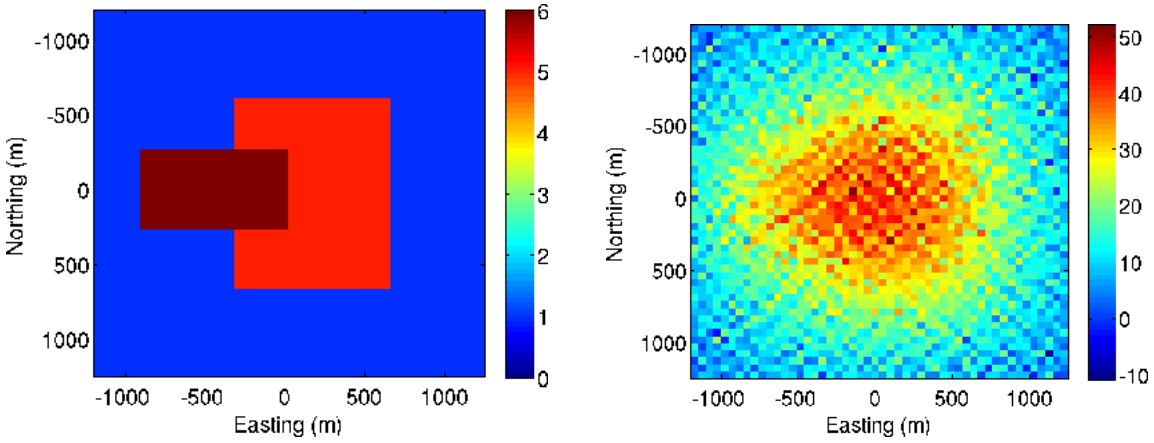


Figure B.7: Convergence curves for CCD (solid red), ADMM with exact solver (blue), FISTA (green) for (a) $\lambda = .05$; (b) $\lambda = 0.1$; (c) $\lambda = 1$; (d) $\lambda = 100$. Limited-memory Compressive Conjugate Directions with $m = 100$ achieves convergence rate comparable to ADMM with exact minimization of (B.63).

ccd/. X2admmlam005conv,X2admmlam01conv,X2admmlam1conv,X2admmlam100conv

amplitude, muted below a quarter of the Nyquist wavenumber, was added to the clean data to produce the noisy displacement measurements of Figure B.8(b).



(a) (b)

Figure B.8: (a) A blocky true pressure model (MPa) [**CR**]; (b) the resulting surface displacements (mm) with added random Gaussian noise with $\sigma = 15\%$ of data amplitude [**CR**]. ccd/. X4true,X4data

Figure B.9(a) shows the result of the limited-memory Compressive Conjugate Directions Algorithm 4 with $m = 100$, after a total of 100 combined applications of operator $\mathbf{A}$ and its adjoint. For the same number of operator applications, Figure B.9(b) shows the best result of the ADMM with restarted Conjugate Gradients Algorithm 5. The corresponding results after 1000 applications of $\mathbf{A}$ and $\mathbf{A}^T$ are shown in Figures B.9(c) and B.9(d), respectively.

The Compressive Conjugate Directions method resolves key model features faster than the ADMM using iterative solution of (B.63) restarted at each ADMM iteration. This advantage of our method is particularly pronounced when the intermediate least-squares problem (B.63) is ill-conditioned—compare Figures B.10(a),B.10(b) with Figures B.10(c),B.10(d). To accurately resolve the blocky pressure model of Figure B.8(a), the Compressive Conjugate Directions technique requires about a tenth of operator $\mathbf{A}$ and adjoint applications compared with Algorithm 5 when (B.63) is poorly conditioned. And again, as in the previous example, there is a trade-off between the convergence rate of the Compressive Conjugate Directions and the condition number

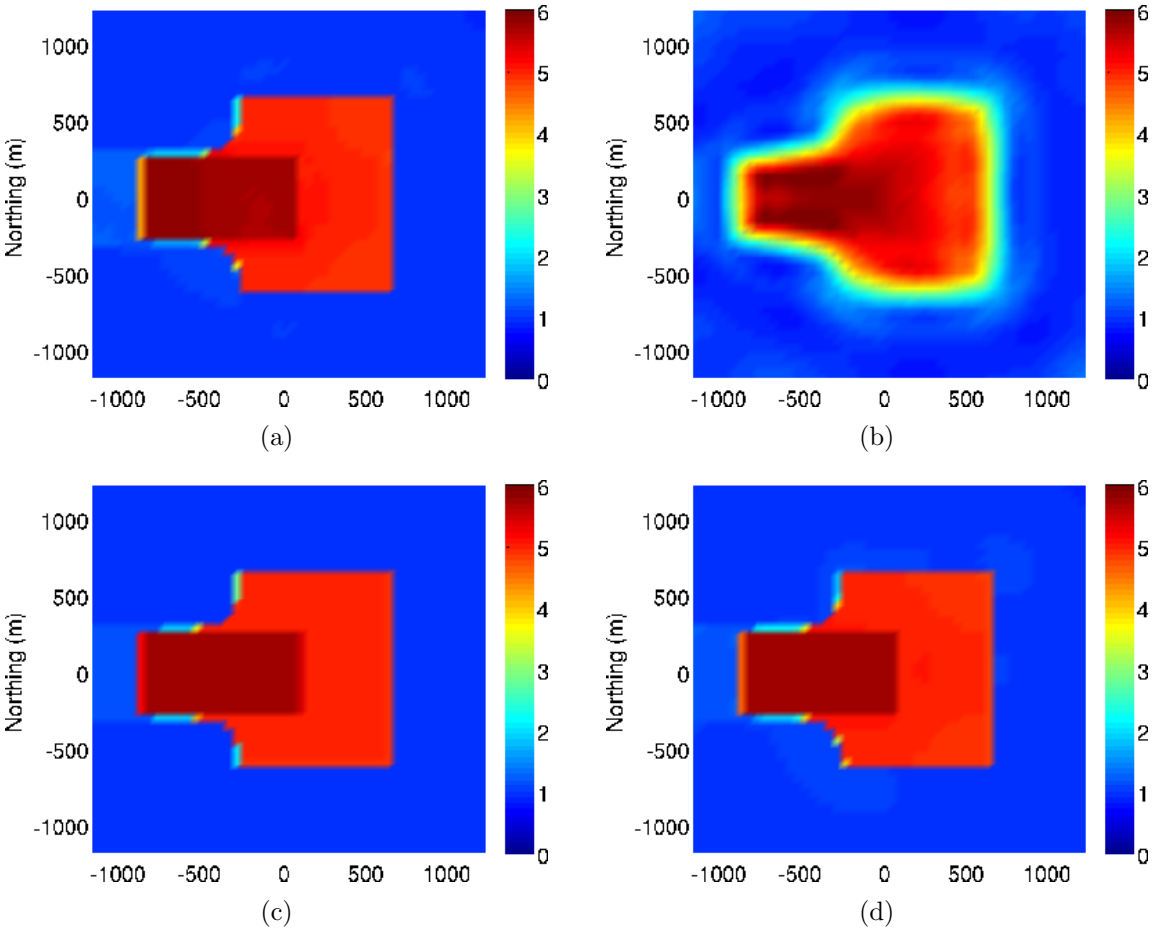


Figure B.9: Inversion results after (a) 100 iterations (operator and adjoint applications) of CCD with $\lambda = 10$ [**CR**]; (b) 100 iterations of RCG with $\lambda = 10$ [**CR**]; (c) 1000 iterations of CCD with $\lambda = 10$ [**CR**]; (d) 1000 iterations of RCG with $\lambda = 10$ [**CR**]. In all tests, CCD is the limited-memory Compressive Conjugate Directions method of Algorithm 4; RCG is ADMM with restarted Conjugate Gradients of Algorithm 5 showing the most accurate model reconstruction among the outputs for different N_c–see Figures B.10(a),B.10(b),B.10(c),B.10(d).
ccd/. X4it100lam10ccd,X4it100lam10rcg,X4lam10ccd,X4lam10rcg

of (B.63): values of λ that result in more poorly-conditioned (B.63) yield the fastest convergence.

DISCUSSION

Compressive Conjugate Directions provides an efficient implementation of the Alternating Direction Method of Multipliers in $L_1 - TV$ regularized inversion problems (B.1) with computationally expensive operators $\mathbf{A}$. By accumulating and reusing information on the geometry of the intermediate quadratic objective function (B.63), the method requires only one application of the operator $\mathbf{A}$ and its adjoint per ADMM iteration while achieving accuracy comparable to that of the ADMM with exact minimization of (B.63). The method does not improve the worst-case asymptotic convergence rate of the ADMM. However, it can be used for fast recovery of spiky or blocky solution components. The method trades the computational cost of applying operator $\mathbf{A}$ and its adjoint for extra memory required to store previous conjugate direction vectors (B.61).

Our numerical experiments involving problems of geomechanical inversion demonstrated a trade-off between the number of ADMM iterations required to achieve a sufficiently accurate solution approximation, and condition number of the intermediate least-squares problem (B.63). Understanding the extent to which this phenomenon applies to solving (B.1) with other classes of modeling operators $\mathbf{A}$ requires further analysis.

Generalizations

The primary focus of this work are $L_1 - TV$ regularized inversion problems (B.1). However, the Steered Conjugate Directions Algorithm 2 can be combined with the Method of Multipliers to solve more general problems of large-scale equality-constrained optimization.

