

RICE UNIVERSITY

Born Waveform Inversion in Shot Coordinate Domain

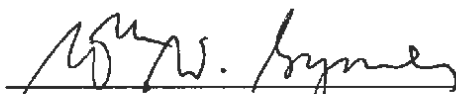
by

Yin Huang

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:



William W. Symes, *Chair*

Noah G. Harding Professor of Computational and Applied Mathematics



Béatrice M. Rivière

Noah G. Harding Professor of Computational and Applied Mathematics



Colin A. Zelt

Professor of Department of Earth Science

HOUSTON, TEXAS
FEBRUARY 2016

ABSTRACT

Born Waveform Inversion in Shot Coordinate Domain

by

Yin Huang

The goal of this thesis is to integrate Born waveform inversion, variable projection algorithm and model extension concept to get a method that can improve the long scale background model updates reliably and efficiently from seismic data.

Born waveform inversion is a partially linearized version of full waveform inversion based on Born (linearized) modeling, in which the earth model is separated into a smooth long scale background model and an oscillatory short scale reflectivity and both are updated to fit observed trace data. Because kinematic variables (background model) are updated, Born waveform inversion has the same feature as full waveform inversion: very sensitive to initial model when solved by gradient based optimization method (almost the only possible method because of the problem scale). Extended Born waveform inversion allows reflectivity to depend on additional parameters, potentially enlarging the convexity domain by enlarging the searching model space and permitting data fit throughout the inversion process and in turn reducing the sensitivity to initial model.

Extended or not, the Born waveform inversion objective function is quadratic in the reflectivity, so that a nested optimization approach is available: minimize over reflectivity in an inner stage, then minimize the

background-dependent result in a second, outer stage, which results in a reduced objective function of the background model only (VPE method). This thesis integrates the nested optimization approach into an inversion scheme and analyzes that the accuracy of the solution to the inner optimization is crucial for a robust outer optimization and both model extension and the nested optimization are necessary for a successful Born waveform inversion. And then we propose a flexibly preconditioned least squares migration scheme (FPCG) that significantly improves the convergence of iterative least squares migration and produces high resolution images with balanced amplitude. The proposed scheme also improves the efficiency of the solution to the inner stage of the nested optimization scheme and the accuracy of the gradient, and thus potentially improves the efficiency of the VPE method. However, a theoretical error estimate in the gradient computation of the VPE method is still hard to obtain, and we explain the reason and illustrate with numerical examples.

ACKNOWLEDGEMENTS

This thesis would not have been possible without support, encouragement and guidance of many people.

First, I would like to express my sincere gratitude to my advisor Dr. William Symes for his consistent support, inspiration and guidance throughout my research work. In particular, his deep insights on connections of Mathematics and geophysical research, his broad knowledge on both subjects shaped and enlightened my understanding of the subjects. He was also very supportive for me to do internships, through which I not only gained a perspective on the general area of geophysical research, but also learned necessary methods and skills that are useful in my thesis work.

Besides my advisor, I am also very grateful to my thesis committee: Dr. Béatrice Rivière and Dr. Colin Zelt, for their insightful comments and encouragement and for questions that make me think deeply.

My sincere thanks also goes to Dr. Anatoly Baumstein and Dr. Yaxun Tang, who provided me an opportunity to join their team as intern and allowed me to test some of the ideas in chapter 2 of this thesis in their team. Their comments and suggestions are very helpful to me.

I would also like to thank Dr. Rami Nammour for his mentorship during my internship where I developed not only valuable experience, but also an algorithm in chapter 3.

I thank sponsors of The Rice Inversion Project (TRIP) for their continued support and thank group members of TRIP and my friends at CAAM department for the stimulating discussions and all the fun we have had in the last few years.

Last but not the least, I would like to thank my family: my husband and my parents for supporting me spiritually throughout writing this thesis and my life in general.

Contents

Abstract	ii
Acknowledgements	iv
1 Introduction	1
2 Born Waveform Inversion via Variable Projection and Shot Coordinate Domain	7
2.1 Summary	7
2.2 Introduction	8
2.3 Theory	13
2.4 Numerical Examples	19
2.5 Discussion	31
3 Flexibly Preconditioned Extended Least Squares Migration in Shot Record Domain (submitted to <i>Geophysics</i>)	33
3.1 Summary	33
3.2 Introduction	34
3.3 Theory	39
3.4 Numerical examples	50
3.5 Discussion	67
4 Variable Projection Extended Born Waveform Inversion Enhanced by Preconditioned Least Squares Migration	69
4.1 Summary	69
4.2 Introduction	70
4.3 Theory	72
4.4 Numerical Examples	79

4.5 Discussion	100
5 Discussion	101
A Power of Laplace operator Λ and implementation	104
B Computation of the gradient of J_{VPE}	106
References	108

List of Figures

1.1	FWI objective function shape demonstration (taken from [Warner and Guasch, 2014a]). 0% is the target model.	3
2.1	Marmousi example: Soure Ricker wavelet with 6 Hz peak frequency.	20
2.2	Marmousi example: Target background model (a) target reflectivity model (b)	20
2.3	Marmousi example: Born shot record with index 41.	21
2.4	Marmousi example: value of VPE objective along line segment $m =$ $(1 + \sigma)m_{\text{target}} - \sigma m_0$ with m_{target} shown in Figure 2.2a and $m_0 =$ $(1.5\text{km/s})^2$ with $\sigma = 0$ the position of target background model.	22
2.5	Marmousi example: Relative error between $\langle \nabla J_{\text{VPE}}[m], dm \rangle_s$ and $\frac{J_{\text{VPE}}[m + h * dm] - J_{\text{VPE}}[m]}{2h}$ with the first one as the value of reference. $m = 0.7m_{\text{target}} + 0.3m_0$ and $dm = 0.1 (\text{km/s})^2$ with m_{target} shown in Figure 2.2a and $m_0 = (1.5$ $\text{km/s})^2$	23
2.6	Truncated Marmousi example: reflectivity model (a) and Born shot record, shot index 41 (b).	25
2.7	Truncated Marmousi example: true velocity-squared background model (a); initial background model (b); inverted background model from 7 steps of VPE method (c) and 350 steps of EBWI method (d)	26
2.8	Truncated Marmousi example: Inverted reflectivity model at true back- ground model (a); initial background model (b); background model with 7 steps of VPE method by solving equation(2.9) (c) and 350 steps of EBWI method (d)	27
2.9	Truncated Marmousi example: Common image gathers at true back- ground model (a); initial background model (b); b ackground model with 7 steps of VPE method by solving equation 2.9 (c) and 350 steps of EBWI method (d).	28

2.10	Marmousi example: Inverted background model from 18 steps of VPE method (a) and 18 steps of VP method (b).	29
2.11	Marmousi example: Inverted reflectivity model at true background model (a); initial background model (b); background model with 18 steps of VPE method (c) and 18 steps of VP method (d) by solving equation 2.9.	29
2.12	Marmousi example: Common image gathers at true background model (a); initial background model (b); background model with 18 steps of VPE method (c) by solving equation 2.9.	30
2.13	Marmousi example: Trace comparison of real data (red), predicted data by VPE (green) and VP (blue) methods for far (top) and near (bottom) offsets.	31
3.1	Slow Gaussian anomaly model: (a) background m , (b) perturbation δm	41
3.2	Convergence curve of steepest descent method for the slow Gaussian anomaly problem: (a) normalized data misfit (b) normalized gradient length.	42
3.3	(a) source wavelet with corner frequency 1 Hz, 7 Hz, 28 Hz, 35 Hz, used in Gaussian anomaly examples (b) 3 of 81 Born shot records for slow Gaussian anomaly example (Figures 3.1a and 3.1b.)	51
3.4	Slow Gaussian anomaly example: convergence curves of numerical methods, (a) normalized data misfit and (b) normalized gradient length.	52
3.5	Slow Gaussian anomaly example: inverted model perturbation cube after 20 Hessian applications using FPCG (a) and using CG with windowing (b).	53
3.6	Slow Gaussian anomaly example: Stacked image after 20 Hessian applications using FPCG (a) and using CG with windowing (b).	53
3.7	Slow Gaussian anomaly example: data residual, same 3 shots as in Figure 3.3b after 20 Hessian applications using FPCG (a), and using CG with windowing (b).	54
3.8	Fast Gaussian anomaly example: (a) background model, and (b) 3 of 81 Born shot records.	55
3.9	Fast Gaussian anomaly example: convergence curves of numerical methods for (a) normalized data misfit and (b) normalized gradient length.	55

3.10	Fast Gaussian anomaly example: inverted model perturbation cube after 20 Hessian applications using FPCG (a) and using CG with windowing (b).	55
3.11	Fast Gaussian anomaly example: Stacked image after 20 Hessian applications using FPCG (a) and using CG with windowing (b).	56
3.12	Fast Gaussian anomaly example: Data residual after 20 Hessian applications using FPCG (a) and using CG with windowing (b).	56
3.13	Marmousi example: (a) background model, (b) model perturbation.	57
3.14	Marmousi example: (a) source wavelet with corner frequencies = 1 Hz, 6 Hz, 25 Hz, and 30 Hz. (b) Born shot record, shot index 65.	58
3.15	Marmousi example: convergence curves of numerical methods, (a) normalized data misfit and (b) normalized gradient length.	58
3.16	Marmousi example: inverted model perturbation cube after 20 Hessian applications using FPCG (a) and using CG with windowing (b).	59
3.17	Marmousi example: Stacked image after 20 Hessian applications using FPCG (a) and using CG with windowing (b).	60
3.18	Marmousi example: Vertical trace comparison at $x = 4000$ m of target model perturbation with inverted model perturbation cube from the 44-th shot (a) and with stacked inverted image (b) by FPCG and CG with windowing methods after 20 Hessian applications.	60
3.19	Marmousi example: Spectrum comparison of the vertical trace shown in Figure 3.18 of target model perturbation with inverted model perturbation cube from the 44-th shot (a) and with stacked inverted image (b) by FPCG and CG with windowing methods after 20 Hessian applications.	61
3.20	Marmousi example: data residual, same shot record as in Figure 3.14b, after 20 Hessian applications using FPCG (a), and using CG with windowing (b).	62
3.21	Marmousi example: data residual after dip filtering with 20 Hessian applications using FPCG (a), and using CG with windowing (b).	63
3.22	Marmousi example: Relative power spectra of data residual (a) and data residual after dip filtering (b) with respect to observed data.	63
3.23	Marmousi example, incorrect background model.	64

3.24	Marmousi example, incorrect background model: convergence curves of numerical methods, (a) normalized data misfit and (b) normalized gradient length.	65
3.25	Marmousi example, incorrect background model: inverted model perturbation cube after 20 Hessian applications using FPCG (a) and using CG with windowing (b).	65
3.26	Marmousi example, incorrect background model: Stacked image after 20 Hessian applications using FPCG (a) and using CG with windowing (b).	66
3.27	Marmousi example, incorrect background model: data residual, same shot record as in Figure 3.14b, after 20 Hessian applications using FPCG (a), and using CG with windowing (b).	67
4.1	Target reflectivity δm (a) and source wavelet with corner frequency 1 Hz, 7 Hz, 28 Hz, 35 Hz (b).	79
4.2	Windowing operator W	80
4.3	Model $\delta m + r_k$ (a) and spectrum of model r_k (b) with $k = 30$	80
4.4	With bandpass filtered source wavelet: (a) $F[m](\delta m + r_k)$ and (b) $N[m](\delta m + r_k)$ with $\delta m + r_k, k = 30$ shown in Figure 4.3a.	81
4.5	Model $\delta m + r_k$ (a) and spectrum of model r_k (b) with $k = 70$	81
4.6	With bandpass filtered source wavelet: (a) $F[m](\delta m + r_k)$ and (b) $N[m](\delta m + r_k)$ with $k = 70$	82
4.7	With bandpass filtered source wavelet: Green curve in both subfigures is the relative error $\frac{\ r_k\ }{\ \delta m\ }$. (a) the relative error of Born modeling operator $\frac{\ F[m]r_k\ }{\ F[m]\delta m\ }$ and (b) the relative error of normal operator $\frac{\ N[m]r_k\ }{\ N[m]\delta m\ }$ with $r_k = W \sin(kz)$, $m = (2\text{km/s})^2$	82
4.8	Source wavelet after twice integral over time of wavelet in Figure 4.1b.	83
4.9	With twice integrated source wavelet: Green curve in both subfigures is the relative error $\frac{\ r_k\ }{\ \delta m\ }$. (a) the relative error of Born modeling operator $\frac{\ F[m]r_k\ }{\ F[m]\delta m\ }$ and (b) the relative error of normal operator $\frac{\ N[m]r_k\ }{\ N[m]\delta m\ }$ with $r_k = W \sin(kz)$, $m = (2\text{km/s})^2$	84
4.10	(a) target background model and (b) target reflectivity.	85

4.11 (a) source wavelet with corner frequency 1 Hz, 7 Hz, 28 Hz, 35 Hz and (b) 3 of 81 Born shot records.	85
4.12 Data integrated over t twice.	86
4.13 At true background model: convergence curves of numerical methods, (a) normalized residual and (b) normalized gradient length.	87
4.14 At true background model: Stacked image after 20 Hessian applications using FPCG (a) and using CG with windowing (b).	87
4.15 At true background model: Common image gathers after 20 Hessian applications using FPCG (a) and using CG with windowing (b).	88
4.16 At true background model: data residual after 20 Hessian applications using FPCG (a) and using CG with windowing (b).	88
4.17 At constant background model $((2 \text{ km/s})^2)$: convergence curves of nu- merical methods, (a) normalized residual and (b) normalized gradient length.	89
4.18 At constant background model $((2 \text{ km/s})^2)$: Stacked image after 20 Hessian applications using FPCG (a) and using CG with windowing (b).	90
4.19 At constant background model $((2 \text{ km/s})^2)$: Common image gathers after 20 Hessian applications using FPCG (a) and using CG with win- dowing (b).	91
4.20 At constant background model $((2 \text{ km/s})^2)$: data residual after 20 Hessian applications using FPCG (a) and using CG with windowing (b).	91
4.21 Value of VPE objective by different numerical methods and different number of inner iterations along line segment $m = \sigma m_{\text{target}} + (1 - \sigma)m_0$ with m_{target} shown in Figure 4.10a and $m_0 = (2\text{km/s})^2$ with $\sigma = 0$ the position of target background model.	92
4.22 Value of VPE objective by FPCG method with different penalty weight α and different number of iterations along line segment $m = \sigma m_{\text{target}} +$ $(1 - \sigma)m_0$ with m_{target} shown in Figure 4.10a and $m_0 = (2\text{km/s})^2$ with $\sigma = 0$ the position of target background model.	93
4.23 At constant background model $(m = (2 \text{ km/s})^2)$: output of $D\bar{F}[m]^T[\delta\bar{m}, \bar{F}[m]\delta\bar{m} -$ $d]$ with $\delta\bar{m}$ the solution of problem 4.6 after 20 Hessian applications using FPCG (a) and using CG with windowing (b).	94
4.24 At constant background model $(m = (2 \text{ km/s})^2)$: gradient computed by equation 4.9 with $\delta\bar{m}$ the solution of problem 4.6 after 20 Hessian applications using FPCG (a) and using CG with windowing (b).	95

4.25	97
4.26	Fast Gaussian anomaly dm	97
4.27	Output of $DF[m][dm, \delta m + r_k]$ with $k = 30$, $m = (2\text{km/s})^2$ and δm shown in Figure 4.10b with bandpass filtered source wavelet (a) and source wavelet integrated over time twice (b).	97
4.28	Output of $DF[m][dm, \delta m + r_k]$ with $k = 70$, $m = (2\text{km/s})^2$ and δm shown in Figure 4.10b with bandpass filtered source wavelet (a) and source wavelet integrated over time twice (b).	98
4.29	Green curve in both subfigures is the relative error $\frac{\ r_k\ }{\ \delta m\ }$. the relative error of $DF[m]$ operator $\frac{\ DF[m][dm, r_k]\ }{\ DF[m][dm, \delta m]\ }$ with $r_k = W \sin(kz)$, $m = (2\text{km/s})^2$ and dm the fast Gaussian anomaly shown in Figure 4.26 when using (a) bandpass filtered source wavelet (Figure 4.1b) and (b) twice integral of bandpass filtered source wavelet (Figure 4.8).	99
4.30	Quotient of L_2 norms of $DF[m][dm, r_k]$ and $F[m]r_k$, when using (a) bandpass filtered source wavelet (Figure 4.1b) and (b) twice integral of bandpass filtered source wavelet (Figure 4.8).	99

Introduction

Seismic inversion in exploration geophysics is a process to determine subsurface structure of the Earth from seismic data, usually collected at the surface and sometimes in wells. Full waveform inversion (FWI) [Tarantola, 1984a, Virieux and Operto, 2009] is defined as a least squares data-fitting problem, with the prediction operator related to the solution operator of the initial and boundary values problem of a wave equation with varying coefficients. The structure of the earth is described by these variable coefficients or *models*, which will be determined by FWI. Models could be density, wave speed, bulk modulus, etc, according to the wave equation we choose to simulate the wave propagation.

This thesis will discuss methods to solve partially linearized version of FWI, namely Born waveform inversion that aims at separately finding both the low frequency components (*background model*) of the model, which is hard to achieve for classic FWI method, and high frequency components (*reflectivity*) of the model, which FWI is capable of once the low frequency components is close to correct. We will combine (a) *model extension* concept: introduce additional degrees of freedom to the reflectivity and (b) *variable projection* method: solve for reflectivity in an inner optimization and then update the background model in an outer optimization, to obtain an effective and robust algorithm to solve Born waveform inversion. We use optimal

pseudodifferential scaling method to construct a flexibly preconditioned conjugate gradient scheme that produces high quality reflectivity efficiently, that is improve the convergence rate of the inner optimization problem and thus speed up the updates of the background model as well. These terminologies will be explained in detail later.

Seismic data usually is collected from many seismic experiments that covers the same domain of interest. Each seismic experiment uses a localized energy source (shot) to initiate the motion of the ground (generate waves to propagate through the earth). These waves travel through the ground and/or water and are refracted and/or reflected by discontinuities and are recorded at the surface by geophones (land) or hydrophones (marine). The recorded data is usually the displacement of particles or pressure field. Data is usually recoded separately for each shot.

Different frequency components in the model are responsible for different aspects in the data: high frequency components of the model (reflectivity) produce reflections in the data and low frequency components of the model (background model) control when these reflections appear in the data. FWI aims to find all these components as a whole model from observed seismic data. However, without good enough initial model (close enough to the true model), FWI usually fails to recover the low frequency components [Plessix, 2009]. Without correct low frequency components, reflectors will all mispositioned. Because of the problem scale, FWI is usually solved using gradient based optimization methods, with the gradient formulated as the application of the adjoint of the Born approximation (Frechet derivative of the forward modeling operator) to the data residual. It is widely believed (without direct validation) that FWI has a lot of local minima that are far from its global minimizer, which is what people suspect when the optimization algorithm stopped. Figure 1.1 demonstrates the shape of a classic FWI objective function.

The background model is usually hard to determine and methods intending to

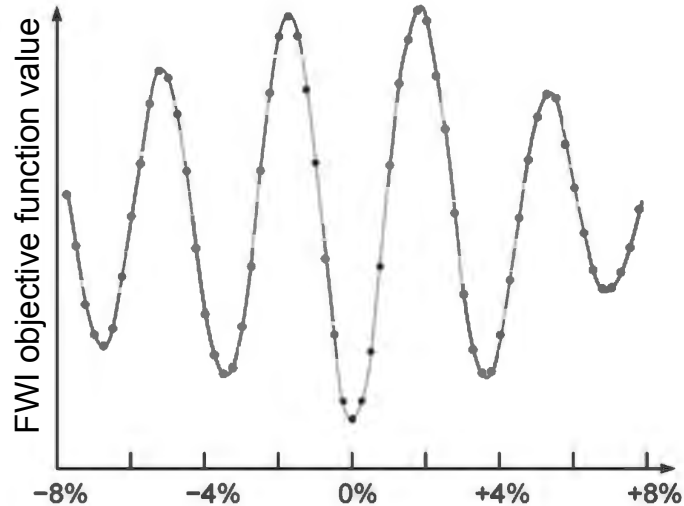


Figure 1.1: FWI objective function shape demonstration (taken from [Warner and Guasch, 2014a]). 0% is the target model.

identify the background model are called Velocity Analysis, with Velocity referring to the low frequency (long scale) background model. Velocity Analysis is a name for a very broad range of methods. The one that is relevant to this thesis is Wave Equation based Migration Velocity Analysis (WEMVA) [Shen, 2012, Shen, 2004], which is also formulated as an optimization problem, with the objective function quantifies the image. An image is a rough approximation to the high frequency reflectivity model at a given background model. For a given background model, an image is produced by applying the adjoint of Born modeling operator to the data. If the data is divided into subsets (for example, each subset is corresponding to a shot) and we produce one different image for each subset, the discrepancy between these images can indicate the next update to the background model. This discrepancy can be measured by an objective function, and minimized to update the background model. Objective functions using differential operators to measure image discrepancy have been called Differential Semblance (DSO) [Symes, 1986, Symes, 1990, Symes, 1993, Jervis et al., 1996, Mulder and ten Kroode, 2002, Chauris and Noble, 2001]. For example, if the data is

grouped by shot coordinates, the images corresponding to different shots should be the same at the same physical location and in this case DSO measures the difference of the images of adjacent shots. DSO objective functions are different for different ways of dividing the data. The DSO term is also part of the objective function we will investigate in this thesis.

The model extension concept comes exactly from this idea: find different model to fit different subsets of data, so that the model depends on an additional axis. Instead of introducing model extension to classic FWI [Sun and Symes, 2012], we introduce model extension to the linearized version of FWI as many authors have done [Symes and Carazzone, 1991, Kern and Symes, 1994, Biondi and Almomin, 2014]: (a) replace forward modeling operator with Born modeling operator (first order derivative of the forward modeling operator) by viewing the background model as a reference model and the reflectivity as a perturbation; (b) allow only the reflectivity to be extended (c) add a DSO term to penalize the degree of extension. We call the resulting objective function the extended Born waveform inversion.

Both the extended reflectivity and background model will be updated in the linearized version of FWI with model extension. In chapter 2, I will discuss the variable projection method to optimize this objective function. This is a nested optimization method that updates the reflectivity in an inner stage, and then updates the background model in an outer stage. The variable projection extended Born waveform inversion tends to be more effective for the background model updates than similar inversion methods either without variable projection or without model extension, according to our numerical examples. However, in order for the method to converge to the target model, a large number of iterations is necessary for the inner problem. Computation of an accurate gradient is essential for convergence, and also depends on the accurate enough solution of the inner problem.

The inner problem (for a fixed background model, finding an extended reflectivity model to fit the data) is a version of least squares migration problem [Bourgeois et al., 1989, Chavent and Plessix, 1999, Nemeth et al., 1999, Tang, 2009, Aoki and Schuster, 2009, Dutta et al., 2014, Luo and Hale, 2014, Valenciano et al., 2015]. In chapter 3, we propose a flexibly preconditioned least squares migration algorithm using flexibly preconditioned conjugate gradient method [Notay, 2000, Knyazev and Lashuk, 2006] and optimal pseudodifferential scaling [Nammour and Symes, 2009, Nammour, 2009] as a preconditioner to solve the inner problem (without DSO penalty term) in chapter 2. The proposed method could improve the efficiency of the short scale model recovery and the quality of the resulting reflectivity or image when the background model could correctly predict the travel time. When the background model is wrong, the algorithm could converge fast as well and generate image gathers that more clearly indicate the correctness of the background model and could be used to improve the performance of the image based velocity analysis, and more importantly the variable projection EBWI discussed in chapter 2.

In chapter 4, the flexibly preconditioned least squares migration in shot record domain proposed in chapter 3 is integrated into the nested optimization scheme to speed up the inner optimization problem over the reflectivity model. With preconditioning, the solution of the inner problem could achieve the same accuracy with less computational cost than the non-preconditioned method used in chapter 2. Preconditioning also produces a better shaped objective function than without preconditioning: the global minimum of the reduced problem (outer problem over background model) is obtained at the target background model, while there are still models that have smaller objective function value than the target model if the inner problem is solved by a non-preconditioned method even with much more computational cost. Numerical test shows that a more accurate gradient is obtained with preconditioning than

without preconditioning at least on the direction we tested. However, a small error in the inner problem solution does not necessary imply a small error in the gradient computation, at least in its current setting. We explain the reason and support with a numerical example.

At last, in chapter 5 I will summarize methods have been talked and examined in the thesis and discuss some potential future improvements and possible applications.

A version of chapter 2 has been accepted and presented at the SEG International Exposition and 85th Annual Meeting, 2015 as an expanded abstract. Chapter 3 as a paper (has coauthor: Rami Nammour) has been submitted to *Geophysics*. A version of Chapter 4 will be submitted to the SEG 86th Annual Meeting, 2016 soon. Each chapter of 2, 3 and 4 is self-contained, with literature review, theoretical development, numerical examples, and discussion, and could be read independently of the other Chapters.

Born Waveform Inversion via Variable Projection and Shot Coordinate Domain

A version of this chapter has been presented at SEG International Exposition and 85th Annual Meeting, 2015.

2.1 Summary

Born waveform inversion is a partially linearized version of full waveform inversion based on Born (linearized) modeling, in which the earth model is separated into a smooth background model and a short scale reflectivity, and both are updated to fit observed trace data. Because kinematic variables (velocity) are updated, the region of convexity in model space is small (on the order of a wavelength), just as it is for Full Waveform Inversion. Extended Born waveform inversion allows reflectivity to depend on additional parameters, potentially enlarging the convexity domain to be wavelength-independent and enhancing the likelihood of convergence to a kinematically accurate model, by permitting data fit throughout the inversion process. Extended or not, the Born waveform inversion objective function is quadratic in the

reflectivity, so that a nested optimization approach is available: minimize over reflectivity in an inner stage, then minimize the background-dependent result in a second, outer stage. This paper uses a 2D acoustic modeling, reflectivity permitted to depend on shot coordinates (shot record extension), a differential semblance penalty to control this dependence, and the variable projection variant of nested optimization. Our examples suggest that neither extended modeling nor variable projection alone are sufficient to enable convergence to a global best-fitting model, but the two together are quite effective.

2.2 Introduction

Seismic full waveform inversion (FWI) is used to infer the interior structure of the earth from observed seismic waves by posing model-based data fitting as a nonlinear least squares problem. Studied in the 1980's by Tarantola and others [Tarantola, 1984a], it has recently become a viable model building strategy [Virieux and Operto, 2009] as a result of the fast development in high performance computing and advances in seismic acquisition technology. Because of the band-limited feature of seismic data, the FWI objective function has a rather small domain of convexity, of diameter proportional to a dominant wavelength in the data: unless the initial model predicts arrival time of major events to within roughly a half-wavelength, convergence of local iterative methods to a global minimizer is doubtful, and model estimates obtained with reasonable computation effort may share few features with global minimizers [Gauthier et al., 1986].

Many methods have been introduced and investigated to enhance the convergence of iterative methods FWI. The availability of long-offset and low frequency data reduces the sensitivity of FWI to initial model [Mora, 1987, Vigh et al., 2013], and in fact this is the approach taken in contemporary production FWI.

This paper will discuss fundamental modifications to FWI that enhance the potential for global convergence. Our work takes place in the context of a partly linearized version of the FWI problem, in which the model is separated into a smooth (or long scale) background model and an oscillatory (or short scale) reflectivity [Symes, 2009], and replace the forward modeling operator of FWI with its linearization (Born approximation), treating the reflectivity as a perturbation. The Born modeling operator is linear in the reflectivity, but nonlinear in the background model, so the FWI problem has been partly linearized by this transformation. We call the least-squares data-fitting problem, based on the Born modeling operator, the Born waveform inversion (BWI) problem. It is essentially FWI with single scattering.

We study two additional modifications to the BWI problem. The first is model extension, or addition of non-physical axes to the model parameters. Addition of suitable axes to the reflectivity enables data fit throughout the inversion, with a differential semblance penalty [Symes, 2008b] to penalize the extension and move the extended model towards physicality. Second, we use a nested optimization approach: eliminate the reflectivity via solution of an inner (quadratic) optimization with the background model fixed, then update the background model in an outer loop to minimize the differential semblance penalty. We formulate the nested optimization by *variable projection* [Golub and Pereyra, 2003], as it generates more convenient computations than alternatives.

Our numerical results indicate that both variable projection and model extension appear to be necessary for reasonably fast global convergence of Born waveform inversion: neither technique is sufficient on its own.

We call BWI with model extension extended Born waveform inversion (EBWI), to distinguish it from (a) least squares migration, with or without reflectivity extension, in which only reflectivity is inverted, but the background model is held fixed, and

(b) BWI without extension (essentially the same as the algorithm called Reflection Waveform Inversion (RWI) by [Xu et al., 2012]), in which physical (non-extended) reflectivity is inverted together with the background model. The salient feature of extended linearized modeling, namely that the data can be fit relatively well with any background model, implies that some sort of penalty must be added to EBWI, to measure how far the extended reflectivity is from a physical model. The background model is then adjusted to minimize this penalty. Several reflectivity extensions have been investigated in the context of EBWI, each with its appropriate penalty term. In the work reported here, we use the shot record model extension, in which each shot is modeled with an independent reflectivity (but all with the same background model). Several methods for measuring the degree of model extension have been suggested. We could minimize the difference of images from neighboring shots, which is the very first version of differential semblance optimization (DSO) [Symes, 1986, Symes, 1990, Symes and Carazzone, 1991, Symes, 1993, Kern and Symes, 1994, Jervis et al., 1996, Mulder and ten Kroode, 2002, Chauris and Noble, 2001]. This penalty is essentially the only choice for the shot-record extension leading to a smooth objective function, amenable to gradient-based optimization [Stolk and Symes, 2003]. We could maximize the similarity of images along shot record direction, by evaluating the stack power [Ronen and Claerbout, 1985, Chavent and Jacewitz, 1995]. [Chauris et al., 1998] compared the behavior of DSO and stack power methods: their conclusion, consistent with [Stolk and Symes, 2003], was that stack power tends to exhibit small-scale oscillations and have the same small domain of convexity as FWI, whereas DSO appears to be smooth convex over a much larger set of models. Recently [Chauris and Plessix, 2013] proposed a differential waveform inversion scheme, where a inverted reflectivity from one shot is used to compute the predicted data at the next shot, and that data residual will be used to update the background model. With Born

modeling, this approach is essentially the same as DSO. [Chauris and Plessix, 2013] conjecture that with full waveform (rather than Born) modeling, this method may converge to a correct velocity model even with strong surface related multiples in the data.

A different approach to reflectivity extension uses the subsurface offset and/or scattering angle as the additional axis. See [Shen and Symes, 2008, Sun and Symes, 2012, Biondi and Almomin, 2012, Weibull and Arntsen, 2014, Lameloise et al., 2014, Biondi and Almomin, 2014] for recent examples of EBWI (and its nonlinear analogue using full waveform modeling) based on these extensions, and [Symes, 2008b] for an overview of earlier work and the comparison of surface vs. subsurface approaches to reflectivity extension.

Reflectivity extension is not the only possible approach to enabling data fit throughout an FWI process. Several authors have investigated the use of additional source parameters for this purpose [Plessix et al., 2000, Plessix, 2000, Luo and Sava, 2011, van Leeuwen and Herrmann, 2013, Warner and Guasch, 2014b]. Since the source acts linearly within the wave modeling system, the form of these algorithms has much in common with EBWI as presented here, although the modeling details and extension penalty are quite different.

For any optimization problem, such as the DSO approach to EBWI, in which some of the variables appear in a simpler way (for example, quadratically) than others, a nested optimization approach is natural: eliminate the “simple” variables first via an inner optimization, then optimize a function function of the “complicated” variables in an outer optimization loop. In EBWI, we can optimize first over (extended) reflectivity, to produce a background model dependent reflectivity, then update the background model by minimizing the “reduced” extension penalty (differential semblance, for example). This nested approach has also be applied to BWI [Clément and Chavent, 1993, Plessix et al., 1999, van Leeuwen and Mulder, 2009, Xu et al., 2012].

The Variable Projection Method of [Golub and Pereyra, 2003] is an attractive approach to nested optimization of such “separable” least squares problem, in that the gradient computation of the reduced objective is relatively simple. Variable projection differs from other approaches to nested optimization by using the same objective, a linear combination of data misfit and extension penalty, for both the inner and outer optimizations. [van Leeuwen and Mulder, 2009] and [Xu et al., 2012] use variable projection for BWI. [Kern and Symes, 1994] showed how to combine variable projection and extended modeling, and described an adjoint state method for background model gradient computation. We integrate these components into an inversion algorithm. We quantify the suggestion of [Kern and Symes, 1994], that the accuracy of the solution of the inner problem, i.e. the minimization of the variable projection cost function over reflectivity for fixed background model, tends to be crucial for the reduced objective function to have global minimum at the target background model, and for accuracy in the gradient computation.

We illustrate these conclusions with numerical examples based on the Marmousi velocity model [Bourgeois et al., 1991]. The target background model is created by smoothing the Marmousi velocity model. The target reflectivity model is the difference of the original model and a less stringent smoothing. We truncate and scale the model and data in two ways, creating two inversion examples. Based on these examples, we illustrate that the inner optimization needs to be solved accurately enough by plotting the objective function values along a line segment, on which the target background model lies. The first inversion example provides a relatively good starting model; nonetheless, its results suggest that without variable projection, neither Born waveform inversion nor its extended variant are likely to be successful, whereas variable projection added to either approach yields accurate inversion. The second inversion example poses a more difficult problem, that of convergence from a con-

stant background velocity. Only the variable projection extended approach appears to produce constructive velocity and reflectivity updates.

The rest of the paper is organized as follows: In the theory section, we start from the classic full waveform inversion, and introduce the Born waveform inversion and model extension: specifically in shot record domain, discuss variable projection optimization method and gradient computation. In the numerical results section, we first illustrate that the inner optimization needs to be solved accurately enough by plotting the reduced objective function values along a line segment, on which the target background model lies, then apply this optimization method to two examples created from the Marmousi velocity model [Bourgeois et al., 1991] and compare its performance with several relevant methods. We end the paper by discussing the issue of this method and possible future improvements.

2.3 Theory

Classic full waveform inversion problem aims to find a model m that predicts the observed seismic data d by minimizing the following objective function:

$$J_{\text{FWI}}[m] = \frac{1}{2} \|\mathcal{F}[m] - d\|^2. \quad (2.1)$$

Denote by \mathbf{x} the physical position in 2D or 3D. Model m is a function of \mathbf{x} . We use $M = \{m(\mathbf{x})\}$ to denote the physical model space, which is an admissible subset of bounded function space. For example, if m is velocity, $m(x)$ is positive and has an upper bound. Denote by $D = \{d(\mathbf{x}_s, \mathbf{x}_r, t)\}$ the data space, with \mathbf{x}_s , \mathbf{x}_r positions of sources and receivers and t the time. $\|\cdot\|$ is the L^2 norm in data space. $\mathcal{F} : M \rightarrow D$ is a forward modeling or wave propagation operator.

We will use acoustic constant density wave equation to model the wave propaga-

tions in this paper:

$$\left(\frac{1}{c(\mathbf{x})^2} \frac{\partial^2}{\partial t^2} - \Delta \right) u(\mathbf{x}, t; \mathbf{x}_s) = \omega(t) \delta(\mathbf{x} - \mathbf{x}_s). \quad (2.2)$$

Let $m = c^2$. Then $\mathcal{F}[m] = u(\mathbf{x}_r, t; \mathbf{x}_s)$, i.e. the solution pressure field of wave equation 2.2 sampled at receiver positions \mathbf{x}_r .

It is well known that FWI objective function is highly nonlinear and requires a very good initial model to converge to a global best fitting model. The goal of this paper is to discuss several methods to modify FWI to achieve global convergence.

2.3.1 Born Waveform Inversion

Instead of using the forward modeling operator \mathcal{F} in FWI objective function 2.1 to predict the observed data, a Born modeling operator, which only models the primary reflections, is used to predict the data in Born waveform inversion.

Born modeling separates the model into a background model, still denote by m and a perturbation δm . For wave propagation, linearized modeling is generally most accurate when the background contains the long scales in the model and the perturbation the short (wavelength) scales [Symes, 2009]. We shall call the perturbation the *reflectivity*, as the background is transparent hence δm is responsible for reflections. Denote the linearized modeling operator by $F[m] = D\mathcal{F}[m]$, which is the Frechet derivative of the forward modeling operator. The least squares objective of Born waveform inversion (BWI) is:

$$J_{\text{BWI}}[m, \delta m] = \frac{1}{2} \|F[m]\delta m - d\|^2 \quad (2.3)$$

For each m , $F[m]$ is an operator on perturbational models $\delta m(\mathbf{x}) \in M$. The value of $F[m]\delta m$ is the solution of the following perturbed wave equation for a given

perturbation $\delta m = \delta c^2$ sampled at receiver positions \mathbf{x}_r .

$$\left(\frac{1}{c(\mathbf{x})^2} \frac{\partial^2}{\partial t^2} - \Delta \right) \delta u(\mathbf{x}, t; \mathbf{x}_s) = \frac{\delta c^2(\mathbf{x})}{c^2(\mathbf{x})} \Delta u(\mathbf{x}, t; \mathbf{x}_s), \quad (2.4)$$

with $u(\mathbf{x}, t; \mathbf{x}_s)$ the solution of equation 2.2.

Different from FWI, which attempts to fit all components: direct arrivals, primary reflection, diving waves and multiple reflections, in the data, BWI only aims at fitting the primary reflection data. All other components, such as direct arrivals and multiple reflections are considered as coherent noise. Preprocessing techniques, such as demultiple, may be needed to apply to the data (real data and full waveform synthetic data) before starting Born modeling based inversion.

2.3.2 Model Extension

We refer to the model space M as the the physical model space, to distinguish it from the extended model space $\bar{M} = \{\bar{m}(\mathbf{x}, \mathbf{h})\}$; in all cases, the physical model space is identified as a subspace of the extended model space - the identification is characteristic for each type of extended model. The variable \mathbf{h} is a (scalar or vector) parameter, such as shot coordinate, subsurface offset, or scattering angle, which characterizes additional degrees of freedom in the extended model space. For the shot record model extension, which will be used in this paper, $\bar{m}(\mathbf{x}, \mathbf{h})$ defines a member of M if it is independent of \mathbf{h} . For the subsurface offset model extension [Rickett and Sava, 2002], physical models are focused at $\mathbf{h} = 0$, that is, take the form $\bar{m}(\mathbf{x}, \mathbf{h}) = m(\mathbf{x})\delta(\mathbf{h})$.

The extended Born modeling operator is denoted $\bar{F}[m]$: note that background models are not extended. For shot record model extension $\mathbf{h} = \mathbf{x}_s$, the value of $\bar{F}[m]$

applied to $\delta\bar{m} = \delta\bar{c}^2$ is the solution of the following wave equation sampled at \mathbf{x}_r :

$$\left(\frac{1}{c(\mathbf{x})^2} \frac{\partial^2}{\partial t^2} - \Delta \right) \delta u(\mathbf{x}, t; \mathbf{x}_s) = \frac{\delta\bar{c}^2(\mathbf{x}, \mathbf{x}_s)}{c^2(\mathbf{x})} \Delta u(\mathbf{x}, t; \mathbf{x}_s), \quad (2.5)$$

with $u(\mathbf{x}, t; \mathbf{x}_s)$ the solution of equation 2.2. Note the only difference between equation 3.3 and equation 3.2 is that the perturbation $\delta\bar{c}^2(\mathbf{x}, \mathbf{x}_s)$ depends on shot record, i.e. perturbation is allowed to be different for different shot. And consequently, the adjoint operator $\bar{F}[m]^T$ will output a prestack image, i.e. a reflectivity model that also depends on \mathbf{x}_s , while the adjoint operator $F[m]^T$ outputs a stacked image.

The extended Born waveform inversion (EBWI) problem is: given data d , find $m \in M$, $\delta\bar{m} \in \bar{M}$ that minimizes

$$J_{\text{EBWI}}[m, \delta\bar{m}] = \frac{1}{2} \|\bar{F}[m]\delta\bar{m} - d\|^2 + \frac{\alpha^2}{2} \|A\delta\bar{m}\|^2. \quad (2.6)$$

The second term in this sum involves an operator A whose null space is precisely the physical models M . Such operators are usually called *annihilator* [Brandsberg-Dahl et al., 2003]. Minimizing it drives extended models toward physical (non-extended) models.

For shot coordinate model extension, $\mathbf{h} = \mathbf{x}_s$ and

$$A = \nabla_{\mathbf{x}_s}.$$

Thus physical models are exactly those extended models that are constant in shot coordinate. The weight α in equation 2.6 controls emphasis on physicality: as $\alpha \rightarrow \infty$, the minimization of J_{EBWI} resembles more and more minimization of J_{BWI} [Gockenbach et al., 1995].

2.3.3 Variable Projection

Both objective functions 2.3 and 2.6 will suffer from the same local minima problem as the FWI objective function 2.1 if we use gradient based optimization methods to solve them. BWI and EBWI objective function are quadratic in the reflectivity model. Thus we could use a nested optimization approach: namely the variable projection method [Golub and Pereyra, 2003] to solve them.

The variable projection reduced objective for Born waveform inversion, $J_{VP}[m]$ is the least value attained by $J[m, \delta m]$ over the model space of reflectivity δm :

$$J_{VP}[m] = \min_{\delta m \in M} J_{BWI}[m, \delta m]. \quad (2.7)$$

[van Leeuwen and Mulder, 2009] and [Xu et al., 2012] have proposed closely related objective functions.