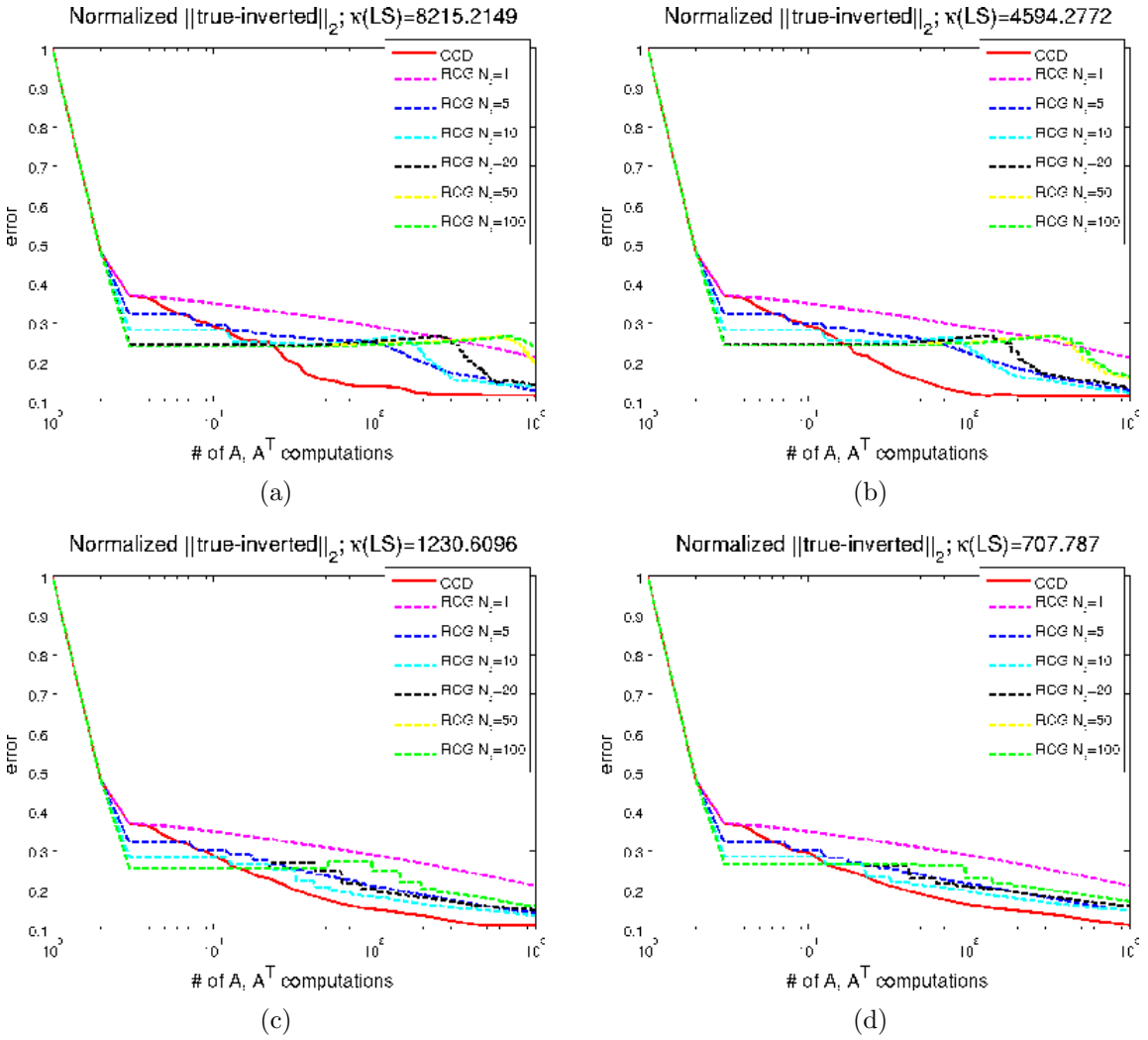


Figure B.10: Convergence rates for CCD and RCG with various N_c for (a) $\lambda = 5$ [**CR**]; (b) $\lambda = 10$ [**CR**]; (c) $\lambda = 50$ [**CR**]; (d) $\lambda = 100$ [**CR**]. ccd/. X4lam5conv,X4lam10conv,X4lam50conv,X4lam100conv

For example, consider the problem

$$\begin{aligned} &\|\mathbf{A}\mathbf{u}-\mathbf{d}\|_2^2 \;\rightarrow\; \min, \\ &\mathbf{B}\mathbf{u}-\mathbf{c} \;=\; \mathbf{0}, \\ &\mathbf{u} \;\in\; \mathbb{R}^N,\; \mathbf{d} \;\in\; \mathbb{R}^M,\; \mathbf{A}:\mathbb{R}^N \rightarrow \mathbb{R}^M,\; \mathbf{B}:\mathbb{R}^N \rightarrow \mathbb{R}^K, \end{aligned} \tag{B.67}$$

where $\mathbf{A}$ is a computationally expensive operator. Many "coupled" systems governing two or more physical parameters can be described mathematically as a constrained problem (B.67). Of special interest are cases when $K \;\ll\; \min\{N, M\}$—e.g., large-scale optimization problems with a localized constraint. Applying the Augmented Lagrangian Method of Multipliers to (B.67), after re-scaling the multiplier vector, we get

$$\begin{aligned} \mathbf{u}_{k+1} \;&= \operatorname{argmin} \|\mathbf{A}\mathbf{u}-\mathbf{d}\|_2^2 + \frac{\lambda}{2}\|\mathbf{c}-\mathbf{B}\mathbf{u}+\mathbf{b}_k\|_2^2, \\ \mathbf{b}_{k+1} \;&= \mathbf{b}_k \;+\; \mathbf{c}-\mathbf{B}\mathbf{u}_{k+1}. \end{aligned} \tag{B.68}$$

As before, the minimization on the first line of (B.68) is equivalent to solving a system of normal equations with a fixed left-hand side and changing right-hand sides. Combining the dual-variable updates from (B.68) with Algorithm 2, we get Algorithm 6.

Algorithm 6 Steered Conjugate Directions + Method of Multipliers for solving (B.67)

$\mathbf{u}_0 \leftarrow \mathbf{0}^N,\; \mathbf{b}_0 \leftarrow \mathbf{0}^K,\; \mathbf{v}_0 \leftarrow \begin{bmatrix} \mathbf{d} \\ \sqrt{\lambda}\,(\mathbf{c}+\mathbf{b}_0) \end{bmatrix}$
$\mathbf{p}_0 \leftarrow \mathbf{F}^T\mathbf{v}_0,\; \mathbf{q}_0 \leftarrow \mathbf{F}\mathbf{p}_0,\; \delta_0 \leftarrow \mathbf{q}_0^T\mathbf{q}_0$
for $k = 0, 1, 2, 3, \ldots$ **do**
    for $i = 0, 1, \ldots, k$ **do**
        $\tau_i \leftarrow \mathbf{q}_i^T\mathbf{v}_k/\delta_i$
    end for
    $\mathbf{u}_{k+1} \leftarrow \sum_{i=0}^{k} \tau_i \mathbf{p}_i$
    $\mathbf{b}_{k+1} \leftarrow \mathbf{b}_k + \mathbf{c} - \mathbf{B}\mathbf{u}_{k+1}$
    $\mathbf{v}_{k+1} \leftarrow \begin{bmatrix} \mathbf{d} \\ \sqrt{\lambda}\,(\mathbf{c}+\mathbf{b}_{k+1}) \end{bmatrix}$
    $\mathbf{r}_{k+1} \leftarrow \mathbf{v}_{k+1} - \sum_{i=0}^{k} \tau_i \mathbf{q}_i$

Operator $\mathbf{F}$ in Algorithm 6 is given by (B.19) with $\alpha = 1$. A limited-memory

Algorithm 6 Steered Conjugate Directions + Method of Multipliers (continued)

11: $\mathbf{w}_{k+1} \leftarrow \mathbf{F}^T\mathbf{r}_{k+1}$
12: $\mathbf{s}_{k+1} \leftarrow \mathbf{F}\mathbf{w}_{k+1}$
13: **for** $i = 0, 1, \ldots, k$ **do**
14: $\beta_i \leftarrow -\mathbf{q}_i^T\mathbf{s}_{k+1}/\delta_i$
15: **end for**
16: $\mathbf{p}_{k+1} \leftarrow \sum_{i=0}^{k}\beta_i\mathbf{p}_i + \mathbf{w}_{k+1}$
17: $\mathbf{q}_{k+1} \leftarrow \sum_{i=0}^{k}\beta_i\mathbf{q}_i + \mathbf{s}_{k+1}$
18: $\delta_{k+1} \leftarrow \mathbf{q}_{k+1}^T\mathbf{q}_{k+1}$
19: **if** $\delta_{k+1} = 0$ **then** ▷ Use condition "$\delta_{k+1} <$ tolerance" in practice
20: $\delta_{k+1} \leftarrow 1$, $\mathbf{p}_{k+1} \leftarrow \mathbf{0}^N$, $\mathbf{q}_{k+1} \leftarrow \mathbf{0}^{M+K}$
21: **end if**
22: Exit loop if $\|\mathbf{u}_{k+1} - \mathbf{u}_k\|_2/\|\mathbf{u}_k\|_2 \leq$ target accuracy
23: **end for**

version of Algorithm 6 is obtained trivially by adapting Algorithm 4. We envisage potential utility of Algorithm 6 in applications where storing a set of previous conjugate direction vectors (B.61) is computationally more efficient that iteratively solving the quadratic minimization problem in (B.68) from scratch at each iteration of the method of multipliers.

The Compressive Conjugate Directions Algorithm 4 can be extended for solving non-linear inversion problems with L_1 and *isotropic* total-variation regularization. Likewise, the Steered Conjugate Directions Algorithm 6 can be adapted to solving general equality-constrained non-linear optimization problems. A nonlinear theory and further applications of these techniques will be the subject of our next work.