Similarly, define a variable projection objective function for extended Born modeling by

$$J_{VPE}[m] = \min_{\delta \bar{m} \in M} J_{EBWI}[m, \delta \bar{m}]. \quad (2.8)$$

The analysis of VP objective function is similar, but simpler than VPE objective function. Thus we use $J_{VPE}[m]$ as an example. The value of J_{VPE} at a given background velocity m is the minimum value of J_{EBWI} over $\delta \bar{m}$ for fixed m, d . Since J_{EBWI} is quadratic in $\delta \bar{m}$, its minimum value $J_{VPE}[m]$ is attained at $\delta \bar{m}$ which solves the *normal equation*

$$N[m]\delta \bar{m} = \bar{F}[m]^T \delta d, \quad (2.9)$$

with $N[m] = \bar{F}[m]^T \bar{F}[m] + \alpha^2 A^T A$.

The numerical solution $\delta \bar{m}$ of equation 2.9 involves an iterative process, which was referred to as PICLI method in [Ehinger and Lailly, 1993] for shot coordinate

dependent operator A . In results reported in the numerical examples section, equation 2.9 both with and without model extension is solved using conjugate gradient (CG) method.

Under some circumstances, the normal operator $N[m]$ is closely related to a *pseudo-differential operator*, a type of oscillatory integral operator. This is principal because the modeling Hessian $\bar{F}[m]^T \bar{F}[m]$ has this property when the source of the forward map \mathcal{F} is an impulse, which leads to the smoothness of the VPE objective function 2.8. Thus a gradient based optimization method is preferable for the background model update. This relationship is contingent on kinematic and dynamic assumptions on the model and data: for example, for some extensions (shot-record) multiple ray paths connecting sources and receivers with scattering points may not occur [Rakesh, 1988a, Symes, 1998].

Assume that the inner product in the background model space takes the form $\langle \bar{m}_1, \bar{m}_2 \rangle_{\bar{M}} = \langle \bar{m}_1, \Lambda \bar{m}_2 \rangle$, where $\langle \cdot, \cdot \rangle$ is the ordinary Euclidean inner product and Λ is a weight or roughening operator, chosen to enforce smoothness (slow variation) of the background models, since the Born modeling operator is more accurate when the background model is smooth. In the examples below, Λ is a power of the Laplace operator

$$\Lambda^{-1} = (I - L)^{-1},$$

with

$$L = \left(\omega_x^2 \frac{\partial^2}{\partial x^2} + \omega_z^2 \frac{\partial^2}{\partial z^2} \right). \quad (2.10)$$

(see Appendix A for the implementation of Λ). Assuming also that $\delta \bar{m}$ solves the normal equation 2.9, then the gradient of J_{VPE} (see Appendix B for derivation) has the following form

$$\nabla J_{\text{VPE}}[m] = \Lambda^{-1} D\bar{F}[m]^T [\delta \bar{m}, \bar{F}[m] \delta \bar{m} - d] \quad (2.11)$$

The $D\bar{F}[m]^T$, which is the transposed second order derivative of the extended forward modeling operator $D^2\bar{\mathcal{F}}[m]^T$, has been called the tomographic or WEMVA operator [Biondi and Sava, 2004, Biondi and Almomin, 2012]. Note that both $D\bar{F}[m]$ and $D\bar{F}[m]^T$ depend on the background model parameter. The gradient of $J_{VP}[m]$ has the same form, but without model extension.

2.4 Numerical Examples

We create a Born model $(m, \delta m)$ by modifying the 2D Marmousi model [Bourgeois et al., 1991]. In this case, m is the velocity-squared field, and δm is its perturbation. We extended the water layer to 450 m depth, then smoothed the velocity-squared field with a moving box average to produce the background velocity-squared model. The reflectivity (δm) is the result of subtracting from the Marmousi model a less aggressive smoothing. Both m and δm were resampled to the 16 m \times 16 m simulation grid. Then Born data were computed by solving the perturbational wave equation 3.2 using a centered finite difference scheme of order 2 in time and 4 in space. The source is a (finite difference version of) an isotropic point radiator with 6 Hz peak frequency Ricker wavelet as in Figure 3.3a. Trace data were synthesized for 110 shots starting from 1 km with spacing 64 m, depth 6 m. The receiver spread is symmetric about zero offset, with 481 receivers spaced 16 m apart. Receiver depth is 10 m.

Before we show our inversion results, several aspects of the VPE method need to be addressed. First, the formulation of the VPE method has several parameters, among which α (the weight on DSO term) is the most important one. We set α to be 0.01 in all cases, on the basis of trial-and-error: the number of CG iterations used to solve 2.9 is fixed at 50 and we plot the value of J_{VPE} with different value of α over a line segment of the background model and choose the one that results in a convex curve. This α is then adopted to EBWI problem.

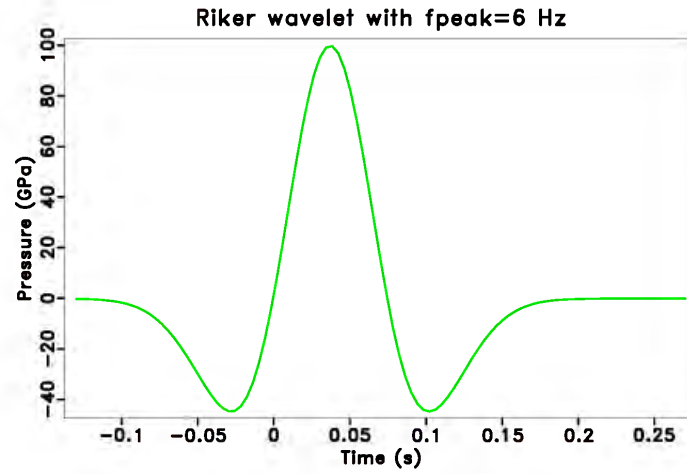


Figure 2.1: Marmousi example: Source Ricker wavelet with 6 Hz peak frequency.

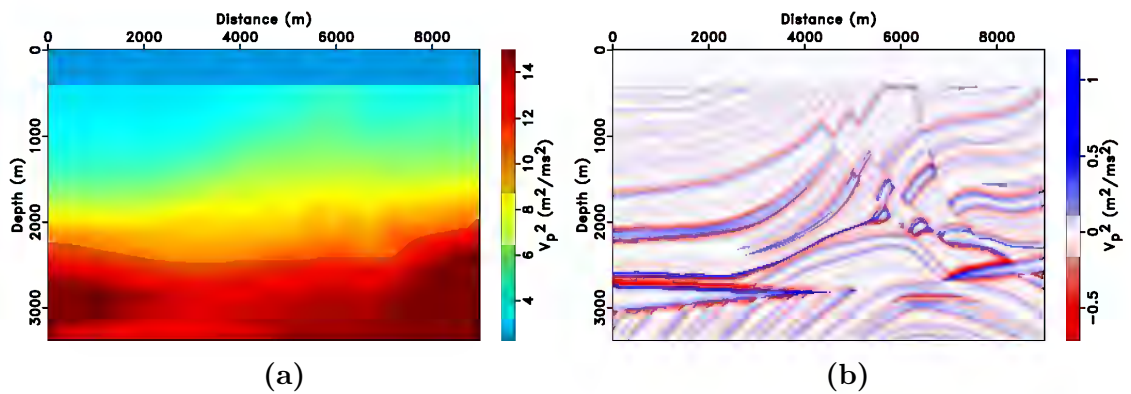


Figure 2.2: Marmousi example: Target background model (a) target reflectivity model (b)

Next parameters are ω_x and ω_z in the smoothing operator Λ^{-1} in equation 2.10. Values of ω_x and ω_z control the degree of smoothness in the corresponding directions to the gradient computed as in equation 2.11. Our rule is to choose one that is close to the dominant wavelength in the data. We fix $\omega_x = \omega_z = 400$ m in all of the following examples.

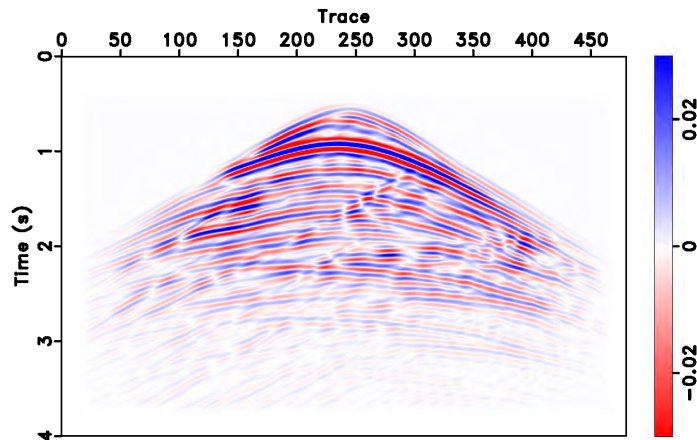


Figure 2.3: Marmousi example: Born shot record with index 41.

Then, we illustrate that it tends to be necessary to solve the normal equation 2.9 accurately enough with two tests: (a) plot the VPE objective function value along a line segment of the background model space and (b) compare the directional derivatives of J_{VPE} computed using analytic gradient formula 2.11 and finite difference approximation. However, a quantitative criterion for the accuracy requirement might be problem dependent. Background model in Figure 2.2a and reflectivity model in Figure 2.2b are used to produce Born trace data (Figure 2.3 shows the shot record with index 41), which is used as the observed data in equation 2.8. We use conjugate gradient method to solve the linear problem 2.9 and evaluate $J_{\text{VPE}}[m]$ along a line segment

$$m = (1 + \sigma)m_{\text{target}} - \sigma m_0$$

in the background model space, with m_{target} the target model shown in Figure 2.2a

and $m_0 = (1.5\text{km/s})^2$. $\sigma = 0$ is the position of target background model and $\sigma = -1$ is the constant model. Different number of conjugate gradient iterations are performed. With only 10 steps of iteration, a lot of testing models give smaller objective function value than the target model. By increasing the number of CG iterations, the minimum points along though curves move closer and closer to the target model. With either 50 or 60 iterations of CG, the global minimum along this line segment is attained at the target model. Enough number of iterations, which results in accurately enough solution to the normal equation, appears to be necessary for a successful VPE inversion. And the closer the model is to the target model, the more iterations may be needed for a reasonable update. A similar behavior of the J_{VPE} has been observed for the following truncated Marmousi example.

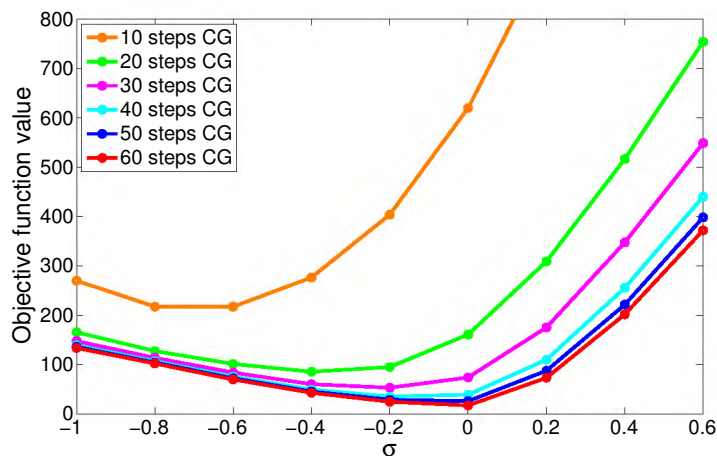


Figure 2.4: Marmousi example: value of VPE objective along line segment $m = (1 + \sigma)m_{\text{target}} - \sigma m_0$ with m_{target} shown in Figure 2.2a and $m_0 = (1.5\text{km/s})^2$ with $\sigma = 0$ the position of target background model.

we plot objective function values with different number of iterations of the inner optimization problem along a line segment, on which the target background model lies. The results indicate that without enough iteration, there are models that have smaller objective function value than the target model, while with enough number of iterations, the minimum objective function value is attained at the target model, at

least along this line segment.

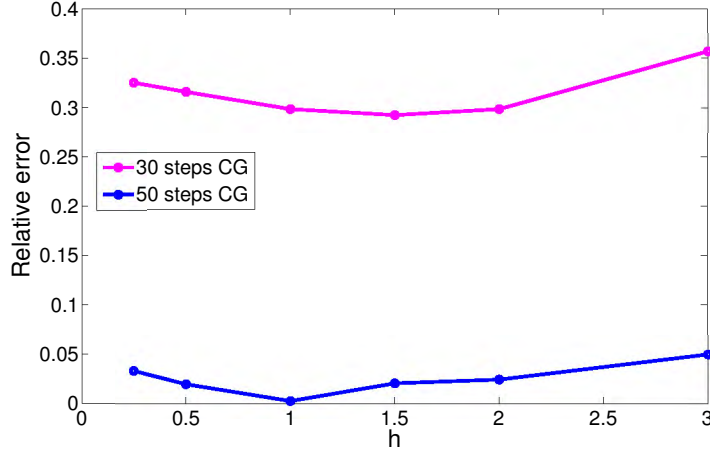


Figure 2.5: Marmousi example: Relative error between $\langle \nabla J_{\text{VPE}}[m], dm \rangle_s$ and $\frac{J_{\text{VPE}}[m + h * dm] - J_{\text{VPE}}[m - h * dm]}{2h}$ with the first one as the value of reference. $m = 0.7m_{\text{target}} + 0.3m_0$ and $dm = 0.1 \text{ (km/s)}^2$ with m_{target} shown in Figure 2.2a and $m_0 = (1.5 \text{ km/s})^2$.

In the gradient accuracy test, we compute the relative error of the following two quantities with different h , with the one on the right hand side as the value of reference

$$\frac{J_{\text{VPE}}[m + h * dm] - J_{\text{VPE}}[m - h * dm]}{2h} \approx \langle \nabla J_{\text{VPE}}[m], dm \rangle_s \quad (2.12)$$

With small number of CG iterations, the relative error could be large, regardless of the value of h . With 50 steps of CG iteration, where the residual in equation 2.9 has been reduced to less than 1 percent of its initial error, the relative error in gradient accuracy test drop to below 1 percent, with $h = 1$ achieving the best precision (below 0.3 percent) as shown in Figure 2.5. Theoretically, Taylor expansion indicates that the order of the finite difference approximation is 2 in h .

$$\frac{J_{\text{VPE}}[m + h * dm] - J_{\text{VPE}}[m - h * dm]}{2h} = \langle \nabla J_{\text{VPE}}[m], dm \rangle_s + C \frac{h^2}{6},$$

with $C = D^3 J_{\text{VPE}}[m_1] dm^3 + D^3 J_{\text{VPE}}[m_2] dm^3$ and m_1, m_2 are some models in the background model space. However, by decreasing the value of h , we do not get the order when h is smaller than 1 in this example. The reason is that the evaluation of J_{VPE} involves CG iterations, which results in an error in the value of J_{VPE} at any background model. What we really get is

$$\frac{J_{\text{VPE}}[m + h * dm] - J_{\text{VPE}}[m - h * dm] + \epsilon_1}{2h} = \langle \nabla J_{\text{VPE}}[m], dm \rangle_s + \epsilon_2 + C \frac{h^2}{6}$$

Note that ϵ_1 and ϵ_2 does not depend on h and depends only on the number of CG iterations, which is fixed. Thus when h decreases, $\frac{\epsilon_1}{2h}$ will increase. And when h is smaller than a certain number, these noise (or error) becomes dominant.

This gradient accuracy test shows that in order to get an accurate enough gradient, and thus a successful inversion, we need to solve the normal equation accurately.

At last, two inversion examples will be presented. The first one compares inversion with and without variable projection. We give only the results for EBWI, as those for BWI are similar. The second example compares extended and non-extended variable projection. on the basis of trial-and-error. Our principal quality control display will include: 1, the stack of the extended reflectivity

$$\delta m_{\text{stack}}(\mathbf{x}) = \sum_h \delta \bar{m}(\mathbf{x}, h)$$

in comparison to the similar stack computed at the target velocity-squared model: if events appear in the same positions with roughly the same amplitude, the inversion is successful; 2, common image gathers (CIGs) [Al Yahya, 1989]: if gathers are as flat and at the same positions as that of the image computed at the true background model, the inverted velocity-squared model is considered to be the global solution.

In the first example, we use 2.6 s data and truncate the model at 2 km depth.

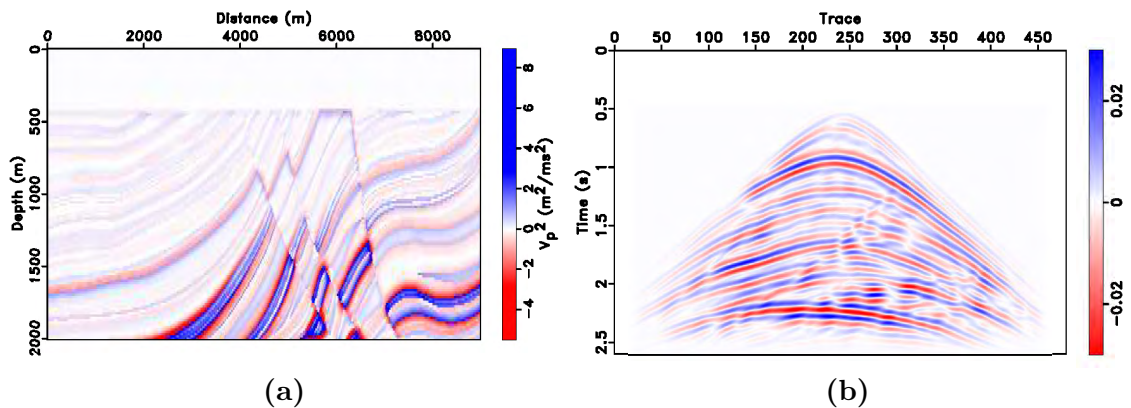


Figure 2.6: Truncated Marmousi example: reflectivity model (a) and Born shot record, shot index 41 (b).

The target (correct) background velocity-squared model for this example appears in Figure 3.1a. Figure 2.6a shows the target reflectivity model used to compute the Born data shown in Figure 3.3b. The initial velocity-squared chosen for the first example (Figure 3.23) is a linear combination: 70% of the target model and 30% homogeneous “water” ((1.5 km/s)²). Because we have limited the depth range to 2 km and the maximum recording time to 2.6 s, the kinematic deviation from the target model is not great: this is a relatively easy velocity estimation problem. The first difference is however great enough that the stack δm_{stack} incorrectly positions reflectors and faults: compare Figure 2.8a the inverted reflectivity model at true background model and Figure 2.8b the inverted reflectivity at initial background model. Common image gathers of the inverted reflectivity at initial background model (Figure 2.8b indicates that the initial background model is smaller than the target background model all over the domain.

We approximate minimization of J_{VPE} by a very crude optimization algorithm: steepest descent with bisection backtracking line search. The gradient is computed by evaluating formula 2.11 using finite difference implementation of the adjoint state method, adapted to compute the tomographic operator $D\bar{F}[m]^T$ [Symes and Santosa, 1988]. After 7 steps of this process, a lot of reflectors which we could not see at the initial

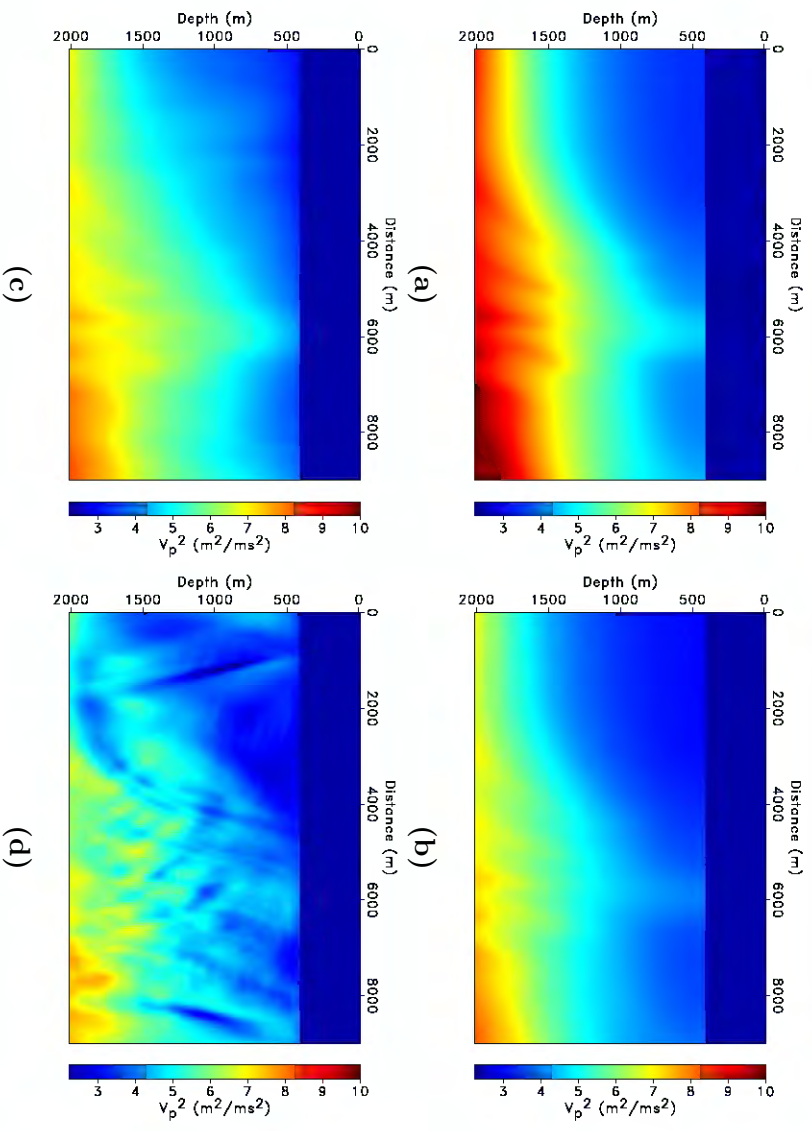


Figure 2.7: Truncated Marmousi example: true velocity-squared background model (a); initial background model (b); inverted background model from 7 steps of VPE method (c) and 350 steps of EBWI method (d)

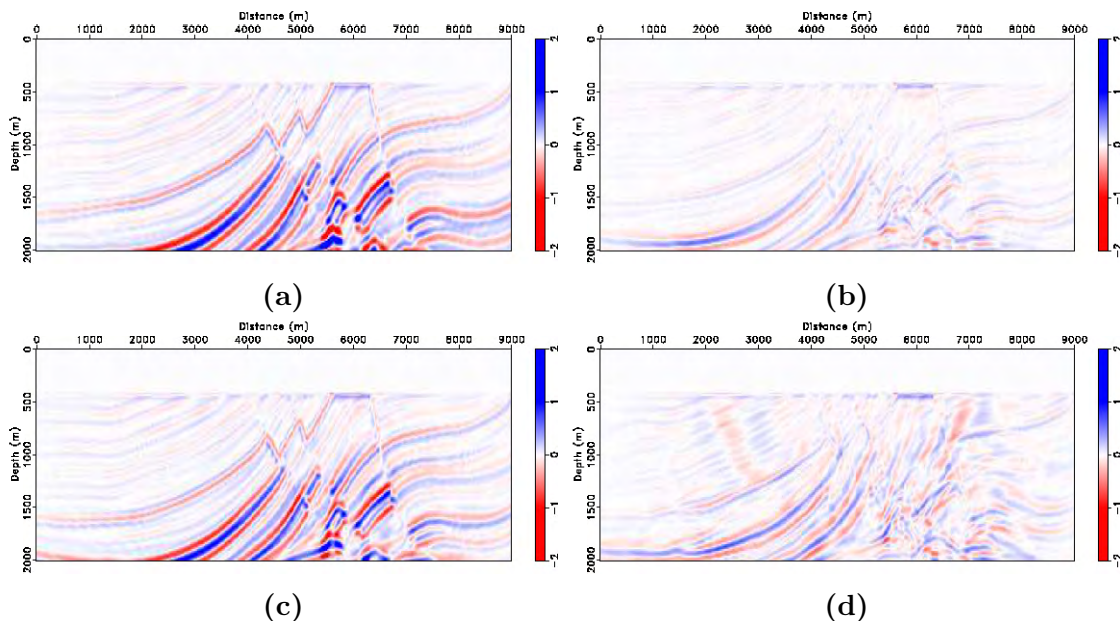


Figure 2.8: Truncated Marmousi example: Inverted reflectivity model at true background model (a); initial background model (b); background model with 7 steps of VPE method by solving equation(2.9) (c) and 350 steps of EBWI method (d)

model appear, and most reflectors positions are correct (Figure 2.8c). Common image gathers at the inverted model (Figure 2.9c) resembles that at the target model (Figure 2.9a) very closely except a small difference at the bottom of the gathers around 2 - 2.5 km positions.

We also attempt minimization of J_{EBWI} (equation 2.6) simultaneously for m and δm , without reduction by variable projection. We use the Limited Memory BFGS algorithm [Nocedal and Wright, 1999], with the same backtracking line search used in the VPE minimization. We re-start the algorithm every 50 steps: each block of 50 steps has cost similar to that of one variable projection iteration, so 350 steps of LBFGS is roughly equivalent to 7 steps of variable projection optimization. The inverted velocity-squared background model obtained at step 350 is displayed in Figure 2.7d. The update of the velocity-squared shows no hint of kinematic correction: all reflectors remain essentially in their initial, incorrect positions (Figure 2.8d). More iterations will not help correcting the positions of reflectors. Common image gathers

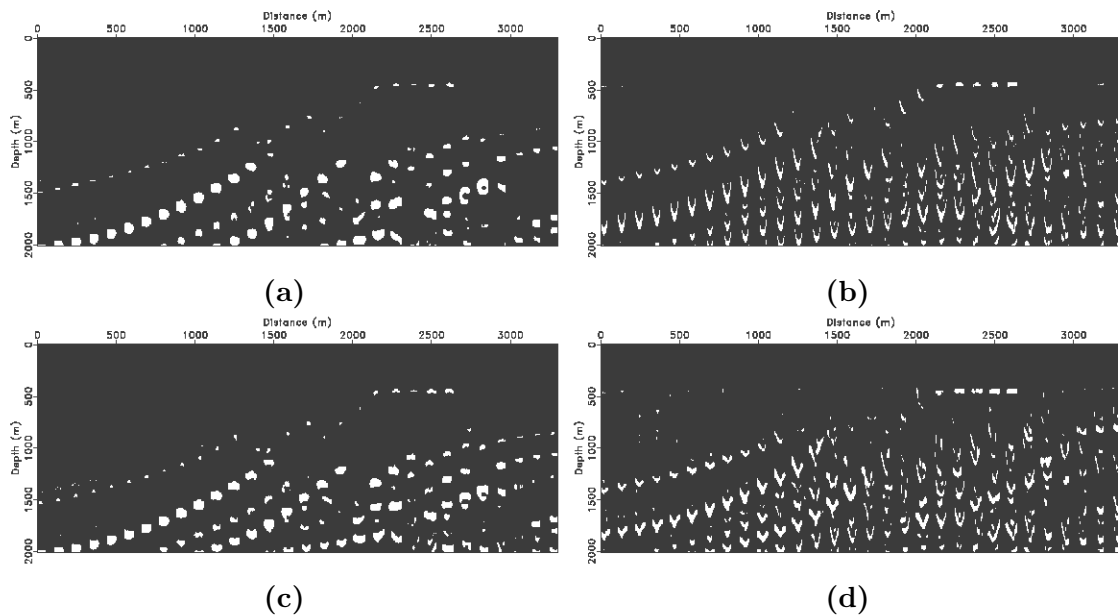


Figure 2.9: Truncated Marmousi example: Common image gathers at true background model (a); initial background model (b); b ackground model with 7 steps of VPE method by solving equation 2.9 (c) and 350 steps of EBWI method (d).

(Figure 2.9d) at inverted background model is very similar to that at the inital model.

From this example, we conclude that minimization of J_{EBWI} for $m, \delta\bar{m}$ jointly is unlikely to succeed, whereas minimization of the variable projection function J_{VPE} produces a useful velocity update. A similar conclusion holds for J_{BWI} and J_{VP} .

The second example compares extended and non-extended variable projection, that is, minimization of J_{VPE} and J_{VP} . Since both appear successful for “easy” velocity estimations, we create a more difficult problem by using more data (4 s as shown in Figure 2.3), the full depth range (3.5 km) in the model (Figure 2.2a), and choosing a more drastically incorrect initial model, namely a homogenous $v = 1.5$ km/s. The stack δm_{stack} of the inverted extended reflectivity at the initial model is weak in amplitude with reflector positions are in error throughout (compare Figure 2.11a and Figure 2.11b).

For both optimizations, we use steepest descent with backtracking line search as described above. The solution of equation 2.9 for both VP and VPE is estimated by

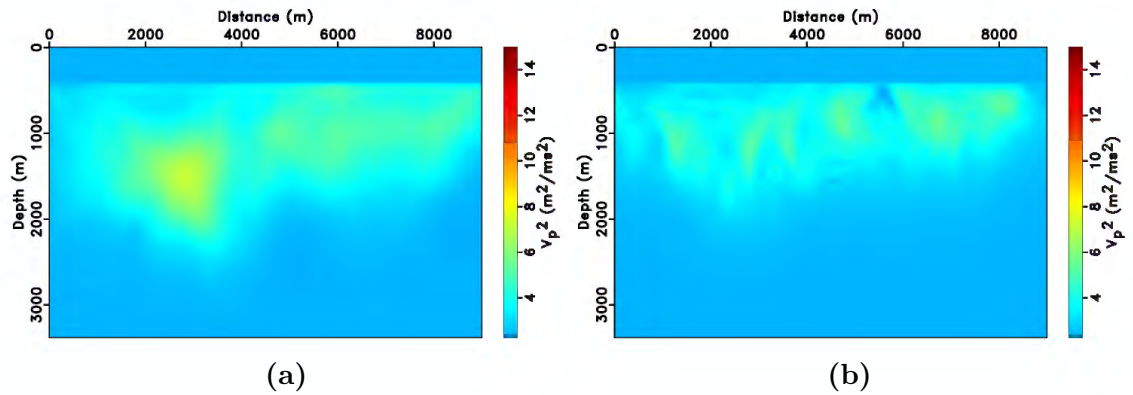


Figure 2.10: Marmousi example: Inverted background model from 18 steps of VPE method (a) and 18 steps of VP method (b).

the CG method with number of iterations increasing with background model update. This is based on our observation that when the background velocity is far from the true model, an inaccurate gradient estimate is still adequate to produce an acceptable update. In all cases, however, the residual error in normal equation 2.9 is reduced to less than 5 percent of its initial size.

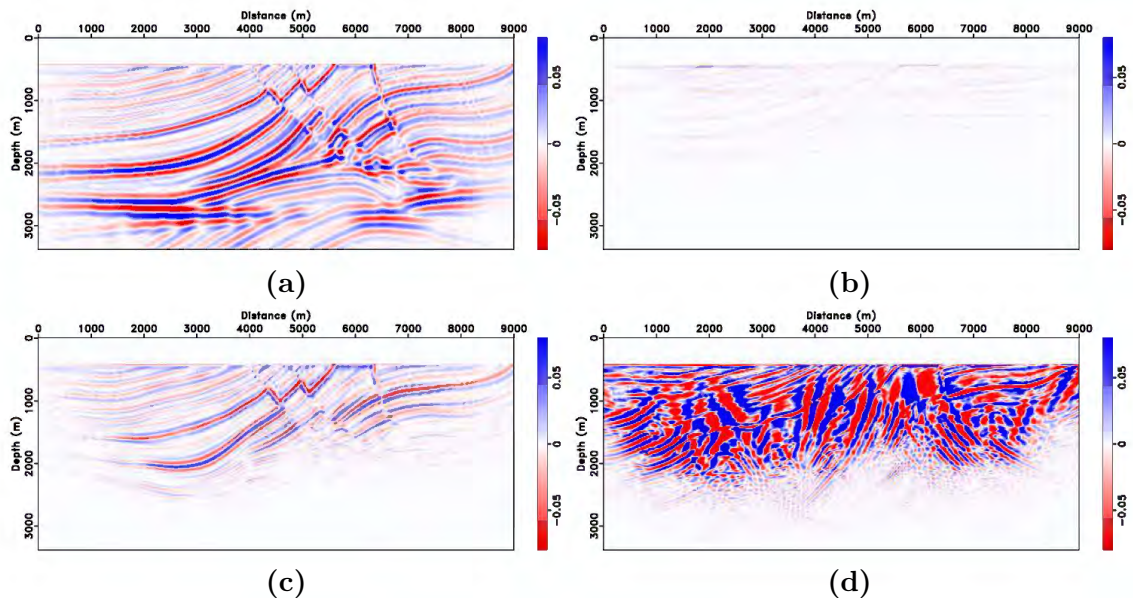


Figure 2.11: Marmousi example: Inverted reflectivity model at true background model (a); initial background model (b); background model with 18 steps of VPE method (c) and 18 steps of VP method (d) by solving equation 2.9.

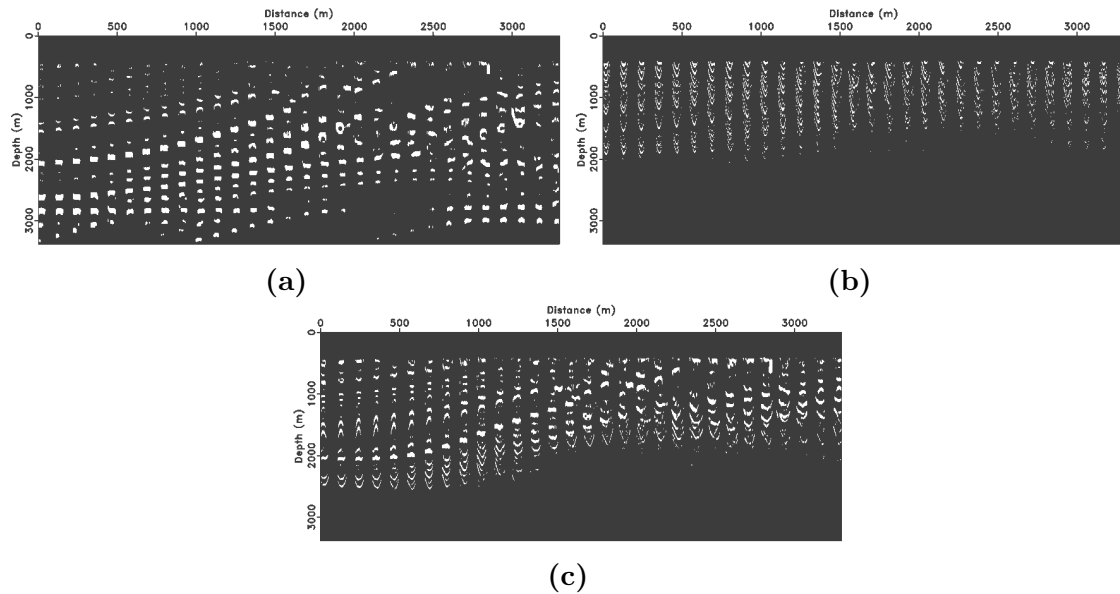


Figure 2.12: Marmousi example: Common image gathers at true background model (a); initial background model (b); background model with 18 steps of VPE method (c) by solving equation 2.9.

After 18 steepest descent steps for J_{VPE} each involving 5 - 30 CG iterations per function evaluation and several backtrack steps, positions of events down to about 1.5 km, somewhat deeper on the left, are largely corrected and amplitudes are considerably improved in the stacked inversion (Figure 2.11c). Common image gathers in Figure 2.12c at the inverted background model are more clear than that at the initial model 2.12b. Although the inverted model at iteration 18 is not close to the true model at below 1.5km, the trend is promising.

The first few VP updates, on the other hand, have partly repositioned the shallowest reflectors, subsequent steps do not improve the kinematics of the image and do not correct deeper events that are well-positioned by VPE (compare Figures 2.11c and 2.11d).

Figure 2.13 shows that VPE method fits the data well for both near and far offset traces, while VP method fails to fit data of medium and far offset traces. VP itself does not solve local minima problem for this example. Model extension appears to

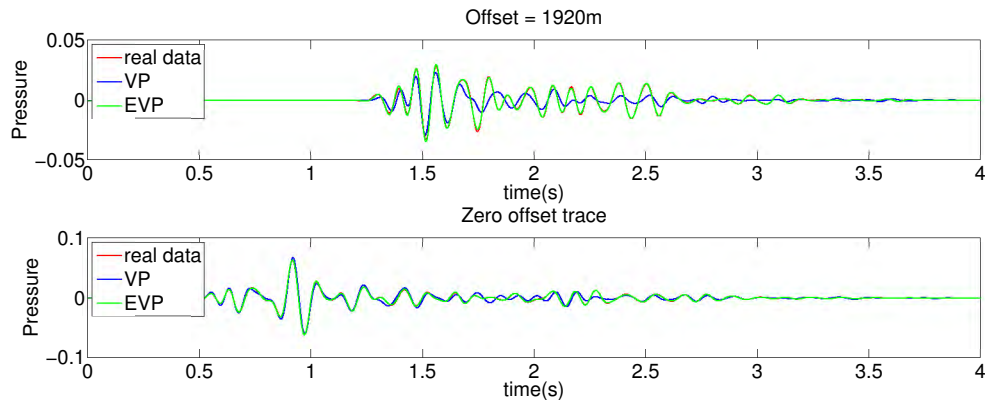


Figure 2.13: Marmousi example: Trace comparison of real data (red), predicted data by VPE (green) and VP (blue) methods for far (top) and near (bottom) offsets.

be necessary for a global convergence.

2.5 Discussion

The examples presented in the last section suggest that both extended modeling, which permits data fit throughout the inversion process, and variable projection, which enforces it, are critical ingredients in waveform inversion. Without both of these ingredients, gradient-based algorithms fail to constructively update kinematically inaccurate initial guesses. In particular, since the variable projection algorithms for Born and extended Born inversion involve similar numbers of migration/modeling pairs, the extended variant would appear to be the superior choice.

The Achilles' heel of this approach is its overall cost: the number of modeling/migration cycles required for the rather simple 2D examples presented here was in the hundreds, and such computational largesse is likely infeasible for industry-scale problems. Most of the cycles in these exercises go into the iterative solution of the normal equation 2.9, which must be fairly precise in order that the global minimum is obtained at the target background model and the error in the gradient formula 2.11 be controlled. Convergence of CG and other iterative methods can be accelerated

through the use of preconditioning, that is, inexpensive approximate inverse operators. Several methods for constructing preconditioners for 2.9 have been proposed [Tang, 2009, Stolk et al., 2009a, Nammour and Symes, 2009,], and should be evaluated for their potential to accelerate the VPE algorithm.

Flexibly Preconditioned Extended Least Squares Migration in Shot Record Domain (submitted to *Geophysics*)

3.1 Summary

This paper demonstrates a method for accelerating the convergence of iterative least squares migration. The algorithm uses a pseudodifferential scaling (dip- and spatially-varying filter) preconditioner together with a variant of conjugate gradient iteration with iterate-dependent (“flexible”) preconditioning. The migration is formulated without the image stack, thus producing a shot dependent image volume that retains offset information useful for velocity updating and amplitude-versus-offset analysis. Numerical experiments show that flexible preconditioning with pseudodifferential scaling not only attains considerably smaller data misfit and gradient error for a given computational effort, but also produces higher resolution image volumes with more balanced amplitude and fewer artifacts than is achieved with a non-preconditioned conjugate gradient method.

3.2 Introduction

While prestack reverse time migration (RTM) [Baysal et al., 1983, Loewenthal and Mufti, 1983, Mulder and Plessix, 2004] can produce accurate images of complex subsurface structure, it can also suffer from unbalanced amplitude and illumination artifacts, acquisition footprint, and imperfect focusing of the seismic wavelet. Least squares RTM, or LSM, an alternate name for iterative least-squares linearized inversion [Bourgeois et al., 1989, Chavent and Plessix, 1999, Nemeth et al., 1999, Tang, 2009, Aoki and Schuster, 2009, Dutta et al., 2014, Luo and Hale, 2014, Valenciano et al., 2015], can resolve the aforementioned problems of RTM (its first iteration) to some extent. Since each iteration requires a migration-modeling pair, the expense of LSM can be considerable, yet achieving its benefits requires that sufficient iterations be performed.

In this paper, we show how to accelerate iterative LSM by combining optimal pseudodifferential scaling [Symes, 2008a, Nammour and Symes, 2009, Nammour, 2009] with the Flexibly Preconditioned Conjugate Gradient (FPCG) algorithm [Notay, 2000, Knyazev and Lashuk, 2006]. Our algorithm is designed to invert each shot record individually, and thus produce image gathers that can be used for velocity and amplitude-versus-offset analysis. We argue that the theoretical properties of this algorithm suggest that it should substantially reduce the number of iterations required for good data fit and small LSM gradient, and illustrate this conclusion with a number of examples showing the benefits of accelerated LSM in better image amplitude and phase for a given computational effort.