Appendix C

Total-variation minimization with bound constraints

"In all pointed sentences, some degree of accuracy must be sacrificed to conciseness."

Samuel Johnson

I present here a powerful and easy-to-implement algorithm for approximate solution of constrained optimization problems that involve L_1/total-variation regularization terms, and both equality and inequality constraints. I discuss the relationship of this method to earlier works of Goldstein and Osher (2009) and Chartrand and Wohlberg (2010), and demonstrate that my approach is a combination of the augmented Lagrangian method with splitting and model projection. I test the method on a geomechanical problem and invert highly compartmentalized pressure change from noisy surface uplift observations.

INTRODUCTION

The primary focus of this work is a class of least-squares fitting problems with a total-variation (TV) regularization and bound model constraints:

$$\begin{aligned} &\||\nabla \mathbf{m}|\|_1 \; + \; \frac{\alpha}{2}\|\mathbf{F}(\mathbf{m}) - \mathbf{d}\|_2^2 \; \rightarrow \; \min, \\ &\mathbf{m}_1 \; \leq \; \mathbf{m} \; \leq \; \mathbf{m}_2. \end{aligned} \tag{C.1}$$

In (C.1) we seek a model vector $\mathbf{m}$ such that forward-modeled data $\mathbf{F}(\mathbf{m})$ match observed data $\mathbf{d}$ in the least squares sense, while imposing blockiness-promoting total-variation (TV) regularization (Rudin et al., 1992) and lower ($\mathbf{m}_1$) and upper ($\mathbf{m}_2$) model bounds. Rather than using a regularization parameter as a coefficient of the regularization term, we use a data-fitting weight α. TV regularization (also know as the Rudin-Osher-Fatemi, or ROF, model) acts as a form of model styling that helps to preserve sharp contrasts and boundaries in the model, even when spectral content of input data has limited resolution. Examples of successful geophysical application of unconstrained TV-regularized optimization are included in (Maharramov and Biondi, 2015b; Maharramov et al., 2015c; Ma et al., 2015a,b). The regularization provided by bounded total-variation sometimes produces sufficient smoothing side-effect on the inverted model that obviates explicit bound constraints. However, many applications still require the imposition of additional constraints regardless of the regularization. For example, reservoir pore-pressure inversion problems often come with *a priori* bounds on the estimated pore pressure change, such as the pore pressure change being non-negative as a result of fluid injection (lower bound) or never exceeding a hydraulic fracturing pressure (upper bound). An example of such inversion for an unconventional reservoir from field tilt measurements is provided by Maharramov and Zoback (2014). TV regularization is a key tool in imaging and denoising applications (Rudin et al., 1992; Chambolle and Lions, 1997; Goldstein and Osher, 2009; Chartrand and Wohlberg, 2010) and require an efficient mechanism for including *a priori* model constraints that can significantly reduce model space (Chartrand and Wohlberg, 2010). While barrier or penalty function methods, such as nonlinear interior-point methods (Nocedal and Wright, 2006), can be used to tackle the

general constrained formulation (C.1), the presence of a non-differentiable L_1-norm total-variation term and non-quadratic penalty terms pose considerable challenges to practical implementation. A log-barrier function such as

$$- \text{ const } \times \sum_{i=1}^{n} \log \frac{m_2^i - m^i}{\delta} + \log \frac{m^i - m_1^i}{\delta}, \tag{C.2}$$

where n is the model space dimension, can be added to the right-hand side of the objective function to keep solution iterates away from the rectangular bounds. However, this adds a non-quadratic term to the objective function. For large-scale inversion problems with $n > 10^5$ (such as typical in geophysical applications) often only iterative gradient-based solution techniques like the nonlinear conjugate gradients (Nocedal and Wright, 2006) are available, and adding non-quadratic terms may significantly affect convergence properties. Note that this is in addition to the challenges associated with handling the non-differentiable TV-regularization term.

Chartrand and Wohlberg (2010) used a splitting approach to decouple the TV-regularized problem from enforcing the constraints. In my approach, we perform three-way splitting of problem (C.1) into a smooth optimization, gradient thresholding and projection steps using the Alternating Direction Method of Multipliers (ADMM) (Boyd et al., 2011). For unconstrained TV-regularized problems this approach is equivalent to the split-Bregman method of Goldstein and Osher (2009). However, we integrate the projection step associated with enforcing the bound constraints into the TV-minimization loop and avoid unnecessary iterations in the minimization of a proximal term (Parikh and Boyd, 2013) associated with the projection.

METHOD

First, we recast the TV-regularization part of (C.1) as a constrained optimization problem following the approach of Goldstein and Osher (2009). I introduce an auxiliary variable $\mathbf{x}$ and operator $\mathbf{\Phi} : \mathbf{m} \rightarrow \mathbf{x}$ such that for isotropic TV regularization

we have a vector of the model-space dimension

$$\mathbf{\Phi}(\mathbf{m}) \;=\; \sqrt{(\nabla_x \mathbf{m})^2 + (\nabla_y \mathbf{m})^2}, \tag{C.3}$$

and for anisotropic regularization a vector twice the model-space dimension

$$\mathbf{\Phi}(\mathbf{m}) \;=\; \begin{bmatrix} \nabla_x \mathbf{m} \\ \nabla_y \mathbf{m} \end{bmatrix}. \tag{C.4}$$

Problem (C.1) can now be reformulated with an additional equality constraint:

$$\begin{aligned} &\|\mathbf{x}\|_1 \;+\; \frac{\alpha}{2}\|\mathbf{F}(\mathbf{m}) - \mathbf{d}\|_2^2 \;\;\rightarrow\;\; \min, \\ &\mathbf{x} \;=\; \mathbf{\Phi}(\mathbf{m}), \\ &\mathbf{m}_1 \;\leq\; \mathbf{m} \;\leq\; \mathbf{m}_2. \end{aligned} \tag{C.5}$$

Problem (C.5) is still a bound-constrained problem. Introducing the projection operator

$$\mathbf{\Pi}(\mathbf{m}) \;=\; \max\{\min\{\mathbf{m}, \mathbf{m}_2\}, \mathbf{m}_1\}, \tag{C.6}$$

where min and max are applied component-wise, we reduce (C.5) to a fully equality-constrained formulation:

$$\begin{aligned} &\|\mathbf{x}\|_1 \;+\; \frac{\alpha}{2}\|\mathbf{F}(\mathbf{m}) - \mathbf{d}\|_2^2 \;\;\rightarrow\;\; \min, \\ &\mathbf{x} \;=\; \mathbf{\Phi}(\mathbf{m}), \\ &\mathbf{m} \;=\; \mathbf{y}, \\ &\mathbf{y} = \mathbf{\Pi}(\mathbf{m}). \end{aligned} \tag{C.7}$$

Following the augmented Lagrangian recipe for (C.7) while assuming the last equality constraint still enforced explicitly, we obtain a sequence of problems (Nocedal and

Wright, 2006)

$$\begin{aligned}
&(\mathbf{x}^{k+1}, \mathbf{m}^{k+1}) \;=\; \text{argmin } \|\mathbf{x}\|_1 \;+\; \frac{\alpha}{2}\|\mathbf{F}(\mathbf{m}) - \mathbf{d}\|_2^2 \;+ \\
&\frac{\lambda}{2}\|\mathbf{x} \;-\; \mathbf{\Phi}(\mathbf{m})\|_2^2 - \boldsymbol{\mu}_k{}^T\,(\mathbf{x} \;-\; \mathbf{\Phi}(\mathbf{m})) + \frac{\delta}{2}\|\mathbf{m} - \mathbf{y}\|_2^2 - \boldsymbol{\nu}_k{}^T\,(\mathbf{m} - \mathbf{y}) \;\rightarrow\; \text{min}, \\
&\boldsymbol{\mu}_{k+1} \;=\; \boldsymbol{\mu}_k - \lambda\left[\mathbf{x}^{k+1} - \mathbf{\Phi}(\mathbf{m}^{k+1})\right], \\
&\boldsymbol{\nu}_{k+1} \;=\; \boldsymbol{\nu}_k - \delta\left[\mathbf{m}^{k+1} - \mathbf{y}\right],\; k = 0, 1, 2, \ldots
\end{aligned} \tag{C.8}$$

Coefficients λ and δ are any positive constants above certain problem-specific "threshold" values (Nocedal and Wright, 2006), and can be selected experimentally. Vectors $\boldsymbol{\mu}_k$ and $\boldsymbol{\nu}_k$ are vectors of multipliers that converge to the set of Lagrange multipliers for the first two equality constraints of problem (C.7). At each step, (C.8) solves an L_1-regularized problem with respect to the combined model vector $(\mathbf{x}, \mathbf{m})$. Introducing new scaled multiplier vectors

$$\mathbf{b}^k \;=\; \frac{\boldsymbol{\mu}_k}{\lambda},\; \mathbf{c}^k \;=\; \frac{\boldsymbol{\nu}_k}{\delta},\; k = 0, 1, 2, \ldots \tag{C.9}$$

a little algebra shows that (C.8) is equivalent to

$$\begin{aligned}
(\mathbf{x}^{k+1},&\mathbf{m}^{k+1}) \;=\; \text{argmin } \|\mathbf{x}\|_1 \;+\; \frac{\alpha}{2}\|\mathbf{F}(\mathbf{m}) - \mathbf{d}\|_2^2 \;+ \\
&\frac{\lambda}{2}\|\mathbf{x} \;-\; \mathbf{\Phi}(\mathbf{m}) - \mathbf{b}^k\|_2^2 + \frac{\delta}{2}\|\mathbf{m} \;-\; \mathbf{y} - \mathbf{c}^k\|_2^2 \;\rightarrow\; \text{min}, \\
&\mathbf{b}^{k+1} \;=\; \mathbf{b}^k + \Phi(\mathbf{m}^{k+1}) - \mathbf{x}^{k+1}, \\
&\mathbf{c}^{k+1} \;=\; \mathbf{c}^k + \mathbf{y} - \mathbf{m}^{k+1},\; k = 0, 1, 2, \ldots
\end{aligned} \tag{C.10}$$

Here we used the fact that adding a constant term $\lambda/2\|\mathbf{b}^k\|_2^2 + \delta/2\|\mathbf{c}^k\|_2^2$ to the objective function obviously does not change the minimizing solution.