Many previous works have addressed computational efficiency of LSM, or linearized inversion. Our work belongs to the *scaling* genre, in which the (computable) action of the normal operator or Hessian $F^T F$ of the scattering operator F on an image or set of images is used to constrain a matrix representation, which is then inverted to approximately solve the normal linear system equivalent to least

squares minimization. The type of algorithm explored in this paper originated in the work of [Claerbout and Nichols, 1994], who computed an approximate inverse Hessian scale factor by point-wise division of migrated image $F^T d$ by the output of the Hessian applied migrated image, that is, $(F^T F)F^T d$. [Rickett, 2003] polished this idea and compared it with other alternatives for replacing the migrated image in the denominator. This type of approximation implicitly presumes that the Hessian acts mostly by space-dependent scaling of its input, that is, as a spatially diagonal operator: in effect, that the Hessian acts mostly to modify amplitudes. However, [Chavent and Plessix, 1999] showed that the Hessian is not spatially diagonal, or even diagonally dominant: that is, the Hessian is also a filter of non-zero width. They suggested taking this non-diagonality into account via a partial mass-lumping method, that is, adding the off-diagonal terms to the diagonal, and then scaling the migrated image by the inverse of resulting diagonal matrix. [Shin et al., 2001] suggested another narrow-band approximation to the normal operator, which they call the pseudo-Hessian matrix. [Guitton and Kaelin, 2006] replaced the diagonal multiplier of [Claerbout and Nichols, 1994] and [Rickett, 2003] with a localized filter, and demonstrated a more accurate approximate inversion in some cases. [Symes, 2008a] explained the non-diagonality observed by [Chavent and Plessix, 1999] and others: under well-understood conditions, the normal operator is the product of a known power of the Laplace operator, which is not a spatially diagonal operator, and another operator that really does act as a spatially diagonal operator, at least so long as the image has a well-defined dip field in most places, as seismic images of sedimentary structures tend to do. This paper also introduced an inverse problem for finding the optimal scaling operator, to replace pointwise division. In essence, optimal scaling is possible because the normal operator (LSM Hessian) is *pseudodifferential*, that is, a space-varying filter behaving like a polynomial in spatial frequency for

large frequencies [Beylkin, 1985, Rakesh, 1988b, ten Kroode et al., 1998, Stolk, 2000]. [Nammour and Symes, 2009, Nammour, 2009] used this observation to extend optimal scaling to imaging problems with conflicting dips, using an algorithm due to [Bao and Symes, 1996] for approximating pseudodifferential operators efficiently. [Herrmann et al., 2008] implemented the same idea using a different approach to computing pseudodifferential operators, based on their approximate diagonalization in curvelet frames. [Demagnet et al., 2012] used the low-rank property implicit in the pseudodifferential nature of the normal operator to approximately constrain a matrix representation by its action on a randomly chosen set of vectors, then efficiently inverting the matrix.

The works cited so far approximate the inverse normal operator, or approximately solve the normal equation. It is natural to think that an approximate inverse could be used to precondition iterative methods for faster convergence, and many of the ideas described in the last paragraph have been used in this fashion, beginning with [Chavent and Plessix, 1999]. We mention in particular [Herrmann et al., 2009], who based a preconditioning construction on the use of curvelet representation as in [Herrmann et al., 2008]. See [Pan et al., 2014] for a comparison of several scaling methods used as preconditioners.

Note that under some circumstances approximate least squares solutions can be constructed via asymptotic analysis. The seminal paper [Beylkin, 1985] led to much work on so-called Generalized Radon Transform inversion, which uses ray tracing quantities to build a Kirchhoff-type integral approximating a pseudo inverse to the scattering operator. See [Métévier et al., 2015, Lameloise et al., 2015] for recent examples of the use of asymptotic inversion based on ray-tracing to accelerate iterative LSM or full waveform inversion. Recently it has been recognized that asymptotic approximate inverses may be constructed without any ray-trace computations

whatsoever: see [Zhang et al., 2003, Stolk et al., 2009a, Hou and Symes, 2015b], and [Hou and Symes, 2015a] for the use of ray-free asymptotics for preconditioning.

The scaling operator concept (and our use of it) does not rely explicitly on high frequency asymptotics or ray tracing, or involve any ray computations. However its justification rests on the pseudodifferential property of the normal operator or Hessian of the scattering operator, and that in turn is a consequence of the ray asymptotics of the wave equation.

Our work differs from that described so far in two main respects. First, we accelerate the solution of the *shot record extension* of linearized (or “Born”) forward modeling, in which each shot is modeled by a reflectivity or model perturbation proper to that shot, while sharing a background or reference model with all other shots. The result of a least-squares Born inversion with this extended modeling principle is an inverted *extended* reflectivity volume, analogous to the prestack shot-record migration image volume - which is, in fact, the result of the first iteration of most iterative algorithms for solving the extended problem. The fixed horizontal position slices of this migrated image volume are known as *shot record common image gathers* or CIGs, and are crucial ingredients in migration-based velocity analysis and amplitude-versus-offset analysis. All of the works cited so far have concerned the non-extended version of the Born inverse problem. Our algorithm iteratively inverts the extended Born modeling operator. The extended reflectivity volume so constructed may be similarly sliced into higher-resolution and more artifact-free version of migration CIGs, as we shall illustrate with several examples.

The other major innovation in our work stems from the nature of scaling algorithms, which do not in general produce an approximate inverse of the modeling operator, but rather an operator that approximately solves the least squares problem with specific data. The scaling operator so produced likely does not solve the least

squares problem with any data other than that used to generate it. Thus use of a scaling operator as preconditioner requires an algorithm that admits a different preconditioner each iteration. In contrast, all of the work cited so far uses conventional (fixed) preconditioned iterations. Fortunately, algorithms using iteration dependent preconditioning, called Flexibly (or Variably) Preconditioned, have been studied in the computational mathematics literature [Notay, 2000, Knyazev and Lashuk, 2006]. We combine a flexibly preconditioned conjugate gradient iteration with the pseudodifferential approximation of [Bao and Symes, 1996] and the optimal coefficient selection of [Symes, 2008a] and [Nammour and Symes, 2009, Nammour, 2009] to substantially accelerate the conjugate gradient algorithm for shot-record Born inversion, and improve the quality of its output in the bargain.

We note that the preconditioning strategy presented in [Hou and Symes, 2015a] and related algorithms are based on the use of *subsurface offset extension*, which involves more expensive modeling and migration in comparison with ordinary modeling and migration, but functions well with highly refractive velocity models (“complex structure”, see [Stolk et al., 2009b]). In contrast, while the shot-record extension leads to accurate imaging only for mild lateral heterogeneity (no multipathing, see [Stolk and Symes, 2004]), it is no more expensive than standard Born modeling and RTM. Optimal pseudodifferential scaling doubles the cost of each iteration of LSM, but considerably reduces the required number of iterations, for significant overall reduction in computational cost, as we shall show.

The rest of the paper is organized as follows: In the theory section, we will briefly discuss the LSM in shot record domain, illustrate its ill-condition property, then review the pseudodifferential scaling optimization and describe a flexibly preconditioned conjugate gradient (FPCG) method to solve the extended LSM problem in shot record domain. In the numerical results section, we apply the FPCG method to

two simple synthetic problems and to a Born inversion problem derived from the Marmousi model, and compare its efficiency with ordinary conjugate gradient method. We end with a discussion of some possible future developments and applications of FPCG-accelerated extended LSM.

3.3 Theory

We choose constant density acoustics for modeling wave propagation in the derivation of the following theory and algorithm. The constant density acoustic wave equation with squared velocity $m(\mathbf{x}) = c^2(\mathbf{x})$ and isotropic radiator source with wavelet $w(t)$ at source position \mathbf{x}_s is

$$\left(\frac{\partial^2}{\partial t^2} - m(\mathbf{x})\nabla^2 \right) u(\mathbf{x}, t; \mathbf{x}_s) = w(t)\delta(\mathbf{x} - \mathbf{x}_s). \quad (3.1)$$

The shot-record extended Born approximation [Kern and Symes, 1994, Symes, 2008b] allows the model perturbation δm to depend on the shot-record parameter as well as spatial location \mathbf{x} . Since we use shot position \mathbf{x}_s to parametrize the isotropic point radiator in equation 3.1, the extended model perturbation takes the same functional form $\delta m(\mathbf{x}; \mathbf{x}_s)$ as the image volume of prestack migration sorted by shot record. The corresponding perturbational (or Born) wave equation is

$$\left(\frac{\partial^2}{\partial t^2} - m(\mathbf{x})\nabla^2 \right) \delta u(\mathbf{x}, t; \mathbf{x}_s) = \delta m(\mathbf{x}; \mathbf{x}_s)\nabla^2 u(\mathbf{x}, t; \mathbf{x}_s) \quad (3.2)$$

The *extended Born modeling operator* is defined by sampling the perturbation field δu at the receiver positions \mathbf{x}_r corresponding to each source:

$$F[m]\delta m(\mathbf{x}_r, t; \mathbf{x}_s) = \delta u(\mathbf{x}_r, t; \mathbf{x}_s). \quad (3.3)$$

The adjoint operator F^T of the extended Born modeling operator F is the reverse time migration (RTM) with the standard cross-correlation imaging condition [Tarantola, 1984b], but without the final stack over shot position: thus the output of F^T is a shot-dependent image volume field of the same type as δm . The corresponding adjoint operator of the non-extended Born modeling operator stacks the output of the extended adjoint over shot position to produce an image (function of spatial position).

With these notational conventions, the linearized least squares, or LSM, problem takes the following form: given a data perturbation d , find a model perturbation δm to minimize

$$\frac{1}{2} \|F[m]\delta m - d\|^2. \quad (3.4)$$

As (3.4) is quadratic in δm , the global minimizer is obtained at any stationary point. Thus the minimizer of the problem (3.4) is equivalent to the solution of the normal equations

$$F^T F \delta m = F^T d, \quad (3.5)$$

Since we will fix the background model m in almost all the cases, we write the extended Born modeling operator as F instead of $F[m]$ for simplicity. Due to the size of the problem, the computation of $(F^T F)^{-1}$, or solution of equation 3.5, by any variant of Gaussian elimination is infeasible even for 2D field-scale examples. Thus iterative methods are used to solve equation 3.5. Most of these converge at rates that depend on the spectrum of the normal operator $F^T F$, and are faster when the spectrum has small extent. However $F^T F$ tends to be ill-conditioned (have a spectrum spread over a large interval) for several reasons. The limited aperture of field survey geometry implies that some structural perturbations have no appreciable impact on the Born seismogram, and imply existence of very small eigenvalues. The dynamic range of geometric spreading also tends to spread the spectrum. Most dramatic in effect

is the frequency-dependent scaling between model and data perturbation: in 2D, an oscillatory localized plane wave component of δm , of spatial frequency k , results in an output of $F^T F$ amplified by $O(k)$. The analogous amplification in 3D is $O(k^2)$. This frequency-dependent scaling has been understood since [Beylkin, 1985]. In effect, the 2D version of $F^T F$ has the square root of the Laplace operator as a factor, whereas the 3D version has the Laplace operator itself. In the continuum limit, the spectrum is without bound; for discretization, the largest eigenvalue grows in proportion to the Nyquist frequency.

Ill-conditioning due primarily to the frequency-dependent scaling of $F^T F$ has a disastrous effect on convergence of iterative methods for solution of equation 3.5. While the analysis simply suggests that the condition of equation 3.5 grows worse with increasing frequency, numerical experiment suggests that it is quite bad already for length and frequency scales characteristic of exploration seismology. We illustrate this effect with a simple example, in which we attempt to invert for a model perturbation δm representing a single reflector in the presence of a slow Gaussian anomaly in the background model, as shown in Figures 3.1a and 3.1b.

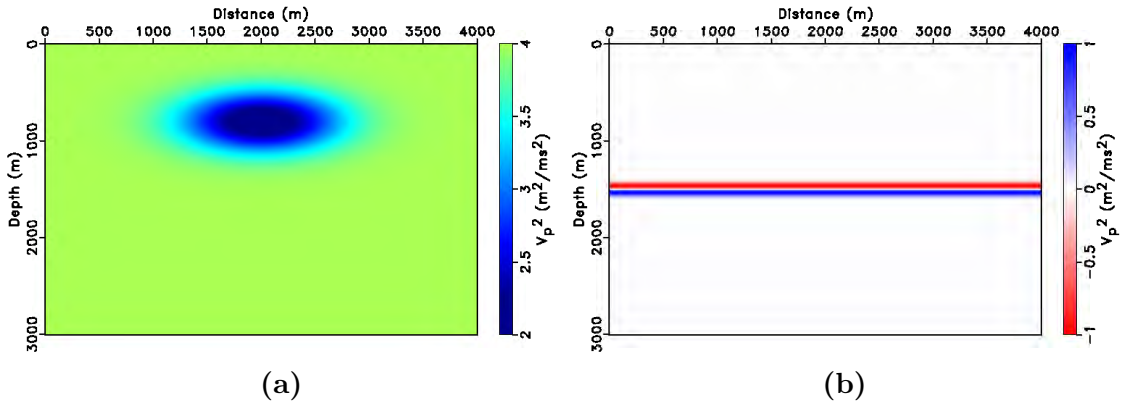


Figure 3.1: Slow Gaussian anomaly model: (a) background m , (b) perturbation δm .

We use point sources with impulse wavelets bandpass filtered with corner frequencies 1 Hz, 7 Hz, 28 Hz, 35 Hz. A total of 81 point sources are placed at depth 20 m

with a spacing of 50 m. 401 fixed receivers with a spacing 10 m are placed at depth 10 m. Data is recorded until 2.4 s.

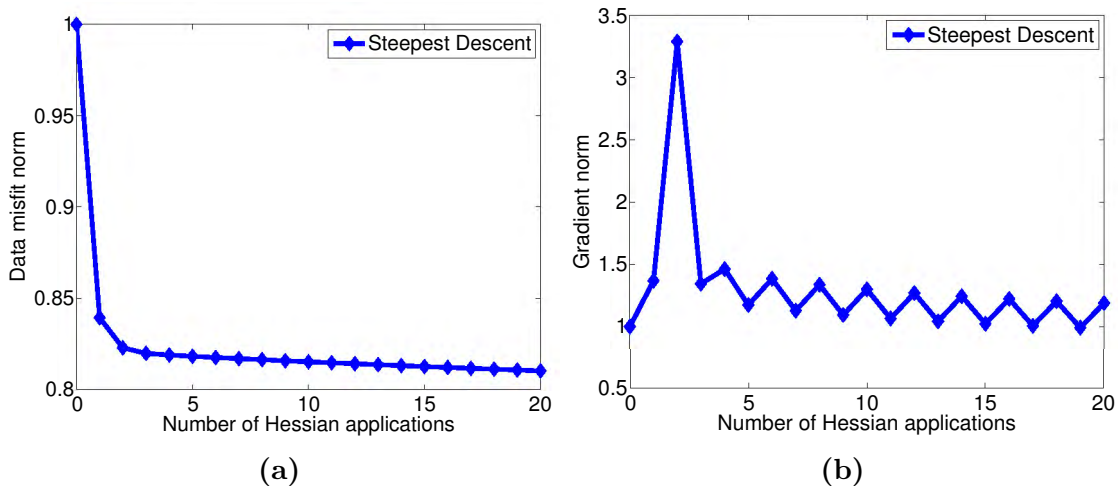


Figure 3.2: Convergence curve of steepest descent method for the slow Gaussian anomaly problem: (a) normalized data misfit (b) normalized gradient length.

We use a very simple iterative algorithm, the steepest descent method with fixed step length chosen as a crude estimate of the reciprocal maximum eigenvalue of $F^T F$ [Boyd and Vandenberghe, 2004]. Although this choice of step length guarantees decrease of the objective function value at each iteration, very slow convergence rate is expected if the condition number of $F^T F$ is large. The data misfit history (Figure 3.2a) shows that there is no big decrease in the error after 3 steps of iteration. The zigzag decrease of the normal residual (length of the difference between the left hand side and right hand side in equation 3.5, Figure 3.2b) is typical of the steepest descent method and indicates the ill-conditioning property of the normal operator $F^T F$.

This simple example strongly indicates that a reformulation of equation 3.5 is required for reasonably fast convergence of iterative solvers.

3.3.1 Pseudodifferential Scaling

The second author introduced an approximate inversion by scaling [Nammour and Symes, 2009, Nammour, 2009], based on a fundamental observation about the normal operator $F^T F$ established by [Beylkin, 1985, Rakesh, 1988b, ten Kroode et al., 1998, Stolk, 2000], namely that $F^T F$ is a *pseudodifferential operator* (Ψ DO) under some conditions, a synopsis of which appears at the end of this section. This fact underlies the effectiveness of migration as an imaging algorithm.

We point out several important properties of Ψ DOs:

- their action preserves the location and orientation of localized oscillatory plane waves;
- to each is associated an *order*, which captures the power of frequency with which localized plane waves are scaled; and
- they are expressed via an oscillatory integral, examples below, and take the form of a spatially variable filter.

The first of these facts explains the connection between the Ψ DO nature of the normal operator and the imaging property of the RTM operator F^T : its effect on the data created by $F\delta m$ is to recover localized oscillatory plane waves components of δm in location and orientation.

The oscillatory integral form of the Ψ DO $F^T F$ is

$$F^T F \delta m(\mathbf{x}) \approx \int a(\mathbf{x}, \mathbf{k}) \hat{\delta m}(\mathbf{k}) e^{i\mathbf{k} \cdot \mathbf{x}} d\mathbf{k} \quad (3.6)$$

The function of position and wavenumber $a(\mathbf{x}, \mathbf{k})$ is the *principal symbol* of $F^T F$; a is homogeneous of order s in wavevector \mathbf{k} (identical to the order of the operator, mentioned above) and depends on the background model m . Note that if a were

independent of \mathbf{x} , then equation 3.6 would define a filter, and thus could be computed inexpensively via FFT. However, the dependence on \mathbf{x} is essential: includes geometric amplitudes and other factors. The theory of Ψ DOs [Taylor, 1981] requires that a is *smooth*, that is, with arbitrarily many well-defined derivatives, in both \mathbf{x} and \mathbf{k} . Smoothness of m implies this property for the principal symbol of $F^T F$, under some assumptions that will be detailed below.

The order s is dimension-dependent: for dimension d , the order is $s = d - 1$ [Rakesh, 1988b]. The sense of approximation suggested in equation 3.6 is asymptotic, in the sense that the difference between the two sides is $O(|\mathbf{k}|^{s-1})$, in terms of spatial frequency.

From here on, we will consider the 2D case only. In particular, $s = 1$.

The pseudodifferential scaling problem [Nammour and Symes, 2009, Nammour, 2009] constructs an approximate solution of equation 3.5: for a given image, for example $b = F^T d$ with d the observed data, find a pseudodifferential operator Q with principal symbol $q(\mathbf{x}, \mathbf{k})$ of order -1 ,

$$Qu(\mathbf{x}) = \int q(\mathbf{x}, \mathbf{k}) \hat{u}(\mathbf{k}) e^{i\mathbf{k} \cdot \mathbf{x}} d\mathbf{k}, \quad (3.7)$$

such that

$$QF^T Fb \approx b. \quad (3.8)$$

That the order of Q should be -1 is a consequence of one of the rules for combining Ψ DOs: the product of two Ψ DOs is another Ψ DO, whose order is the sum of the orders of the factors [Taylor, 1981]. Since the identity operator is a Ψ DO of order 0 (principal symbol $\equiv 1!$), and the 2D version of $F^T F$ has order 1, the approximate inverse Ψ DO should have order -1 .

Whatever approach one might use to solve the pseudodifferential scaling problem,

it is necessary to compute the actions of the operators Q and $F^T F$, if only to gauge success. We will use a finite difference time-stepping method to approximate both Born modeling (F) and RTM (F^T). To approximate the action of Q , we use the algorithm of [Bao and Symes, 1996]. This algorithm is based on the observation that since the principal symbol $q_{-1}(\mathbf{x}, \mathbf{k})$ of Q is homogeneous of order -1 in $\mathbf{k} = k\hat{\mathbf{k}} = (k \cos \theta, k \sin \theta)$, $q_{-1}(\mathbf{x}, \mathbf{k}) = k^{-1}q_{-1}(\mathbf{x}, \hat{\mathbf{k}})$ is a 2π -periodic smooth function of θ , hence approximated using its first $N + 1$ Fourier components with exponentially small error in N :

$$q_{-1}(\mathbf{x}, \hat{\mathbf{k}}) \approx \frac{1}{2}c_0(\mathbf{x}) + \sum_{n=1}^N c_n(\mathbf{x})C_n(\hat{\mathbf{k}}) + s_n(\mathbf{x})S_n(\hat{\mathbf{k}}) \quad (3.9)$$

The spatial functions $\{c_n\}$ and $\{s_n\}$ are Fourier coefficients, and $\{C_n\}$ and $\{S_n\}$ are polynomials in the unit vector $\hat{\mathbf{k}}$ defined by the conditions:

$$\begin{aligned} C_n(\hat{\mathbf{k}}) &= \cos n\theta \\ S_n(\hat{\mathbf{k}}) &= \sin n\theta \end{aligned} \quad (3.10)$$

Recursion rules derived from the sum-of-angle formulas are the easiest way to compute $\{C_n\}$ and $\{S_n\}$:

$$\begin{aligned} C_{n+1}(\hat{\mathbf{k}}) &= 2\hat{k}_1 C_n(\hat{\mathbf{k}}) - C_{n-1}(\hat{\mathbf{k}}); C_0 = 1, C_1(\hat{\mathbf{k}}) = \hat{k}_1 \\ S_{n+1}(\hat{\mathbf{k}}) &= 2\hat{k}_2 C_n(\hat{\mathbf{k}}) + S_{n-1}(\hat{\mathbf{k}}); S_0 = 0, S_1(\hat{\mathbf{k}}) = \hat{k}_2 \end{aligned} \quad (3.11)$$

The reader will note that the first of these rules defines the Chebyshev polynomials in \hat{k}_1 .

It remains to discretize the equation 3.7 in \mathbf{x} . [Bao and Symes, 1996] use cubic spline approximation on a coarse nodal grid, taking advantage of the smoothness of $q(\mathbf{x}, \mathbf{k})$ and the implied smoothness of $\{c_n(\mathbf{x})\}$ and $\{s_n(\mathbf{x})\}$. In terms of a B-spline

basis $\{\psi_j(\mathbf{x}) : 1 \leq j \leq J\}$,

$$\begin{aligned} c_n(\mathbf{x}) &= \sum_{j=1}^J c_n^j \psi_j(\mathbf{x}) \\ s_n(\mathbf{x}) &= \sum_{j=1}^J s_n^j \psi_j(\mathbf{x}) \end{aligned} \quad (3.12)$$

Combine equations 3.9 through 3.12 with the definition 3.7 and rearrange to obtain the approximation of the action of Q in terms of the coefficient sequence $\mathbf{c} = \{c_n^j, s_n^j : 0 \leq n \leq N, 1 \leq j \leq J\}$:

$$Q[\mathbf{c}]u(\mathbf{x}) =$$

$$\sum_{j=1}^J \left(\sum_{n=0}^N \left[c_n^j \int k^{-1} C_n(\hat{\mathbf{k}}) \hat{u}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} d\mathbf{k} + s_n^j \int k^{-1} S_n(\hat{\mathbf{k}}) \hat{u}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} d\mathbf{k} \right] \right) \psi_j(\mathbf{x}). \quad (3.13)$$

Each integral inside the square bracket is actually a filter hence can be computed with the aid of the FFT at low cost. There are only $O(N)$ such filters to compute, for a much lower total cost per application than a straightforward gridded quadrature method applied to the definition 3.7. This reduction in cost is possible because the approximation 3.9 has replaced a joint function of \mathbf{x} and \mathbf{k} with a sum of products of \mathbf{x} and \mathbf{k} separately, and each of these can be computed using coarse grid computations by virtue of the smoothness of q . See [Bao and Symes, 1996, Nammour and Symes, 2009] for a careful discussion of the computational cost of this algorithm.

If the goal were to calculate the action of Q given the principal symbol q , then the coefficient array \mathbf{c} would have to be extracted from q by FFT and solution of a large linear system. Instead, [Nammour, 2009] simply reformulates the optimal pseudodifferential scaling problem to extract \mathbf{c} directly from its data, via solution of

an optimization problem: given an image b , for example $b = F^T d$, find coefficients \mathbf{c} to minimize

$$\|b - Q[\mathbf{c}](F^T F)b\|^2, \quad (3.14)$$

Problem 3.14 is a quadratic least squares problem, and could be solved using an appropriate numerical optimization method, for example conjugate gradient method. Note that the number of unknowns here is $(2N + 1)J$, which could be much smaller than the unknowns of extended LSM: based on our experience, $N \approx O(10)$ is adequate, and J can be taken to be 2-3 orders of magnitude smaller than the number of grid points of the model b (or m). Given the work estimate mentioned above for each application of Q , the optimal pseudodifferential scaling problem is cheap to solve, compared with the iterative extended LSM.

[Nammour, 2009] shows that the solution of problem 3.14 satisfies

$$Q[\mathbf{c}]b \approx (F^T F)^\dagger b = (F^T F)^\dagger F^T d \approx \delta m, \quad (3.15)$$

for $b = F^T d$ with $d = F\delta m$. $(F^T F)^\dagger$ denotes the Moore-Penrose pseudoinverse [Moore, 1920].

To summarize, to compute the optimal pseudodifferential scaling,

- Compute extended migration $b = F^T d$;
- Apply normal operator to get re-migrated image $(F^T F)b$;
- Solve problem 3.14 to get \mathbf{c} ;
- Compute $Q[\mathbf{c}]b \approx (F^T F)^{-1}b$.

We end this discussion by mentioning the restrictions that must hold in order that the underlying assumption of this construction, namely that $F^T F$ be a Ψ DO, is valid. Smoothness of the velocity model m is essential: presence of interfaces

typically implies the well-known tendency for reflectors to be imaged in multiple locations, whereas Ψ DOs preserve reflector locations. For the physical (non-extended) Born operator, the necessary additional conditions are rather mild: essentially, no rays can arrive at the receiver or source arrays tangent to the surfaces on which they lie, and scattering over π (diving waves) must also be ruled out (or removed from the output of F) [Stolk, 2000]. For the shot record extension, an additional and rather strict assumption must be satisfied, namely that the ray field associated to m have no caustics in the region where $\delta m \neq 0$. This condition effectively restricts preconditioning by pseudodifferential scaling to models with mild lateral heterogeneity. [Nolan and Symes, 1996, Stolk and Symes, 2004] thoroughly discuss the reasons for this restriction, which is related to failure of shot-record extended prestack migration to act as an imaging operator in the presence of caustics or multipathing.

3.3.2 Flexibly Preconditioned Conjugate Gradient Algorithm

Preconditioned conjugate gradient (CG) iteration [Nocedal and Wright, 1999, Golub and van Loan, 1996] is simply the CG algorithm formulated with an inner product different from the usual Euclidean dot product. Applied to least-squares problems such as equation 3.5, the effect is to alter the application of the adjoint operator (F^T) by appending another operator M . If $MF^T F \approx I$, convergence is much accelerated.

It is tempting to use the approximate inverse Q , computed by solving equation 3.14 as in the previous section, for M in preconditioned CG, however that does not result in accelerated convergence. The reason is that while Q satisfies equation 3.8 approximately, it is not generally an approximation to $(F^T F)^\dagger$ - it only approximates the action of the latter operator on a particular vector.

There is however a variant of preconditioned CG, called *flexibly preconditioned CG* (FPCG), that makes use of this type of approximate solution of the normal equation

as opposed to an approximate inverse of the normal operator. The algorithm is listed explicitly in **algorithm 1**. The action of a conventional preconditioner is replaced by an approximate solution of the normal equation at each step - thus each iteration requires that Q , hence \mathbf{c} , be constructed anew, and an additional application of Hessian per iteration is required. In the algorithm listing, the construction of the iteration-dependent preconditioning step is highlighted in red.

Algorithm 1 A flexibly preconditioned conjugate gradient method

Given data d and $\delta m_0 = 0$, Born modeling operator F , RTM operator F^T

Given maximum iterations `MaxIter` and $i = 0$

$r_0 = d$, $G_0 = F^T r_0$

compute \mathbf{c}_0 by solving equation 3.14 with $b = G_0$

$g_0 = Q[\mathbf{c}_0]G_0$

$p_0 = g_0$

while (not converge && $i < \text{MaxIter}$) **do**

$q_i = Fp_i$

$\alpha_i = \frac{\langle g_i, G_i \rangle}{\langle q_i, q_i \rangle}$

$\delta m_{i+1} = \delta m_i + \alpha_i p_i$

$r_{i+1} = r_i - \alpha_i q_i$

$G_{i+1} = F^T r_{i+1}$

compute \mathbf{c}_{i+1} by solving equation 3.14 with $b = G_{i+1}$

$g_{i+1} = Q[\mathbf{c}_{i+1}]G_{i+1}$

$\beta = \frac{\langle g_{i+1}, G_{i+1} - G_i \rangle}{\langle g_i, G_i \rangle}$

$p_{i+1} = g_{i+1} + \beta p_i$

$i = i + 1$

end while

One thing to note is the formula for the direction update scalar β in algorithm 1 should be the Polak-Ribiere variant [Nocedal and Wright, 1999], as recommended by [Notay, 2000], who proved that as long as the variation in the preconditioner is sufficiently small, the FPCG method will have a convergence rate that is comparable with CG method with an approximate inverse of $F^T F$ as the iteration-independent preconditioner. While the conditions for convergence developed in [Notay, 2000, Knyazev and Lashuk, 2006] do not appear to be verifiable in examples, the algorithm

nonetheless appears to perform reasonably well, as the examples to come will show.

We note that the behavior of the algorithm may be improved by modifying the definition of F so that $F^T F$ is better approximated by a Ψ DO. For example, appropriate mutes and window functions to remove diving wave energy and shallow acquisition footprint will improve the approximate solution by pseudodifferential scaling. High pass filtering may also make this approximation more accurate, since we keep only the leading order terms in frequency throughout our constructions.

3.4 Numerical examples

In this section, we show several numerical results for the FPCG method and comparisons with conjugate gradient and steepest descent methods.

Wave equations are discretized using regular grid finite-difference scheme: 2nd order accurate in time and 4th order accurate in space. We use free surface top boundary condition and place sources and receivers below the surface. The choice of free surface boundary condition and this acquisition geometry introduce two derivatives to the data: vertical derivative over the source position and vertical derivative over the receiver position [Hou and Symes, 2015b], which are approximately equivalent to $\frac{\partial^2}{\partial t^2}$. Thus composing the original Born modeling operator with the square of integral operator over time I_t^2 roughly undo the effect of using this free surface boundary condition and acquisition geometry, where $I_t = \int_t dt$.

The FPCG method used in following examples contains two steps of preconditioning: first, apply I_t^2 to both the input data and the source wavelet; second, apply Algorithm 1 to the integrated data and the integrated source wavelet.

The first two examples concern recovering a single reflector under the presence of a slow and a fast Gaussian anomaly in background model respectively. These two simple synthetic examples are used to demonstrate the behavior of the proposed

algorithm. The last example is the Marmousi model with 450 m extended water layer, which is used to illustrate the capability and highlight the superiority of the proposed method when a model has complex structures.

For each example, we show the convergence curves of solving extended LSM using FPCG method and other methods, display the inverted extended image $\delta m(\mathbf{x}, \mathbf{x}_s)$. We also show the stacked image $\sum_{\mathbf{x}_s} \delta m(\mathbf{x}, \mathbf{x}_s)$, the target data, and the data residual (the difference of the predicted and target data).

3.4.1 Slow Gaussian anomaly model

The slow Gaussian anomaly model is illustrated in Figures 3.1a and 3.1b. The acquisition geometry is as described in the last section. Source and 3 shots of the data are shown in Figures 3.3a and 3.3b.

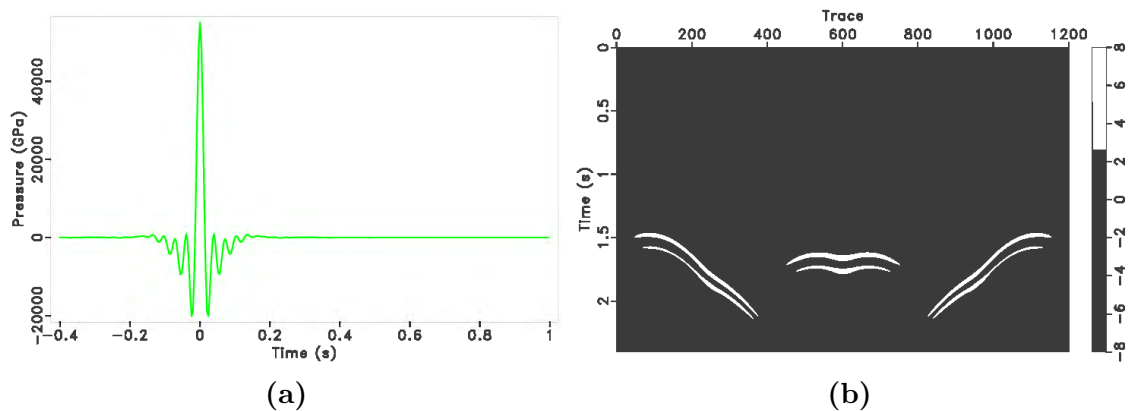


Figure 3.3: (a) source wavelet with corner frequency 1 Hz, 7 Hz, 28 Hz, 35 Hz, used in Gaussian anomaly examples (b) 3 of 81 Born shot records for slow Gaussian anomaly example (Figures 3.1a and 3.1b.)

Data shown in Figure 3.3b is used as the observed data in this example (81 shots - only 3 are shown here). Figure 3.4a shows the normalized data misfit (the value of the objective function in equation 3.4) as a function of number of Hessian applications. Figure 3.4b plots the normalized normal residual (the L2 norm of the difference of the right hand side and the left hand side with or without preconditioning of equation

3.5). CG with windowing is the conjugate gradient method with the forward map F replaced by FW , the composition of F and a windowing operator W which multiplies a smooth cutoff function that sets $\delta m(\mathbf{x}, \mathbf{x}_s)$ to zero from the surface to depth = 200 m: this is the depth range in which imaging is dominated by acquisition footprint and large amplitude, making the Hessian more ill-conditioned. The improvement of conditioning resulting from windowing out these effects is evident in the convergence plots, but is not sufficient to bring the residual below 40% of its initial value in 20 iterations. Unwindowed CG converges even more slowly.

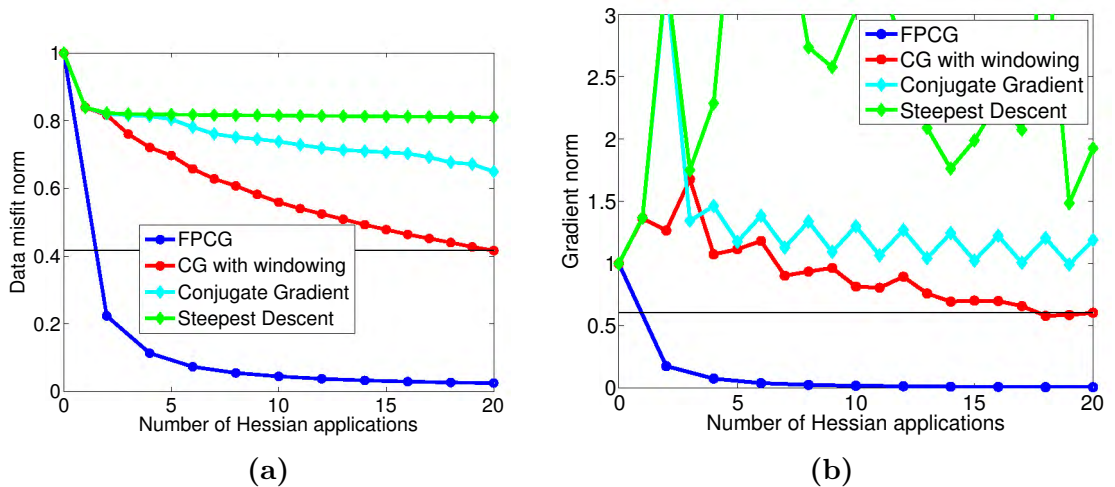


Figure 3.4: Slow Gaussian anomaly example: convergence curves of numerical methods, (a) normalized data misfit and (b) normalized gradient length.

In contrast, FPCG (Algorithm 1 applied to integrated data and the integrated source wavelet, without windowing or muting) achieves 4% of the initial residual RMS in 10 iterations. Note that each iteration of FPCG requires a second application of the Hessian, hence is twice as expensive as each iteration of CG. The convergence plots are arranged to take this expense differential into account: the horizontal axis unit is Hessian applications. In particular, note that FPCG with 2 Hessian applications fits the data better than CG with 20 Hessian applications (and depth windowing as described above, Figure 3.4a). In this example, FPCG is the only method to

approximate a stationary point of the least squares objective, reducing the gradient length to a small fraction (0.8%) of its original value in 20 iterations (Figure 3.4b).

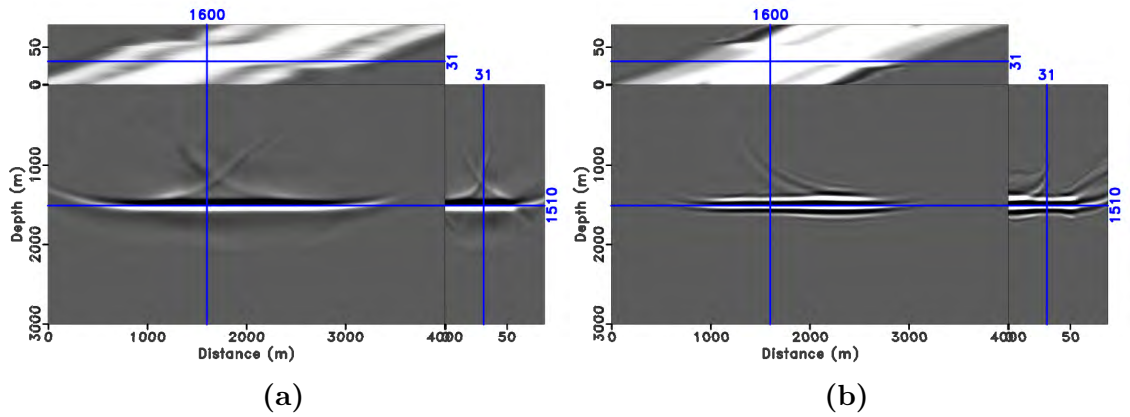


Figure 3.5: Slow Gaussian anomaly example: inverted model perturbation cube after 20 Hessian applications using FPCG (a) and using CG with windowing (b).

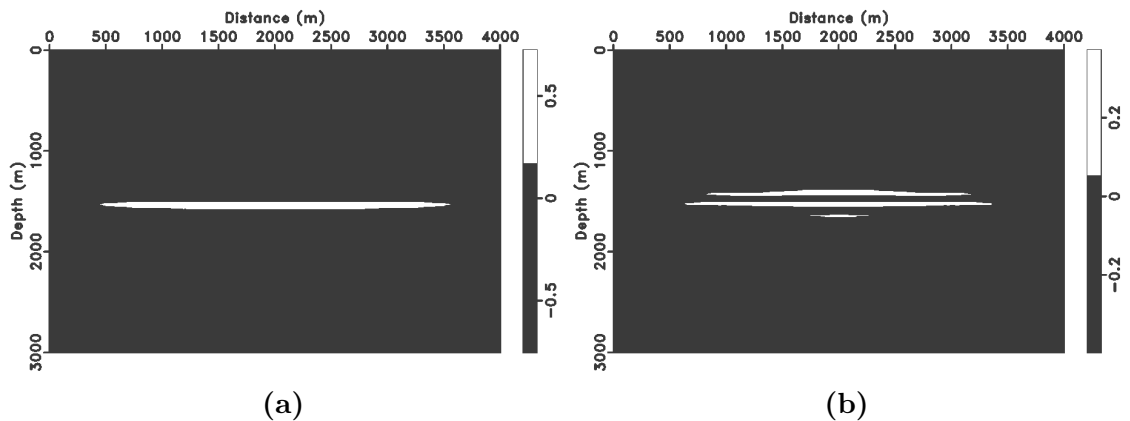


Figure 3.6: Slow Gaussian anomaly example: Stacked image after 20 Hessian applications using FPCG (a) and using CG with windowing (b).

Figures 3.5a and 3.5b show the inverted extended images or extended model perturbations, the solutions of equation 3.5 using FPCG and CG method with windowing. They are 2D views of 3D volumes. The physical model axes are labeled as "Distance" and "Depth". The third axis without label is the shot record. Figures 3.6a, 3.6b, 3.7a and 3.7b compare the stacked images and data residuals of FPCG method and CG method with windowing. Figures 3.5a and 3.5b and stacked images

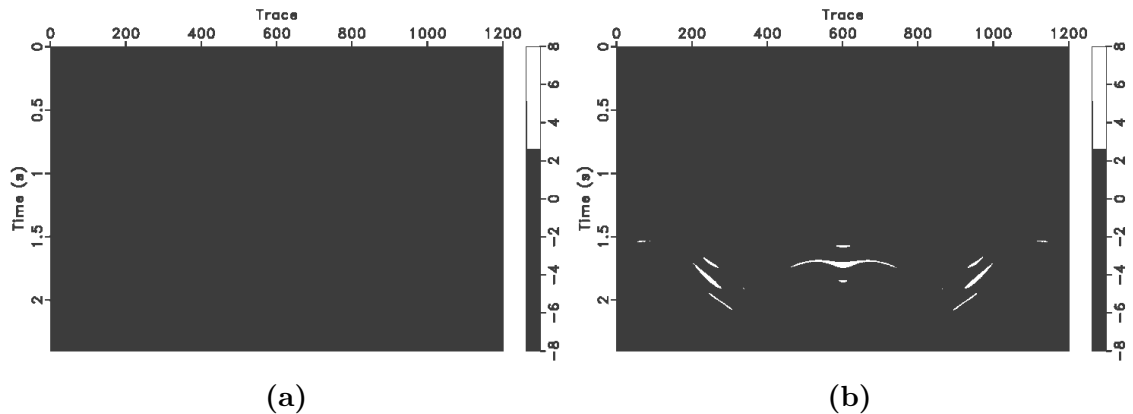


Figure 3.7: Slow Gaussian anomaly example: data residual, same 3 shots as in Figure 3.3b after 20 Hessian applications using FPCG (a), and using CG with windowing (b).

in Figures 3.6a and 3.6b show that the frequency contents in the final images are quite different with and without preconditioning. Fewer artifacts are observed in the image from FPCG method (Figures 3.5a and 3.6a). It is obvious from Figures 3.7a and 3.7b that preconditioning improves the data fitting significantly as well as helps to attenuate unwanted noise.