Problem (C.7) can be solved by iteratively projecting the current model vector $\mathbf{m}$ onto $\mathbf{y}$, then conducting the iterations (C.10) to convergence, then repeating the process. However, presence of the proximal term $\delta/2\|\mathbf{m} - \mathbf{y}\|_2^2$ in (C.8) due to the constraint $\mathbf{m} = \mathbf{y}$ means that a very accurate solution of (C.10) at early iterations is wasteful and unnecessary. I instead carry out a single iteration of (C.10) followed by

the model projection:

$$\begin{aligned}
(\mathbf{x}^{k+1},\mathbf{m}^{k+1}) \;&=\; \text{argmin}\; \|\mathbf{x}\|_1 \;+\; \frac{\alpha}{2}\|\mathbf{F}(\mathbf{m})-\mathbf{d}\|_2^2 \;+ \\
&\frac{\lambda}{2}\|\mathbf{x} \;-\; \mathbf{\Phi}(\mathbf{m})-\mathbf{b}^k\|_2^2+\frac{\delta}{2}\|\mathbf{m} \;-\; \mathbf{y}^k-\mathbf{c}^k\|_2^2 \;\rightarrow\; \min, \\
&\mathbf{b}^{k+1} \;=\; \mathbf{b}^k+\Phi(\mathbf{m}^{k+1})-\mathbf{x}^{k+1}, \\
&\mathbf{c}^{k+1} \;=\; \mathbf{c}^k+\mathbf{y}^k-\mathbf{m}^{k+1}, \\
&\mathbf{y}^{k+1} \;=\; \mathbf{\Pi}(\mathbf{m}^{k+1}) \;=\; \max\{\min\{\mathbf{m}^{k+1},\mathbf{m}_2\},\mathbf{m}_1\},\; k=0,1,2,\ldots
\end{aligned} \tag{C.11}$$

The iterative process (C.11) still requires solving an L_1-regularized problem. However, the L_1-norm term now involves only the vector $\mathbf{x}$. Therefore, we apply splitting, minimizing

$$\|\mathbf{x}\|_1 \;+\; \frac{\alpha}{2}\|\mathbf{F}(\mathbf{m})-\mathbf{d}\|_2^2 \;+\frac{\lambda}{2}\|\mathbf{x} \;-\; \mathbf{\Phi}(\mathbf{m})-\mathbf{b}^k\|_2^2+\frac{\delta}{2}\|\mathbf{m} \;-\; \mathbf{y}^k-\mathbf{c}^k\|_2^2 \tag{C.12}$$

alternately with respect to $\mathbf{m}$ and $\mathbf{x}$ in an inner loop of $N_1 \geq 1$ cycles. Because the proximal constraint $\mathbf{m}=\mathbf{y}$ renders good fitting accuracy at early stages unnecessary, N_1 can be small. Further we note that the minimization of (C.12) with respect to $\mathbf{x}$ (in a splitting step with $\mathbf{m}$ fixed) is given trivially by the "shrinkage" operator (Goldstein and Osher, 2009):

$$\mathbf{x}^{k+1} \;=\; \text{shrink}\left\{\mathbf{\Phi}(\mathbf{m})+\mathbf{b}^k,\frac{1}{\lambda}\right\}, \tag{C.13}$$

where

$$\text{shrink}\left\{\mathbf{x},\gamma\right\} \;=\; \frac{\mathbf{x}}{|\mathbf{x}|}\max\left(|\mathbf{x}|-\gamma,0\right), \tag{C.14}$$

and is effectively thresholding the model gradient. Our algorithm can be described by the following five steps:

1 Initialization

$$\begin{aligned} \mathbf{m}^0 &= \text{starting guess}, \\ \mathbf{x}^0 &= \mathbf{0}, \\ \mathbf{y}^0 &= \max\{\min\{\mathbf{m}^0, \mathbf{m}_2\}, \mathbf{m}_1\}, \\ \mathbf{b}^0 &= \mathbf{0}, \\ \boldsymbol{c}^0 &= \mathbf{0}, \end{aligned} \tag{C.15}$$

2 **Outer loop**. Repeat steps 3-5 for $k = 0, 1, 2, \ldots$

3 **Inner loop**. Iterate (C.16) $N_1 \geq 1$ times.

$$\begin{aligned} \mathbf{m}^{k+1} &= \text{argmin } \frac{\lambda}{2}\|\mathbf{x}^k - \mathbf{\Phi}(\mathbf{m}) - \mathbf{b}^k\|_2^2 + \frac{\alpha}{2}\|\mathbf{F}(\mathbf{m}) - \mathbf{d}\|_2^2 + \\ &\quad \frac{\delta}{2}\|\mathbf{m} - \mathbf{y}^k - \boldsymbol{c}^k\|_2^2, \\ \mathbf{x}^{k+1} &= \text{shrink}\left\{\mathbf{\Phi}(\mathbf{m}^{k+1}) + \mathbf{b}^k, \frac{1}{\lambda}\right\}, \ \mathbf{x}^k = \mathbf{x}^{k+1}, \end{aligned} \tag{C.16}$$

4 Update the multipliers and project the model onto the bounding rectangle:

$$\begin{aligned} \mathbf{b}^{k+1} &= \mathbf{b}^k + \mathbf{\Phi}(\mathbf{m}^{k+1}) - \mathbf{x}^{k+1}, \\ \boldsymbol{c}^{k+1} &= \boldsymbol{c}^k + \mathbf{y}^k - \mathbf{m}^{k+1}, \\ \mathbf{y}^{k+1} &= \max\{\min\{\mathbf{m}^{k+1}, \mathbf{m}_2\}, \mathbf{m}_1\}. \end{aligned} \tag{C.17}$$

5 Terminate if the target accuracy is reached

$$\frac{\|\mathbf{m}^{k+1} - \mathbf{m}^k\|_2}{\|\mathbf{m}^k\|} \leq \text{target accuracy}. \tag{C.18}$$

or go back to step 2 otherwise.

Optimizing (C.16) with respect to $\mathbf{m}$ is in itself a large-scale optimization problem, nonlinear for a nonlinear modeling operator $\mathbf{F}$. Solving the optimization problem (C.16) exactly is unnecessary because for small k (i.e., at early stages of the inversion) vector $\mathbf{y}^k$ is not the true model, vector $\mathbf{x}^k$ is far from the true model gradient, and

the multipliers $\mathbf{b}^k, \mathbf{x}^k$ could be far from scaled Lagrange multipliers. Therefore, for large-scale problems only a few steps of an iterative method like conjugate gradients need be carried out. As the solution converges to the true solution and critical sharp contrasts in the model are identified, an iterative solver can begin to take advantage of the objective function curvature information collected at previous iterations of the *outer* loop, potentially leading to a significantly faster convergence.

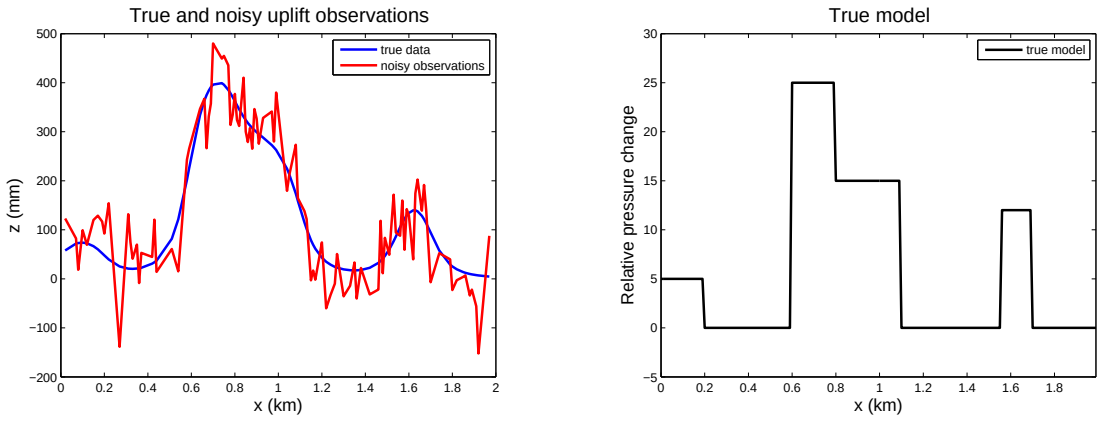
RESULTS

I demonstrate the method on a test problem that simulates vertical surface uplift in response to distributed dilatational sources, mathematically equivalent to surface deformation due to pore pressure change (Segall, 2010). Our modeling operator is defined as

$$\mathbf{F}(\mathbf{m}) \;=\; u(x),\; u(x) \;=\; const \int_0^A \frac{Dm(\xi)d\xi}{\left(D^2 + (x-\xi)^2\right)^{3/2}}, \tag{C.19}$$

where we assume that $\mathbf{m} = m(\xi), \xi \in [0, A]$ is a relative pore pressure change along a linear segment $[0, A]$ of a reservoir at a constant depth D, and $\mathbf{u} = u(x), x \in [0, A]$, within a proportionality factor determined by poroelastic medium properties (Maharramov, 2012), is the induced vertical uplift on the surface. For demonstration purposes we consider a one-dimensional model but the results trivially extend to realistic reservoir and surface geometries. Operator (C.19) is a smoothing operator, and recovering sharp pressure contrasts e.g. due to permeability barriers requires model "styling" or regularization such as blockiness-promoting ROF model. As a true model I used a highly compartmentalized pressure model of Figure C.1(b). In this experiments, I set $D = 100$ m $A = 2$ km, and discretized both the model and data space using a 200-point uniform grid. Random Gaussian noise with $\sigma = 15\%$ of the maximum noise-free data amplitude was added to the clean forward-modeled data to produce the noisy observations shown in Figure C.1(a).