3.4.2 Fast Gaussian anomaly model

The fast Gaussian anomaly model is shown in Figure 3.8a. The acquisition geometry, source wavelet, and the true reflectivity model are the same as the slow Gaussian anomaly model example. 3 shots of the data are shown in Figures 3.8b. Very similar convergence behavior (Figures 3.9a and 3.9b) as in the slow Gaussian anomaly model example is observed in this set of tests. See Figures 3.10, 3.11 and 3.11 for the comparison of inverted images, stacked images and data residuals by FPCG method and CG method.

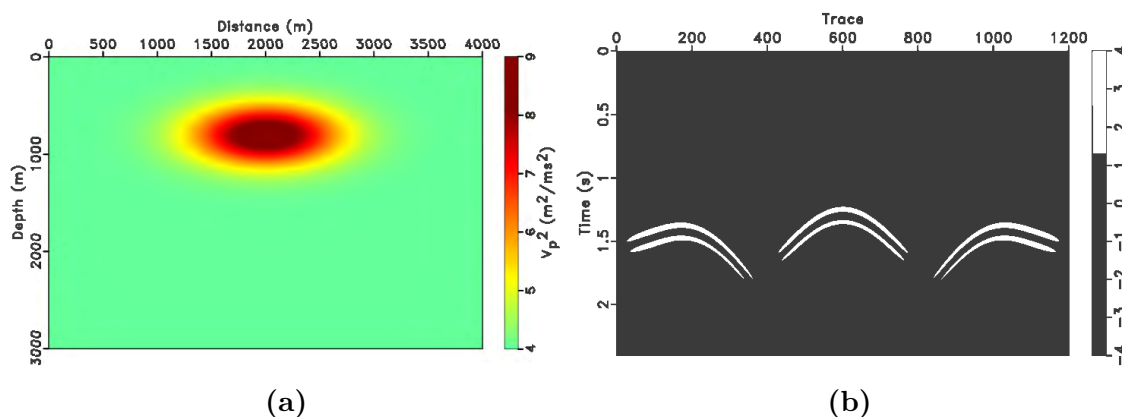


Figure 3.8: Fast Gaussian anomaly example: (a) background model, and (b) 3 of 81 Born shot records.

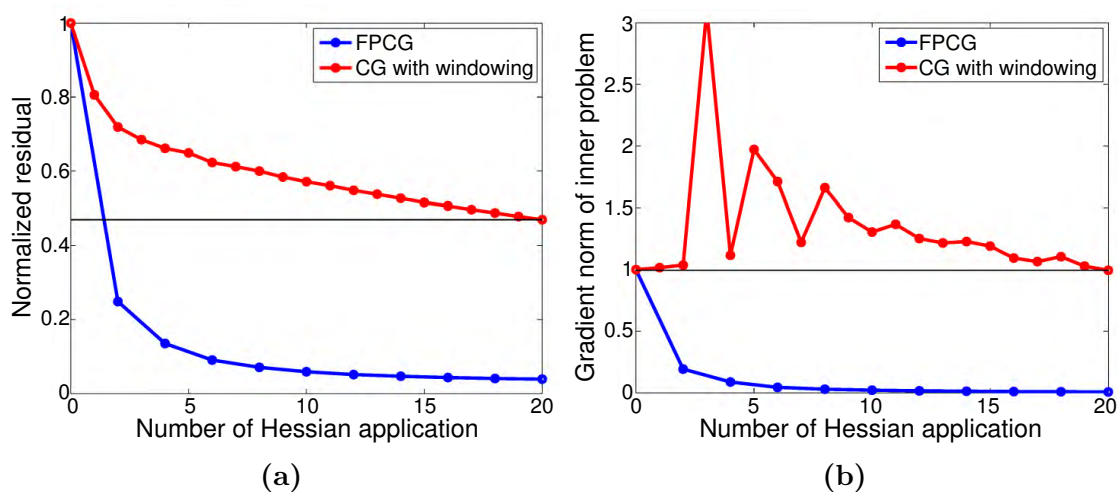


Figure 3.9: Fast Gaussian anomaly example: convergence curves of numerical methods for (a) normalized data misfit and (b) normalized gradient length.

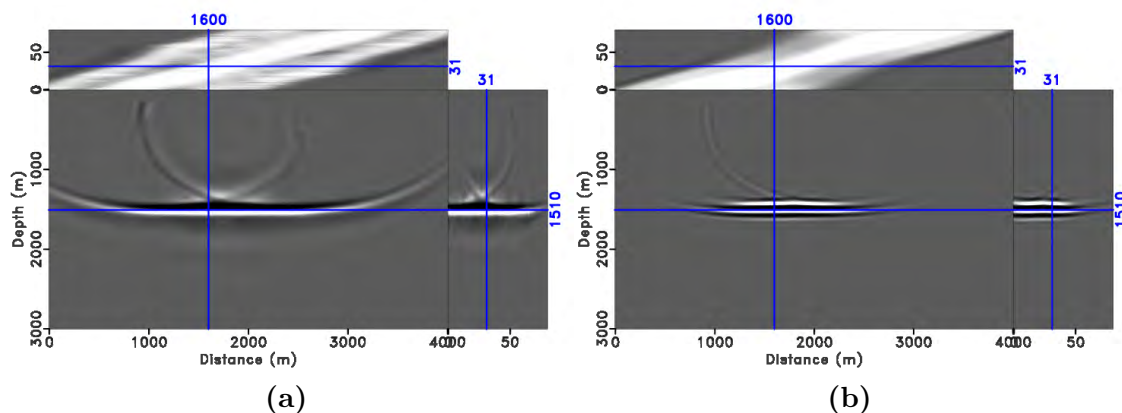


Figure 3.10: Fast Gaussian anomaly example: inverted model perturbation cube after 20 Hessian applications using FPCG (a) and using CG with windowing (b).

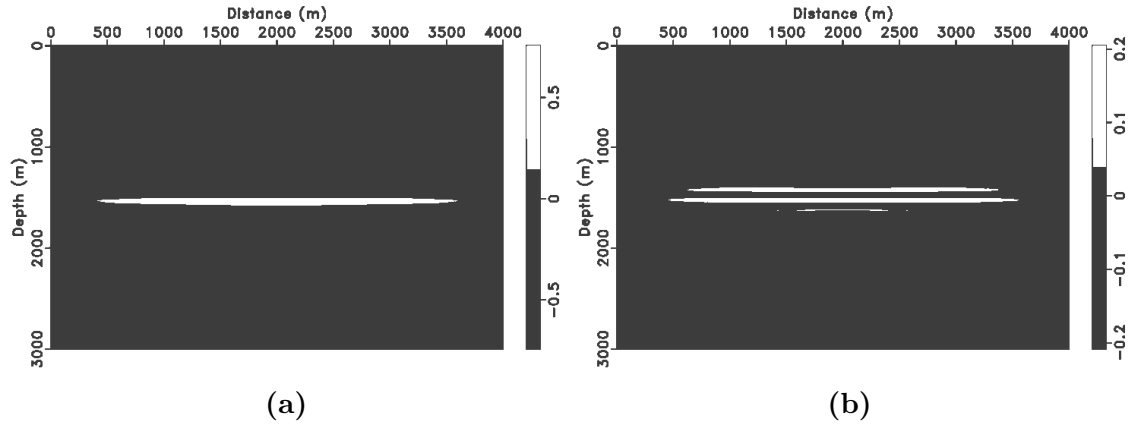


Figure 3.11: Fast Gaussian anomaly example: Stacked image after 20 Hessian applications using FPCG (a) and using CG with windowing (b).

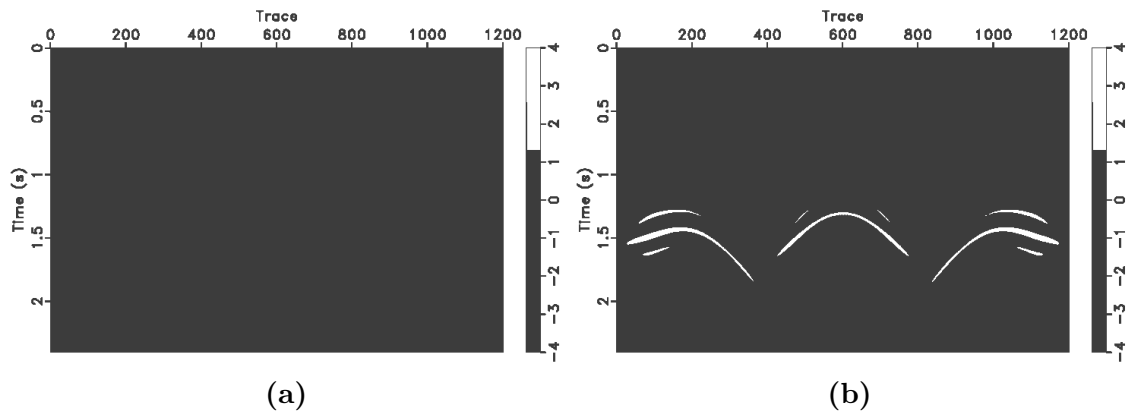


Figure 3.12: Fast Gaussian anomaly example: Data residual after 20 Hessian applications using FPCG (a) and using CG with windowing (b).

3.4.3 Marmousi model

To evaluate the potential of the FPCG algorithm to recover reflectivity with a wide range of dips and strengths, we apply it to a model derived from the Marmousi model [Versteeg and Grau, 1991]. We smoothed the model and added a 450 m deep water layer to produce the smooth background model shown in Figure 3.13a. The difference between the original and a slightly less smoothed model serves as the model perturbation or reflectivity field (Figure 3.13b). The model was gridded at 8 m in both directions, and a finite difference method used to produce Born data (Figures 3.14a, 3.14b). The source wavelet was a zero-phase bandpass filter with corner frequencies 1 Hz, 6 Hz, 25 Hz, 30 Hz. 128 point sources were located at depth 8 m with 48 m spacing. 1151 fixed receivers with 8 m spacing were placed at depth 6 m. Final recording time was 3.4 s.

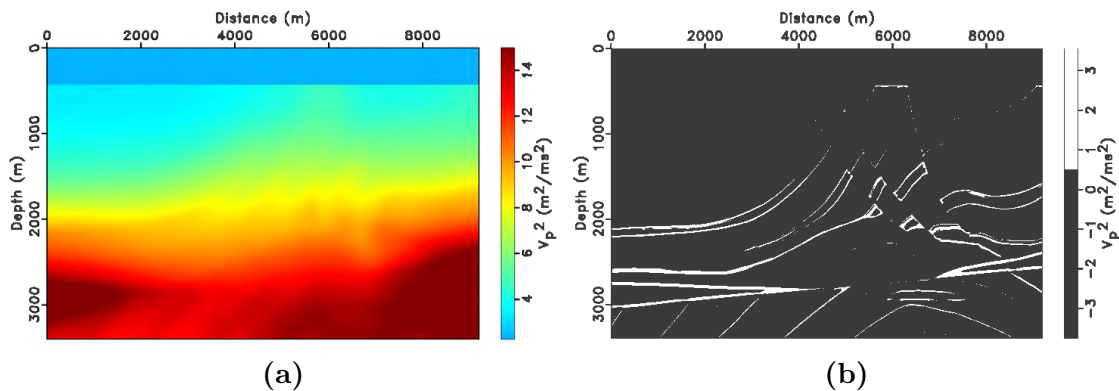


Figure 3.13: Marmousi example: (a) background model, (b) model perturbation.

Although we use Born modeling to generate the observed data, diving waves kinematics are still present since diving waves generated by the background model are also perturbed. We have muted the diving waves, both in the target data and by composing the modeling operator with the same mute.

As in the previous examples, we used both CG with windowing and FPCG to invert this data. The windowing operator in this case zeroes the first 450 m in depth,

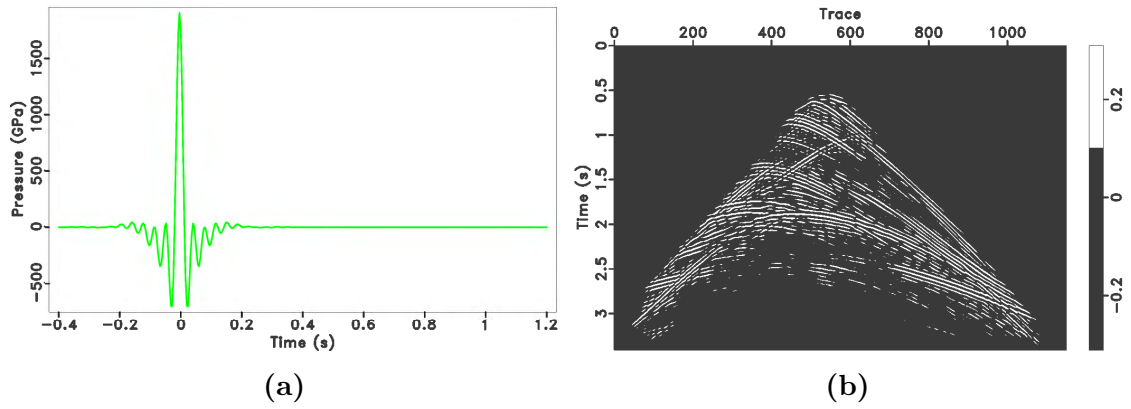


Figure 3.14: Marmousi example: (a) source wavelet with corner frequencies = 1 Hz, 6 Hz, 25 Hz, and 30 Hz. (b) Born shot record, shot index 65.

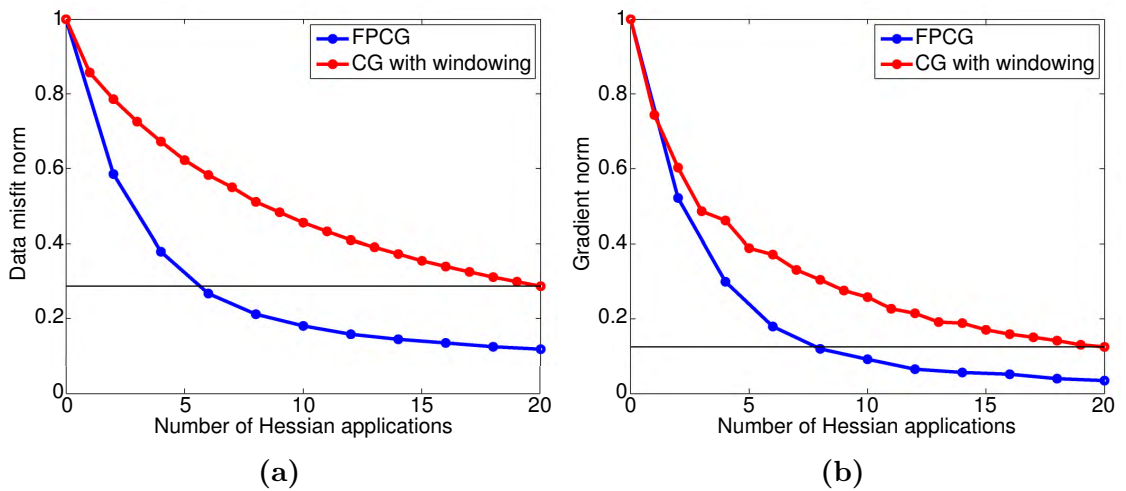


Figure 3.15: Marmousi example: convergence curves of numerical methods, (a) normalized data misfit and (b) normalized gradient length.

with a smooth taper to avoid introducing high-frequency artifacts into the inversion. Figure 3.15a shows the normalized data misfit as a function of Hessian applications. Figure 3.15b plots the normalized normal residual or the gradient of the CG and FPCG method. The FPCG method with 6 Hessian applications fits the data better than the CG method with 20 Hessian applications (Figure 3.15a), similarly for the gradient length.

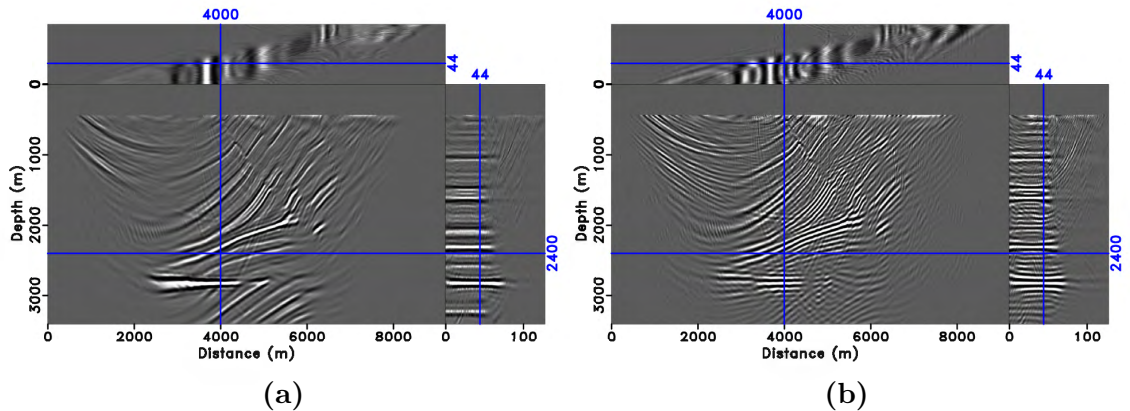


Figure 3.16: Marmousi example: inverted model perturbation cube after 20 Hessian applications using FPCG (a) and using CG with windowing (b).

Figures 3.16a and 3.16b show the inverted extended model perturbations produced by these methods. The FPCG result (Figure 3.16a) shows both more accurate amplitudes and better focused wavelet than does the non-preconditioned CG result (Figure 3.16b). The superiority of the FPCG method is especially evident in the gather panel: the FPCG-produced events are better separated and flatter, with fewer artifacts contaminating the image volume, than are the CG-produced events.

Figures 3.17a and 3.17b compare the stacked version of the inverted extended images of FPCG method and CG method with windowing. We see that preconditioning is effective in mitigating artifacts and producing a better balanced amplitude in the image, as well as improving the resolution of the image especially in the deep part (below 2.5km).

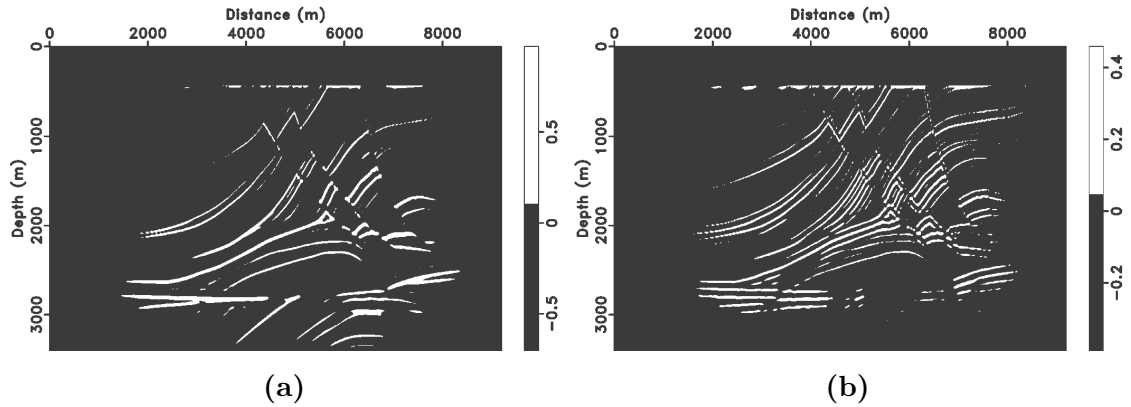


Figure 3.17: Marmousi example: Stacked image after 20 Hessian applications using FPCG (a) and using CG with windowing (b).

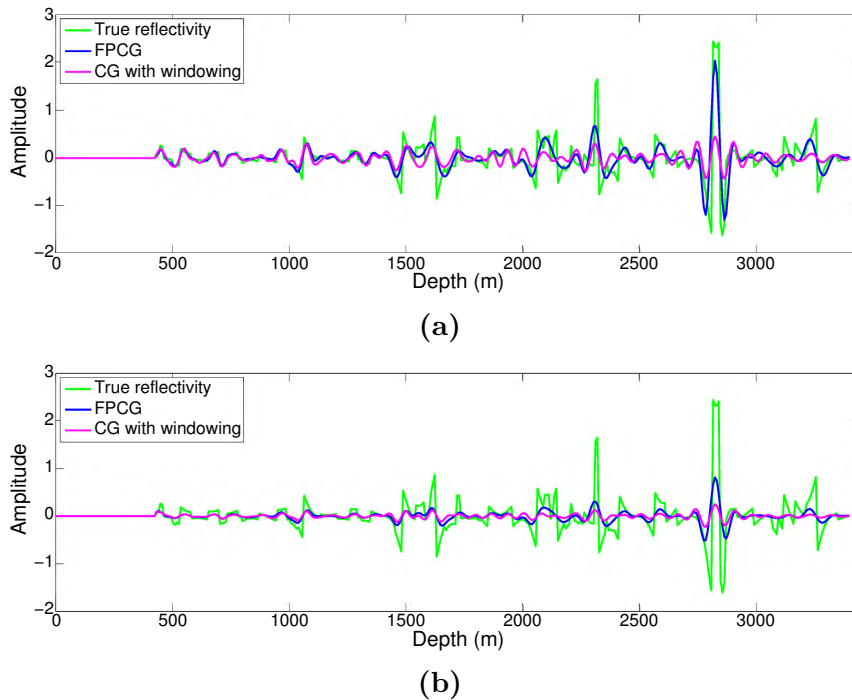


Figure 3.18: Marmousi example: Vertical trace comparison at $x = 4000$ m of target model perturbation with inverted model perturbation cube from the 44-th shot (a) and with stacked inverted image (b) by FPCG and CG with windowing methods after 20 Hessian applications.