The result of a TV-regularized *unconstrained* inversion is shown in Figure C.2(a) against the true model and a Tikhonov-regularized inversion. This result was obtained using the above algorithm by setting $\delta = 0$ (no bound constraints) and using



(a) (b)

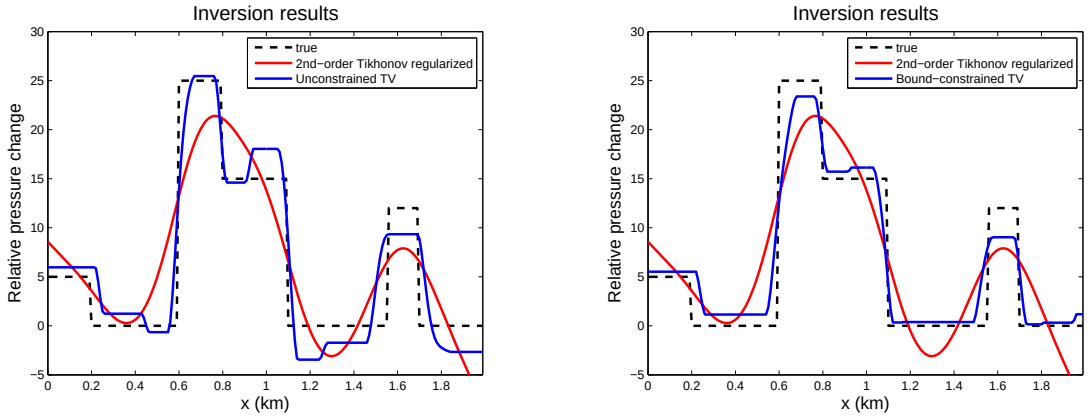
Figure C.1: (a) True and noisy uplift observations. Random Gaussian noise with $\sigma =$ 15% of maximum clean data amplitude was added to the clean data. (b) True model exhibits a highly compartmentalized "blocky" behavior. [**ER**] btv/. tvdata,true

the values of $\alpha = 1$ and $\lambda = 2$. The TV-regularized result captures the compartmentalized picture of pressure distribution better than the highly smoothed Tikhonov regularization result. However, due to absence of bound constraints, lower pressure bounds are not honored, resulting in negative pressure areas that are not present in the true model. The result of running the new bound-constrained TV-regularization algorithm is shown in Figure C.2(b). The imposition of bound constraints not only removed the negative relative pressure areas, but also removed the pressure spike at approximately $x \approx 1$ km in the unconstrained inversion of Figure C.2(a) that apparently had resulted from compensating negative pressures. In both the constrained and the unconstrained runs I conducted 1000 outer loop iterations with 2 inner loops cycles. However, the algorithm converged quickly, with only a few initial iterates outside a tight neighborhood of the final curve, as shown in Figure C.3(b). Finally, note that many practical implementations of bound constraints often resort to a simplistic way of enforcing the constraints: the inverted model is projected onto the bounding rectangle either once after applying a direct unconstrained solver, or

at each iteration of an unconstrained solver. In this case variable $\mathbf{y}$ and the associated quadratic regularization term are not introduced into the objective function. This may result in a violation of the KKT optimality conditions where the bound constraints are active (Nocedal and Wright, 2006), and is demonstrated by the blue plot in Figure C.3(a). While the bound constraints are honored, the solution is both qualitatively and quantitatively far from optimal.

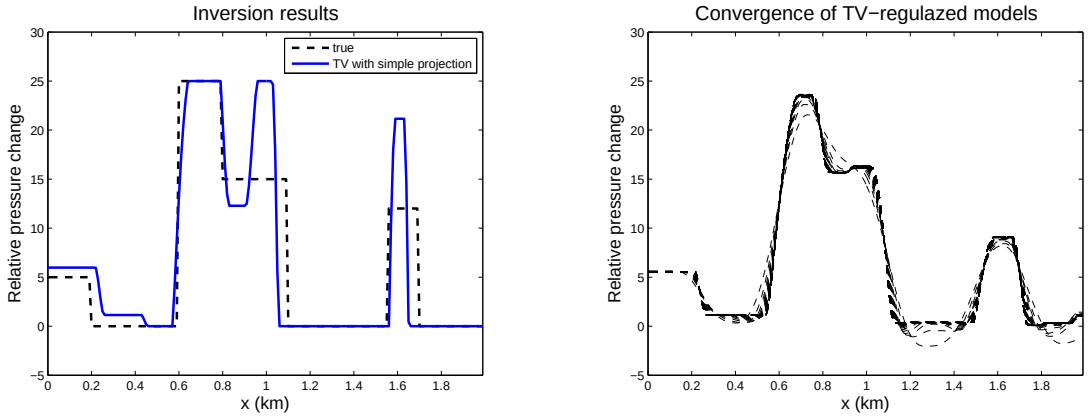
CONCLUSIONS AND PERSPECTIVES

The algorithm for approximate TV-regularized inversion with bound constraints combines the advantages of the blockiness-promoting and edge-preserving ROF model with the ability to impose bound constraints. The splitting mechanism used for enforcing the bound constraints is naturally integrated into the ADMM/split-Bregman iterations and results in no extra computational cost. The method was able to resolve compartmentalized subsurface pressure changes from noisy surface uplift observations despite the highly diffusive nature of the underlying deformation process. The method can be implemented around any large-scale nonlinear solver such as conjugate gradients or quasi-Newton methods. Additional equality and inequality constraints can be incorporated into the algorithm using the general ADMM framework. Solutions provided by this algorithm are only approximations to the true solution of problem (C.1). However, progressively more accurate solutions of (C.1) can be obtained by gradually reducing parameter δ. A variable-δ version of algorithm (C.15–C.18) for finding true solutions of (C.1) merits further investigation.



(a) (b)

Figure C.2: (a) Unconstrained TV-regularized inversion. The algorithm tries to fit the data by allowing negative relative pressure changes. (b) Bound constrained TV-regularized result. Note that enforcing lower bounds resulted in a more accurate shape matching of the true model. [**ER**] btv/. tvinv,boundtvinv



(a) (b)

Figure C.3: (a) Direct imposition of the bound constraints at each iteration of the unconstrained solver resulted in a qualitatively and quantitatively wrong inversion. (b) Convergence of TV-regularized inverted models with bound constraints. The method quickly resolves both sharp contrasts and active bounds as only a few initial curves out of 1000 iterates lie outside a small neighborhood of the final curve. [**ER**] btv/. simplefailboundtvinv,boundtvconv

Appendix D

Useful functions and equations

"It's the little things that smoothes people's roads the most"

Mark Twain

This appendix is a reference of functions and equations used in the thesis.

MINDLIN'S ELASTOSTATIC TENSOR

The elastostatic Green's tensor $g_i^k(x, y, z, \xi, \eta, \zeta)$ has the meaning of the displacement along axis i at point (x, y, z) due to a concentrated force along axis k at point (ξ, η, ζ). The analytical expression for the components of the Green's tensor in the elastic half-space with a free-surface boundary condition are given by the following equations (Mindlin, 1936):

$$
\begin{aligned}
g_1^1 =& w\left(\frac{3-4\nu}{r_1}+\frac{1}{r_2}+\frac{(x-\xi)}{r_1^3}+\frac{(3-4\nu)(x-\xi)^2}{r_2^3}\right)+\\
&+w\left(\frac{2(r_2^2-3(x-\xi)^2)z\zeta}{r_2^5}+\frac{4(1-\nu)(1-2\nu)(r_2^2-(x-\xi)^2-r_2(z+\zeta))}{r_2(r_2-z-\zeta)^2}\right)\\
g_2^1 =& (x-\xi)(y-\eta)w\left(\frac{1}{r_1^3}+\frac{3-4\nu}{r_2^3}-\frac{6z\zeta}{r_2^5}-\frac{4(1-\nu)(1-2\nu)}{r_2(r_2-z-\zeta)^2}\right)\\
g_3^1 =& (x-\xi)w\left(\frac{z-\zeta}{r_1^3}+\frac{(3-4\nu)(z-\zeta)}{r_2^3}-\frac{4(1-\nu)(1-2\nu)}{r_2(r_2-z-\zeta)}-\frac{6z\zeta(z+\zeta)}{r_2^5}\right)
\end{aligned}
$$

$$
\begin{aligned}
g_1^2 =& g_2^1\\
g_2^2 =& w\left(\frac{3-4\nu}{r_1}+\frac{1}{r_2}+\frac{(y-\eta)}{r_1^3}+\frac{(3-4\nu)(y-\eta)^2}{r_2^3}\right)+\\
&+w\left(\frac{2(r_2^2-3(y-\eta)^2)z\zeta}{r_2^5}+\frac{4(1-\nu)(1-2\nu)(r_2^2-(y-\eta)^2-r_2(z+\zeta))}{r_2(r_2-z-\zeta)^2}\right)\\
g_3^2 =& (y-\eta)w\left(\frac{z-\zeta}{r_1^3}+\frac{(3-4\nu)(z-\zeta)}{r_2^3}-\frac{4(1-\nu)(1-2\nu)}{r_2(r_2-z-\zeta)}-\frac{6z\zeta(z+\zeta)}{r_2^5}\right)
\end{aligned}
$$

$$
\begin{aligned}
g_1^3 =& (x-\xi)w\left(\frac{z-\zeta}{r_1^3}+\frac{(3-4\nu)(z-\zeta)}{r_2^3}+\frac{4(1-\nu)(1-2\nu)}{r_2(r_2-z-\zeta)}+\frac{6z\zeta(z+\zeta)}{r_2^5}\right)\\
g_2^3 =& (y-\eta)w\left(\frac{z-\zeta}{r_1^3}+\frac{(3-4\nu)(z-\zeta)}{r_2^3}+\frac{4(1-\nu)(1-2\nu)}{r_2(r_2-z-\zeta)}+\frac{6z\zeta(z+\zeta)}{r_2^5}\right)\\
g_3^3 =& w\left(\frac{3-4\nu}{r_1}+\frac{5-12\nu+8\nu^2}{r_2}+\frac{(z-\zeta)^2}{r_1^3}+\frac{(3-4\nu)(z+\zeta)^2-2z\zeta}{r_2^3}+\frac{6z\zeta(z+\zeta)^2}{r_2^5}\right)
\end{aligned}
$$

$$
\begin{aligned}
w =& \frac{1}{16\pi\mu(1-\nu)} \\
r_1 =& \sqrt{(x-\xi)^2+(y-\eta)^2+(z-\zeta)^2} \\
r_2 =& \sqrt{(x-\xi)^2+(y-\eta)^2+(z+\zeta)^2}
\end{aligned}
$$

Bibliography

Ascher, U. M., 2008, Numerical methods for evolutionary differential equations: Society for Industrial and Applied Mathematics.

Asnaashari, A., R. Brossier, S. Garambois, F. Audebert, P. Thore, and J. Virieux, 2012, Time-lapse imaging using regularized FWI: A robustness study: 82nd Annual International Meeting, SEG, Technical Program Expanded Abstracts, doi:10.1190/segam2012-0699.1, 1–5.

Aster, R., B. Borchers, and C. H. Thurber, 2011, Parameter estimation and inverse problems: Academic Press.

Auer, L., A. M. Nuber, S. A. Greenhalgh, H. Maurer, and S. Marelli, 2013, A critical appraisal of asymptotic 3D-to-2D data transformation in full-waveform seismic crosshole tomography: Geophysics, **78**, no. 6, R235–R247.

Ayeni, G., 2011, Time-lapse seismic imaging by linearized joint inversion: PhD thesis, Stanford University.

Ayeni, G., and B. Biondi, 2012, Time-lapse seismic imaging by linearized joint inversion–A Valhall Field case study: 82nd Annual International Meeting, SEG, Technical Program Expanded Abstracts, doi:10.1190/segam2012-0903.1, 1–6.

Bamberger, A., G. Chavent, and P. Lailly, 1977, Une application de la théorie du contrôle à un problème inverse de sismique: Annales de Geophysique, **33**, 183–200.

Barkved, O. I., and T. Kristiansen, 2005, Seismic time-lapse effects and stress changes: Examples from a compacting reservoir: The Leading Edge, **24**, 1244–1248.

Bauschke, H. H., and P. L. Combettes, 2011, Convex analysis and monotone operator theory in Hilbert spaces: Springer.

Beck, A., and M. Teboulle, 2009a, Fast gradient-based algorithms for constrained

total variation image denoising and deblurring problems: IEEE Transactions on Image Processing, **18**, 2419–2434.

———, 2009b, A fast iterative shrinkage-thresholding algorithm for linear inverse problems: SIAM Journal on Imaging Sciences, **2**, 183–202.

Biot, M. A., 1941, General theory of three-dimensional consolidation: Journal of Applied Physics, **12**, 155–164.

Bioucas-Dias, J. M., and M. A. Figueiredo, 2007, A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration: IEEE Transactions on Image Processing, **16**, 2992–3004.

Björk, A., 1996, Numerical methods for least squares problems: Society for Industrial and Applied Mathematics.

Boyd, S., N. Parikh, E. Chu, B. Peleato, and J. Eckstein, 2011, Distributed optimization and statistical learning via the Alternating Direction Method of Multipliers: Foundations and Trends in Machine Learning, **3**, 1–122.

Boyd, S. P., and L. Vandenberghe, 2004, Convex optimization: Cambridge University Press.

Bruck Jr., R. E., 1977, On the weak convergence of an ergodic iteration for the solution of variational inequalities for monotone operators in Hilbert space: Journal of Mathematical Analysis and Applications, **61**, 159 – 164.

Cameron, M. K., S. B. Fomel, and J. A. Sethian, 2007, Seismic velocity estimation from time migration: Inverse Problems, **23**, 1329.

Chambolle, A., 2004, An algorithm for total variation minimization and applications: Journal of Mathematical Imaging and Vision, **20**, 89–97.

Chambolle, A., and P. L. Lions, 1997, Image recovery via total variational minimization and related problems: Numerische Mathematik, **76**, 167–188.

Chartrand, R., and B. Wohlberg, 2010, Total-variation regularization with bound constraints: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 766–769.

Chavent, G., 1971, Analyse fonctionnelle et identification de coefficients répartis dans les équations aux dérivées partielles: PhD thesis, Université de Paris.

Colton, D., and R. Kress, 1998, Inverse acoustic and electromagnetic scattering theory, Applied Mathematical Sciences, 2nd ed.: Springer.

Combettes, P. L., and V. R. Wajs, 2005, Signal recovery by proximal forward-backward splitting: Multiscale Modeling & Simulation, **4**, 1168–1200.

Daubechies, I., M. Defrise, and C. De Mol, 2004, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint: Communications on Pure and Applied Mathematics, **57**, 1413–1457.

Denli, H., and L. Huang, 2009, Double-difference elastic waveform tomography in the time domain: 79th Annual International Meeting, SEG, Technical Program Expanded Abstracts, 2302–2306.

Douglas, J., and H. H. Rachford, 1956, On the numerical solution of heat conduction problems in two and three space variables: Transactions of the American Mathematical Society, **82**, 421–439.

Du, J., and J. E. Olson, 2001, A poroelastic reservoir model for predicting subsidence and mapping subsurface pressure fronts: Journal of Petroleum Science and Engineering, **30**, 181 – 197.

Duan, R., and V. Rokhlin, 2009, High-order quadratures for the solution of scattering problems in two dimensions: Journal of Computational Physics, **228**, 2152–2174.

Dvorkin, J., M. A. Gutierrez, and D. Grana, 2014, Seismic reflections of rock properties: Cambridge University Press.

Eckstein, J., and D. P. Bertsekas, 1992, On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators: Mathematical Programming, **55**, 293–318.

Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani, 2004, Least angle regression: The Annals of Statistics, **32**, 407–499.

Engquist, B., and A. Majda, 1977, Absorbing boundary conditions for the numerical simulation of waves: Mathematics of Computation, **31**, 629–651.

Etgen, J. T., 2012, 3D wave equation Kirchhoff migration: 82nd Annual International Meeting, SEG, Technical Program Expanded Abstracts, doi:10.1190/segam2012-0755.1, 1–5.

Fichtner, A., 2011, Full seismic modeling and inversion: Springer.

Figueiredo, M. A. T., R. D. Nowak, and S. J. Wright, 2007, Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems: IEEE Journal of Selected Topics in Signal Processing, **1**, 586–597.

Gabay, D., and B. Mercier, 1976, A dual algorithm for the solution of nonlinear variational problems via finite element approximation: Computers & Mathematics with Applications, **2**, 17–40.

Geertsma, J., 1973, Land subsidence above compacting oil and gas reservoirs: SPE-AIME European Spring Meeting.

Gelfand, I. M., and S. V. Fomin, 2000, Calculus of variations: Dover.

Glowinski, R., and P. Le Tallec, 1989, Augmented Lagrangian and operator-splitting methods in nonlinear mechanics: Society for Industrial and Applied Mathematics.

Glowinski, R., and A. Marroco, 1975, Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires: ESAIM: Modélisation Mathématique et Analyse Numérique, **9**, 41–76.

Goldstein, T., B. O'Donoghue, S. Setzer, and R. Baraniuk, 2014, Fast Alternating Direction Optimization Methods: SIAM Journal on Imaging Sciences, **7**, 1588–1623.

Goldstein, T., and S. Osher, 2009, The split Bregman method for L1-regularized problems: SIAM Journal on Imaging Sciences, **2**, 323–343.

Golub, G. H., and C. F. Van Loan, 1996, Matrix computations, 3rd ed.: Johns Hopkins University Press.

Hansen, P. C., 1998, Rank-deficient and discrete ill-posed problems: Numerical aspects of linear inversion: Society for Industrial and Applied Mathematics.

Hastie, T., S. Rosset, R. Tibshirani, and J. Zhu, 2004, The entire regularization path for the support vector machine: Journal of Machine Learning Research, **5**, 1391–1415.

Hatchell, P., and S. Bourne, 2005, Measuring reservoir compaction using time-lapse timeshifts: 75th Annual International Meeting, SEG, Technical Program Expanded Abstracts, 2500–2503.

He, B., and X. Yuan, 2012, On the $O(1/n)$ convergence rate of the Douglas-Rachford

alternating direction method: SIAM Journal on Numerical Analysis, **50**, 700–709.

Herwanger, J. V., 2008, R we there yet?: 70th EAGE Conference and Exhibition, Extended Abstract I029.

Hestenes, M. R., 1969, Multiplier and gradient methods: Journal of Optimization Theory and Applications, **4**, 303–320.

Hinkle, A., and M. Batzle, 2006, Heavy oils: A worldwide overview: The Leading Edge, **25**, 742–749.

Hodgson, N., C. Macbeth, L. Duranti, J. Rickett, and K. Nihei, 2007, Inverting for reservoir pressure change using time-lapse time strain: Application to Genesis Field, Gulf of Mexico: The Leading Edge, **26**, 649–652.

Holt, R. M., O.-M. Nes, and E. Fjaer, 2005, In-situ stress dependence of wave velocities in reservoir and overburden rocks: The Leading Edge, **24**, 1268–1274.

Hudson, T., B. Regel, J. Bretches, P. Condon, J. Rickett, B. Cerney, and P. Inderwiesen, 2005, Genesis Field, Gulf of Mexico, 4D project status and preliminary lookback: 75th Annual International Meeting, SEG, Technical Program Expanded Abstracts, 2436–2439.

Iserles, A., 2008, A first course in the numerical analysis of differential equations: Cambridge University Press.

Ishimaru, A., 1978, Wave propagation and scattering in random media: Academic Press.

Johnston, D., 2013, Practical applications of time-lapse seismic data: Society of Exploration Geophysicists.

Kantorovich, L. V., and G. P. Akilov, 1982, Functional analysis: Pergamon Press.

Kim, S., K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, 2007, An interior-point method for large-scale ℓ_1-regularized least squares: IEEE Journal of Selected Topics in Signal Processing, **1**, 606–617.

Kolmogorov, A. N., and S. V. Fomin, 1999, Elements of the theory of functions and functional analysis: Dover.

Kosloff, D., R. Scott, and J. Scranton, 1980, Finite element simulation of Wilmington oil field subsidence: I. Linear modelling: Tectonophysics, **65**, 339 – 368.

Lions, J. L., 1971, Optimal control of systems governed by partial differential equations: Springer.

Lumley, D. E., 1995, Seismic time-lapse monitoring of subsurface fluid flow: PhD thesis, Stanford University.

Ma, Y., M. Maharramov, R. Clapp, and B. Biondi, 2015a, Illumination compensation by L1 regularization and steering filters: Stanford Exploration Project Report, **158**, 279–290.

———, 2015b, Time lapse seismic imaging with L1 regularization and steering filters: Stanford Exploration Project Report, **158**, 291–310.

Magesan, M., S. Depagne, K. Nixon, B. Regel, J. Opich, G. Rogers, and T. Hudson, 2005, Seismic processing for time-lapse study: Genesis Field, Gulf of Mexico: The Leading Edge, **24**, 364–373.

Maharramov, M., 2012, Identifying reservoir depletion patterns from production-induced deformations with applications to seismic imaging: Stanford Exploration Project Report, **147**, 193–211.

Maharramov, M., and B. Biondi, 2013, Simultaneous time-lapse full waveform inversion: Stanford Exploration Project Report, **150**, 63–70.

———, 2014a, Improved depth imaging by constrained full-waveform inversion: Stanford Exploration Project Report, **155**, 17–21.

———, 2014b, Joint 4DFWI with model-difference regularization: SEG-AGU Summer Research Workshop. Advances in Active+Passive "Full Wavefield" Seismic Imaging: From Reservoirs to Plate Tectonics.

———, 2014c, Joint full-waveform inversion of time-lapse seismic data sets: 84th Annual International Meeting, SEG, Technical Program Expanded Abstracts, doi:10.1190/segam2014-0962.1, 954–959.

———, 2014d, Joint full-waveform inversion of time-lapse seismic data sets: Stanford Exploration Project Report, **152**, 19–28.

———, 2014e, Robust joint full-waveform inversion of time-lapse seismic datasets with total-variation regularization: Stanford Exploration Project Report, **155**, 199–208.

———, 2015a, Multi-scale inversion of subsurface velocity models using cartoon-texture decomposition: SIAM Conference on Mathematical and Computational

Issues in the Geosciences, MS2 Full Waveform Inversion I: Algorithms and Performance.

———, 2015b, Resolving the effects of production-induced overburden dilation using simultaneous TV-regularized time-lapse FWI: Stanford Exploration Project Report, **158**, 1–9.

Maharramov, M., B. Biondi, and M. Meadows, 2015a, Simultaneous TV-regularized time-lapse FWI with application to field data: 85th Annual International Meeting, SEG, Technical Program Expanded Abstracts, doi:10.1190/segam2015-5925444.1, 1236–1241.

Maharramov, M., B. Biondi, and S. Ronen, 2015b, Robust simultaneous time-lapse full-waveform inversion with total-variation regularization of model difference: 77th EAGE Conference and Exhibition, Extended Abstract We P3 09.

Maharramov, M., and S. Levin, 2015, Total-variation minimization with bound constraints: Stanford Exploration Project Report, **158**, 271–278.

Maharramov, M., Y. Ma, and B. Biondi, 2015c, Double-difference time-lapse FWI with a total-variation regularization: Stanford Exploration Project Report, **158**, 263–270.

Maharramov, M., and M. Zoback, 2014, Monitoring of cyclic steam stimulation by inversion of surface tilt measurements: AGU Fall Meeting, Session H23A-0859.

Maharramov, M., and M. D. Zoback, 2015, Characterization of reservoir heterogeneity from surface deformation: SRB Annual Report.

Maslov, V. P., 1976, Operational methods: Mir Publishers.

———, 1989-1990, Asymptotic solutions of equations with slowly varying coefficients: Ecole Polytechnique, Centre de Mathématiques, Séminaire Équations aux dérivées partielles, 1–20.

Mavko, G., T. Mukerji, and J. Dvorkin, 2009, The rock physics handbook: Cambridge University Press.

Meyer, Y., 2001, Oscillating patterns in image processing and nonlinear evolution equations. The Fifteenth Dean Jacqueline B. Lewis Memorial Lectures: American Mathematical Society.

Mindlin, R. D., 1936, Force at a point in the interior of a semi-infinite solid: Physics,

7, 195–202.

Nesterov, Y. E., 1983, A method for solving the convex programming problem with rate of convergence $O(1/k^2)$: Doklady Akademii Nauk SSSR, **269**, 543–547.

Nocedal, J., and S. J. Wright, 2006, Numerical optimization: Springer.

Nur, A. M., 1982, Seismic imaging in enhanced recovery: SPE Enhanced Oil Recovery Symposium, 4-7 April, Tulsa, Oklahoma, SPE-10680-MS.

Nur, A. M., C. Tosaya, and D. Vo-Thanh, 1984, Seismic monitoring of thermal enhanced oil recovery processes: 54th Annual Meeting, SEG, Technical Program Expanded Abstracts, 337–340.

Osborne, M. R., B. Presnell, and B. A. Turlach, 2000, A new approach to variable selection in least squares problems: IMA Journal of Numerical Analysis, **20**, 389–403.

Parikh, N., and S. Boyd, 2013, Proximal algorithms: Foundations and Trends in Optimization, **1**, 123–231.

Passty, G. B., 1979, Ergodic convergence to a zero of the sum of monotone operators in Hilbert space: Journal of Mathematical Analysis and Applications, **72**, 383 – 390.

Qiu, Y., W. Xue, and G. Yu, 2012, A projected conjugate gradient method for compressive sensing, *in* Intelligent Science and Intelligent Data Engineering: Third Sino-foreign-interchange Workshop, IScIDE 2012, Nanjing, China, October 15-17, 2012. Revised Selected Papers: Springer Berlin Heidelberg, 398–406.

Raknes, E., W. Weibull, and B. Arntsen, 2013, Time-lapse full waveform inversion: Synthetic and real data examples: 83rd Annual International Meeting, SEG, Technical Program Expanded Abstracts, doi:10.1190/segam2013-0540.1, 944–948.

Rice, J. R., and M. P. Cleary, 1976, Some basic stress diffusion solutions for fluid-saturated elastic porous media with compressible constituents: Reviews of Geophysics, **14**, 227–241.

Rickett, J., L. Duranti, T. Hudson, and N. Hodgson, 2006, Compaction and 4D time strain at the Genesis Field: 76th Annual International Meeting, SEG, Technical Program Expanded Abstracts, 3215–3219.

Rickett, J., L. Duranti, T. Hudson, B. Regel, and N. Hodgson, 2007, 4D time strain

and the seismic signature of geomechanical compaction at Genesis: The Leading Edge, **26**, 644–647.

Rockafellar, R. T., 1971, Convex analysis: Princeton University Press.

———, 1976, Augmented Lagrangians and applications of the proximal point algorithm in convex programming: Mathematics of Operations Research, **1**, 97–116.

Routh, P., G. Palacharla, I. Chikichev, and S. Lazaratos, 2012, Full wavefield inversion of time-lapse data for improved imaging and reservoir characterization: 82nd Annual International Meeting, SEG, Technical Program Expanded Abstracts, doi:10.1190/segam2012-1043.1, 1–6.

Rudin, L. I., S. Osher, and E. Fatemi, 1992, Nonlinear total variation based noise removal algorithms: Physica D: Nonlinear Phenomena, **60**, 259–268.

Rytov, S. M., Y. A. Kravtsov, and V. I. Tatarski, 1989, Principles of statistical radiophysics 4: Wave propagation through random media: Springer.

Saad, Y., 2003, Iterative methods for sparse linear systems, second edition: Society for Industrial and Applied Mathematics.

Segall, P., 1985, Stress and subsidence resulting from subsurface fluid withdrawal in the epicentral region of the 1983 Coalinga Earthquake: Journal of Geophysical Research: Solid Earth, **90**, 6801–6816.

———, 1992, Induced stresses due to fluid extraction from axisymmetric reservoirs: Pure and Applied Geophysics, **139**, 535–560.

———, 2010, Earthquake and volcano deformation: Princeton University Press.

Segall, P., and S. D. Fitzgerald, 1998, A note on induced stress changes in hydrocarbon and geothermal reservoirs: Tectonophysics, **289**, 117 – 128.

Segall, P., J.-R. Grasso, and A. Mossop, 1994, Poroelastic stressing and induced seismicity near the Lacq gas field, southwestern France: Journal of Geophysical Research, **99**, 15423–15438.

Sirgue, L., O. I. Barkved, J. Dellinger, J. Etgen, U. Albertin, and J. Kommedndal, 2010a, Full waveform inversion: the next leap forward in imaging at Valhall: First Break, **28**, 65–70.

Sirgue, L., J. Etgen, U. Albertin, and S. Brandsberg-Dahl, 2010b, System and method for 3D frequency domain waveform inversion based on 3D time-domain forward

modeling. (US Patent 7,725,266).

Sirgue, L., J. T. Etgen, and U. Albertin, 2008, 3D frequency domain waveform inversion using time domain finite difference methods: 70th EAGE Conference and Exhibition, Extended Abstract F022.

Sirgue, L., and R. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: Geophysics, **69**, 231–248.

Slaney, M., A. C. Kak, and L. E. Larsen, 1984, Limitations of imaging with first-order diffraction tomography: IEEE Transactions on Microwave Theory and Techniques, **32**, 860–874.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266.

Tikhonov, A. N., and V. Y. Arsenin, 1977, Solution of ill-posed problems: Winston & Sons.

Trefethen, L. N., and D. Bau III, 1997, Numerical linear algebra: Society for Industrial and Applied Mathematics.

Uzawa, H., 1958, Iterative methods for concave programming, *in* Studies in Linear and Non-Linear Programming: Stanford University Press.

Vasco, D. W., K. Karasaki, and C. Doughty, 2000, Using surface deformation to image reservoir dynamics: Geophysics, **65**, 132–147.

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

Vladimirov, V. S., 1971, Equations of mathematical physics: Marcel Dekker.

Vogel, C. R., and M. E. Oman, 1996, Iterative methods for total variation denoising: SIAM Journal on Scientific Computing, **17**, 227–238.

Walters, R., and M. Zoback, 2013, Microseismicity and surface deformation of a heavy-oil reservoir undergoing cyclic steam stimulation: 83rd Annual International Meeting, SEG, Technical Program Expanded Abstracts, 2218–2222.

Wang, H., 2000, Theory of linear poroelasticity with applications to geomechanics and hydrogeology: Princeton University Press.

Watanabe, T., S. Shimizu, E. Asakawa, and T. Matsuoka, 2004, Differential waveform tomography for time-lapse crosswell seismic data with application to gas hydrate

production monitoring: 74th Annual International Meeting, SEG, Technical Program Expanded Abstracts, 2323–2326.

Woodward, M. J., 1989, Wave equation tomography: PhD thesis, Stanford University.

Wu, R.-S., and M. N. Toksöz, 1987, Diffraction tomography and multisource holography applied to seismic imaging: Geophysics, **52**, 11–25.

Yang, D., A. E. Malcolm, and M. C. Fehler, 2014, Time-lapse full waveform inversion and uncertainty analysis with different survey geometries: 76th EAGE Conference and Exhibition, Extended Abstract We ELI1 10.

Ying, L., 2015, Sparsifying preconditioner for the Lippmann–Schwinger equation: Multiscale Modeling & Simulation, **13**, 644–660.

Zebker, H. A., P. A. Rosen, R. M. Goldstein, A. Gabriel, and W. C. L., 1994, On the derivation of coseismic displacement fields using differential radar interferometry: the Landers earthquake: Journal of Geophysical Research, **99**, 19617–19634.

Zhang, X., M. Burger, and S. Osher, 2010, A unified primal-dual algorithm framework based on Bregman iteration: Journal of Scientific Computing, **46**, 20–46.

Zheng, Y., P. Barton, and S. Singh, 2011, Strategies for elastic full waveform inversion of timelapse ocean bottom cable (OBC) seismic data: 81st Annual International Meeting, SEG, Technical Program Expanded Abstracts, 4195–4200.

Zoback, M., 2010, Reservoir geomechanics: Cambridge University Press.

Zoback, M. D., and J. C. Zinke, 2002, Production-induced normal faulting in the Valhall and Ekofisk oil fields: Pure and Applied Geophysics, **159**, 403–420.

Zorich, V. A., 2008, Mathematical analysis II (Universitext): Springer.

Index

2.5D inversion, 71
4D analysis, 5

absorbing boundary conditions, 134
adjoint equation, 125
adjoint field, 124
adjoint-state method, 124
ADMM, 41, 149
    with hot-restarted conjugate gradients, 167
Alternating Directions Method of Multipliers, *see* ADMM, 149
Amplitude Versus Offset, *see* AVO, 7
AVO, 7

Biot's theory, 83
bound-constrained optimization, 192
boundary conditions
    absorbing, 13, 134
    free surface, 134

center of dilatation, 84
CGNE, 153
Compressive Conjugate Directions, 41, 152
    limited-memory, 166
condition
    Sommerfeld radiation condition, 13
conjugate directions, 154
Conjugate Directions for Normal Equations, *see* CGNE, 153
CSS, *see* Cyclic Steam Stimulation 95
cycle skipping, 22, 137
Cyclic Steam Stimulation, 95

data misfit, 118
diffractor, 20
dual optimization problem, 148

effective stress, 8
Enhanced Oil Recovery, 8
equality-constrained optimization, 184
equation
    Helmholtz, 13
    Lippman-Schwinger, 14
    wave, *see* wave equation, 120

Fast Iterative Shrinkage -Thresholding Algorithm, *see* FISTA, 148
finite elements, 91
first discretize then optimize, 130
FISTA, 148
fluid substitution, 6
    the effect on elastic rock properties, 6
free surface boundary conditions, 134

full derivative, 122
full-waveform inversion
    4D, *see* time-lapse full-waveform inversion, 8
    time-lapse, 8
    phase-only, 135
function
    objective, *see* objective function 22
    penalty, 22
FWI
    phase-only, 135
FWI misfit function, 30
    phase-only, 21, 30, 135
FWI objective function, 30

GMRES, 155
gradient descent method, 147
    accelerated, 148
Green's tensor
    elastostatic, 84

homotopy methods, 147
hysteresis
    of velocity/stress relationship, 7

image denoising, 168
implicit function theorem, 121
inequality-constrained optimization, 192
interior point methods, 88, 147
inverse problem
    of inverting pressure contrasts, 178
    of inverting pressure from deformation, 83, 170, 178
    of inverting sparse dilatational pseudo-sources, 170
ISTA, 147
Iterative Shrinkage -Thresholding Algorithm, *see*, ISTA, 147

Karush-Kuhn-Tucker optimality conditions, *see* KKT optimality conditions, 199
KKT optimality conditions, 163, 199
Krylov subspace, 154

Lagrange multipliers, 195
Lagrangian, 149
    augmented, 149, 194
Lagrangian multipliers, 149
least squares
    L_1-regularized, 146
    condition number of, 167
    nonlinear, *see* nonlinear least squares, 118
    regularized, 146
    TV-regularized, 146
linearized adjoint equation, 125

matrix
    banded, 132
    Jacobi, 119
    sparse, 132
    triangular, 131
    tridiagonal, 132
medium
    layered, 84
    weakly laterally heterogeneous, 84

method of multipliers, 149
model
    baseline, 21
    monitor, 21
modulus
    bulk, 7
    shear, 7

Nesterov relaxation, 148
nonlinear least squares, 118

objective function, 118
    FWI, 22
    gradient of, 118
    phase-misfit, 135
operator
    condition number, 153
    convolutional, 86
    integral, 86, 173
    projection, 194
    resolvent, 147
    shrinkage, 147
    splitting, 149
        Douglas-Rachford, 149
        three-way, 193
overburden
    dilation, 7

phase
    misfit, 135
    unwrapping, 135
phase misfit residual, 137
poroelastostatic system, 85
    coupled, 85, 91
    semicoupled, 83
preconditioner
    sparsifying, 15
primal-dual optimization method, 148

quadratic penalty function, 167
quasilinear equations, 134
quasistatic poroelastic deformation, 83

R-factor, 64
regularization, 22, 31
    L_1, 31, 146
    L_2, 31
    model, 31
    model difference, 31
    Tikhonov, 25, 88, 198
    total-variation, *see* TV regularization, 88
    TV, 31, 88, 146
reservoir
    caprock, 7
    compaction, 7
    low-permeability barrier, 109
    permeability, 109
    trap, 7
reservoir simulation
    single-phase, 109
ROF model, 192
Rudin-Osher-Fatemi model, *see* ROF model, 192

saddle-point of Lagrangian, 149

scattering
    inverse Rytov, 21
scattering series
    Born, 15
    Rytov, 17
self-adjoint operator, 128
slowness
    acoustic, 13
    perturbation, 13
source, 13
    dilatational, 173
    sparse pseudo-source, 173
Steered Conjugate Directions, 154
    for constrained optimization, 184
strain, 64
    time, 64
subgradient, 147

time marching, 131
time-lapse inverse theory, 5
tomography
    diffraction, 17
    travel-time, 20
total derivative, *see* full derivative, 122
Total-variation minimization
    with bound constraints, 192
TV seminorm, 31
    anisotropic, 169
    isotropic, 31

Uzawa method, 149

variational equations, 122

velocity
    acoustic, 7
    pressure, *see* acoustic velocity, 7
    relationship to confining stress, 7
    relationship to pore pressure, 7
    relationship to strain, 7
    shear, 7

wave equation
    in frequency domain, 120
    in time domain, 130