Figure 3.18a compares the well logs at $x = 4000$ m of inverted model perturbations from the 44-th shot data by FPCG and CG with windowing methods with the true reflectivity model. Both FPCG and CG with windowing recover the amplitude of the shallow part of the model, while FPCG method restores relatively better the amplitude of the deep part of the model than the non-preconditioned method. The spectra of traces of inverted images corresponding to shot 44 are shown in Figure 3.19a. It is clear that for the wavenumber range that could be explained by these methods (between 0.0025 $1/\text{m}$ and 0.013 $1/\text{m}$), the amplitude of the inverted image by FPCG method resembles that of the target image.

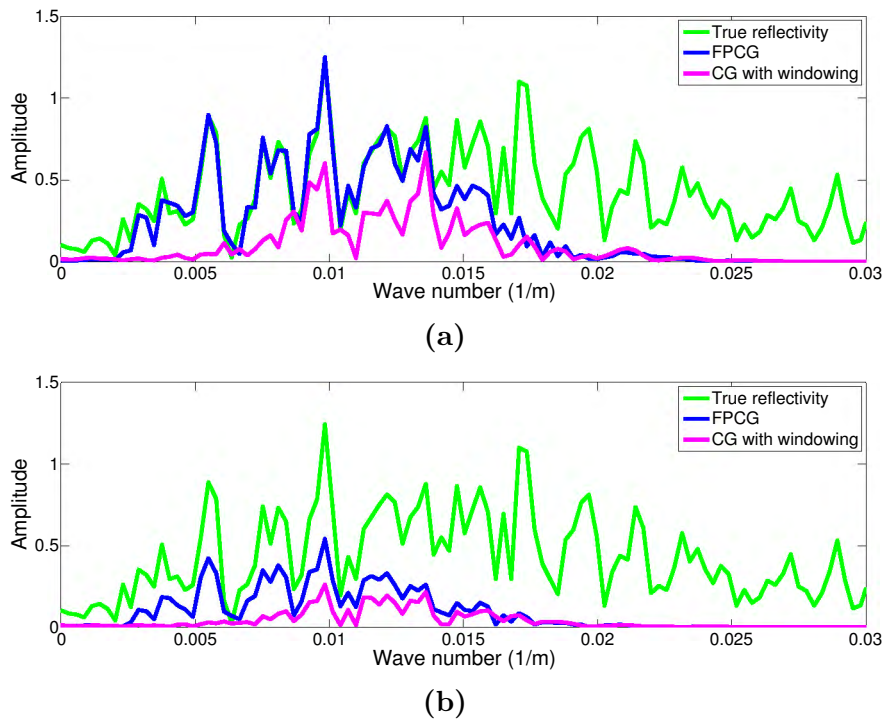


Figure 3.19: Marmousi example: Spectrum comparison of the vertical trace shown in Figure 3.18 of target model perturbation with inverted model perturbation cube from the 44-th shot (a) and with stacked inverted image (b) by FPCG and CG with windowing methods after 20 Hessian applications.

Since the LSM is in the shot record domain, the inverted image volume minimizes the data residual, while the stack of the inverted image does not, although the stacked

image contains a lot of useful informations as well. That is why after stacking, the amplitude does not match the true reflectivity as shown in Figure 3.18b, but the amplitude from FPCG method is still relatively closer to the true one than that from CG with windowing at the deep part of the model (below 2500 m). The corresponding spectra comparison of traces of stacked images are shown in Figure 3.19b. Note that the traces and spectra of stacked images are for visualization purposes and have no quantitative meaning.

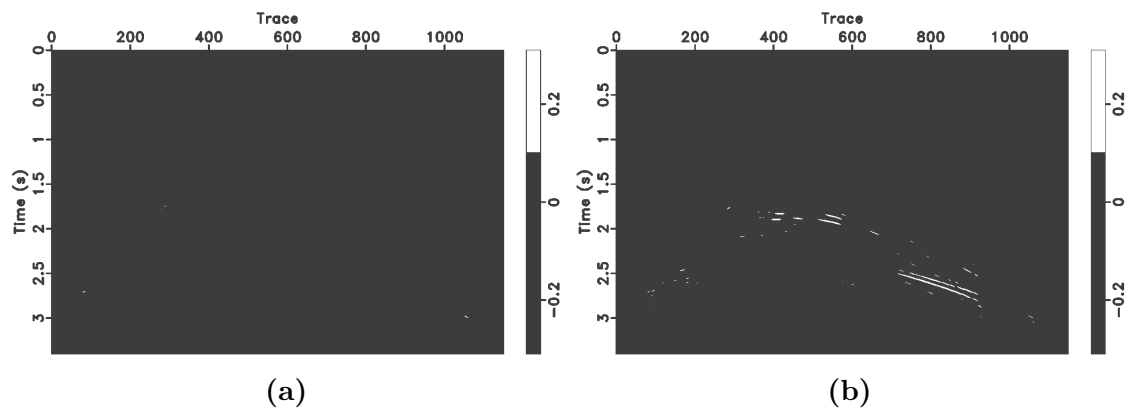


Figure 3.20: Marmousi example: data residual, same shot record as in Figure 3.14b, after 20 Hessian applications using FPCG (a), and using CG with windowing (b).

Data residuals from FPCG method and CG method with windowing are shown in Figures 3.20a and 3.20b respectively. With preconditioning, data residual is small everywhere, while the non-preconditioned CG is struggling to fit the deep part (around and below 2s) of the data. We also see steeply-dipping events in the residual from FPCG method, which is better modeled by the non-preconditioned CG method. These events are close to diving wave perturbations, and are poorly explained by the pseudodifferential scaling operator.

Since these straight-line events are mainly due to refracted waves, with a particular apparent velocity, they can be suppressed through dip filtering. Figures 3.21a and 3.21b show the same data residual, but after application of a dip filter with cut-off velocity 2500 m/s, so that events with slopes larger than 0.0004 s/m are excluded.

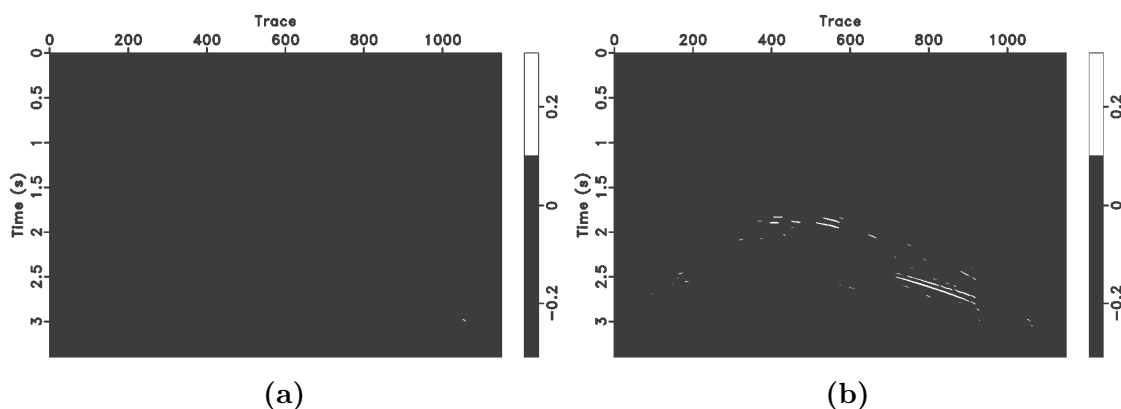


Figure 3.21: Marmousi example: data residual after dip filtering with 20 Hessian applications using FPCG (a), and using CG with windowing (b).

Figures 3.21a and 3.21b show the data residual after application of the dip filter.

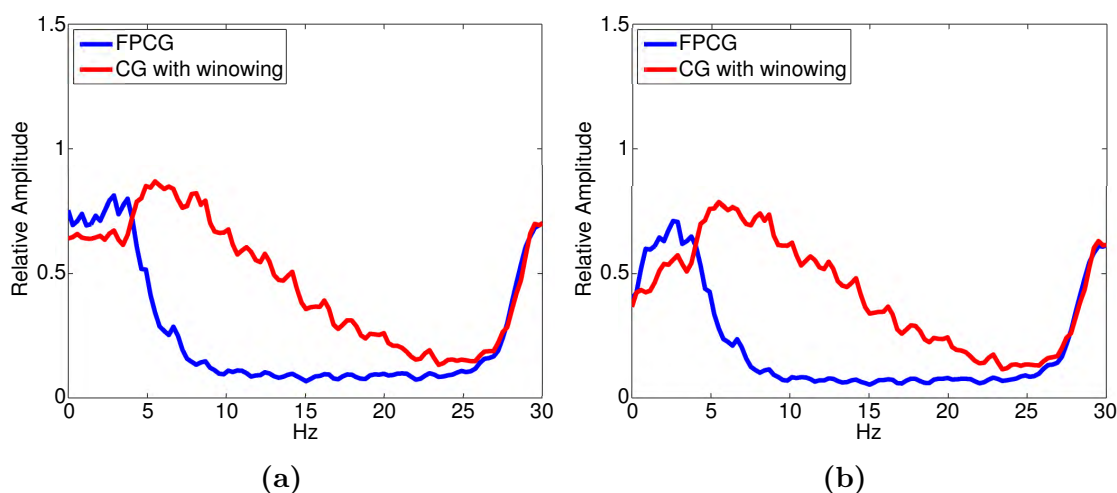


Figure 3.22: Marmousi example: Relative power spectra of data residual (a) and data residual after dip filtering (b) with respect to observed data.

The relative data residual power spectrum gives another way of appreciating the effect of preconditioning. FPCG method suppresses error regardless of frequency components, while CG method fits different frequency components in the data differently, as can be observed in Figure 3.22a. The error spectra with dip filter (Figure 3.22b) are slightly lower in amplitude than those without dip filter.

One advantage cited earlier for extended Born modeling is ability to fit data even

with incorrect velocity. The final example of this section illustrates this property. The smooth background model depicted in Figure 3.23 is obtained by adding 70% of the target model in Figure 3.13a and 30% of the constant model $m = 2.25 \text{ m}^2/\text{s}^2$. Note that the background model is scaled, thus the reflection coefficients at the ocean bottom is different from the target model and the refracted waves generated by the target model will not be predicted correctly. Perturbations of these refracted waves form a part of the data that could not be fit using shot record domain LSM, preconditioned or not, which is a feature of shot record domain model extension. However, for the part of data that could be fitted, using shot record domain model extension, preconditioning will still significantly improve the convergence rate. We show the convergence curves for inversion of the Marmousi-Born data (Figure 3.14b) with this incorrect velocity model in Figure 3.24a. As was the case with correct velocity, a reasonably good fit is obtained, with most of the remaining residual energy located near the diving wave perturbations, and the FPCG iteration is considerably more efficient than the CG iteration, by about the same factor as for the correct velocity case.

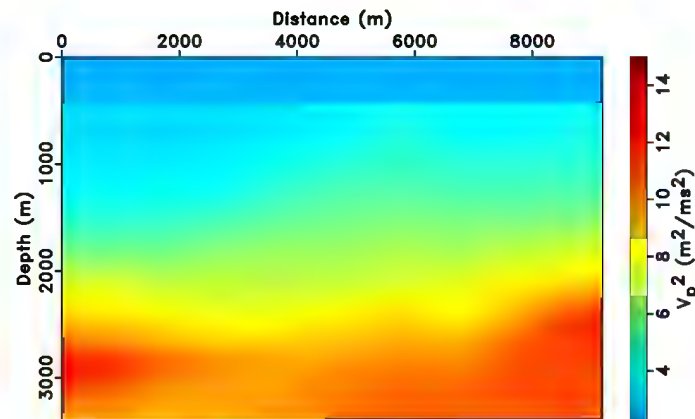


Figure 3.23: Marmousi example, incorrect background model.

Figures 3.25a and 3.25b show the inverted extended model perturbations. The gather information is more valuable than the single shot image in this example, since

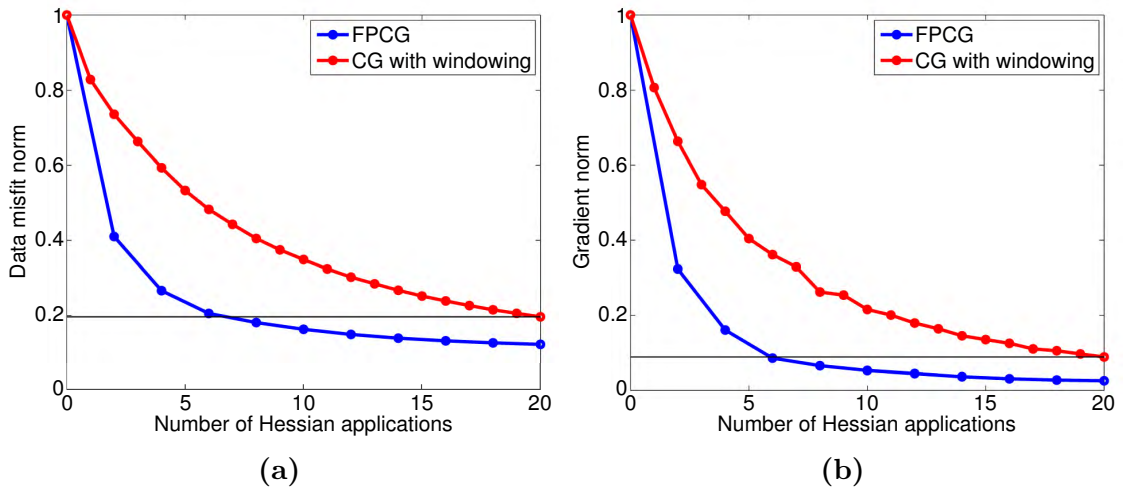


Figure 3.24: Marmousi example, incorrect background model: convergence curves of numerical methods, (a) normalized data misfit and (b) normalized gradient length.

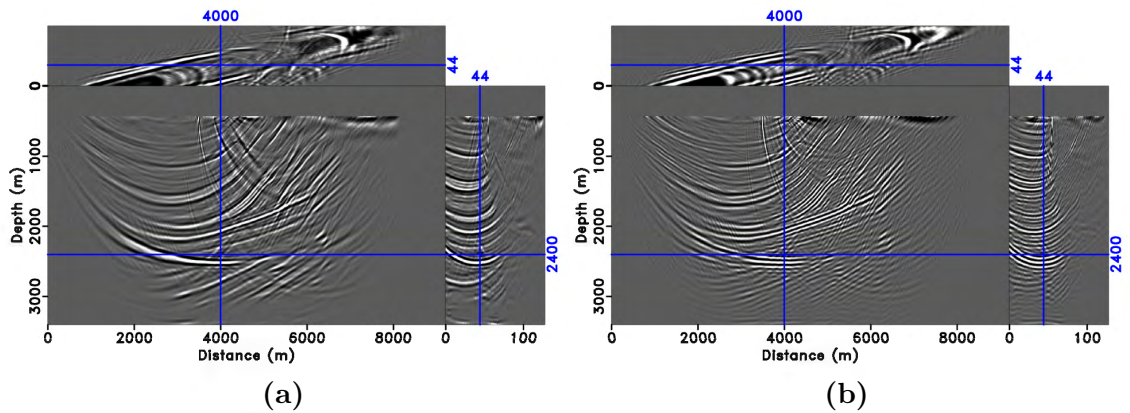


Figure 3.25: Marmousi example, incorrect background model: inverted model perturbation cube after 20 Hessian applications using FPCG (a) and using CG with windowing (b).

the background model is wrong. We see a better focused gather with preconditioning (Figure 3.25a) compared with non-preconditioning (Figure 3.25b). The curvatures in image indicate that the background model is smaller than the true model and thus could be used to update the background model.

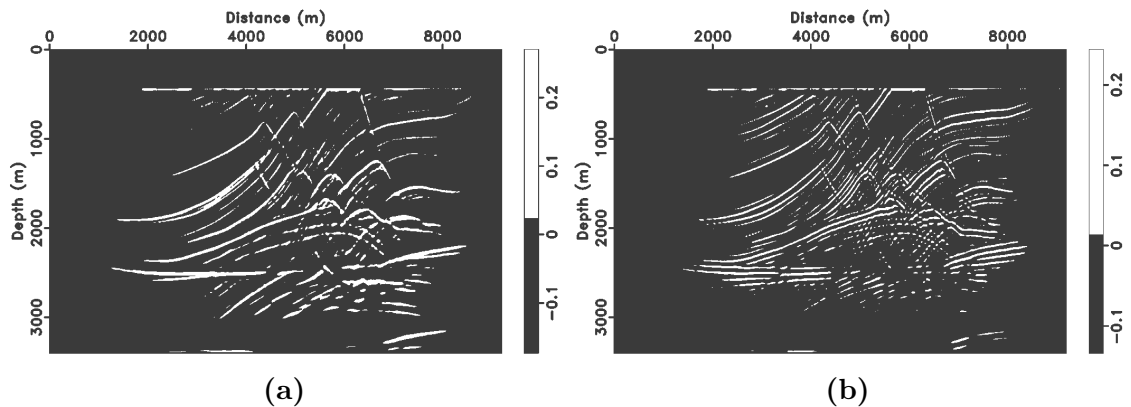


Figure 3.26: Marmousi example, incorrect background model: Stacked image after 20 Hessian applications using FPCG (a) and using CG with windowing (b).

Figures 3.26a and 3.26b compares the stacked images of FPCG method and CG method with windowing at wrong background model (as shown in Figure 3.23). We see that the preconditioning is still helpful in mitigating artifacts in the image, and produces better balanced amplitude as well as improves the resolution of the image, although the reflector positions are wrong.

Finally, the data residuals depicted in Figures 3.27a and 3.27b, using the same grey scales as in Figures 3.20a and 3.20b, confirm the ability of shot record extended Born modeling to fit data very well even with incorrect velocity. Once again, the FPCG residual is considerably smaller than the CG residual, even though the same computational effort has been expended in producing both.

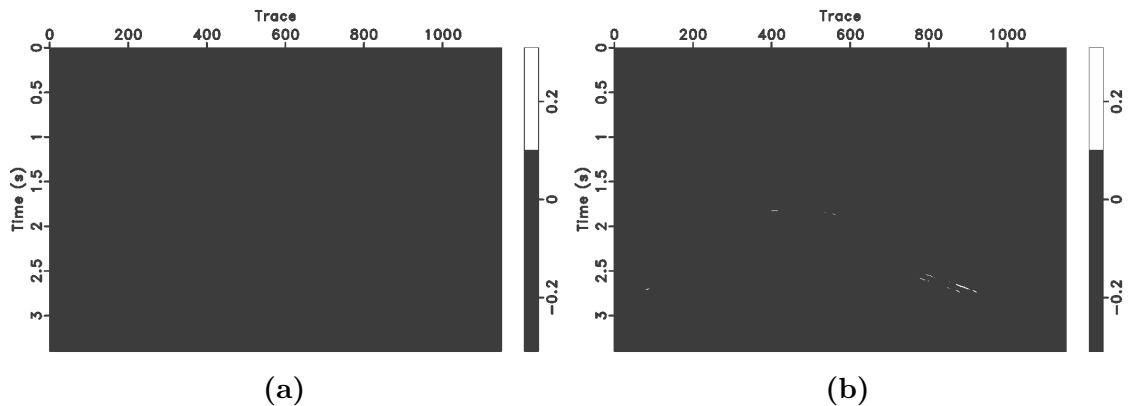


Figure 3.27: Marmousi example, incorrect background model: data residual, same shot record as in Figure 3.14b, after 20 Hessian applications using FPCG (a), and using CG with windowing (b).

3.5 Discussion

The acceleration technique presented here suffers from some strict limitations on its validity, noted in the section on pseudodifferential scaling: for example, it depends on more or less unique energy paths between sources, receivers, and scattering points. However, within the domain of validity - mild lateral velocity variation and slow changes in spatially averaged index of refraction - it provides an inexpensive and effective way to accelerate iterative LSM.

We have examined only the simplest relevant wave physics in this paper, namely constant density acoustics. However, the foundations have been laid long since for extension to variable density acoustics and indeed to elasticity [Beylkin and Burridge, 1990, Burridge et al., 1998, De Hoop and Bleistein, 1997]. This extension amounts to a rational version of amplitude versus offset (AVO) analysis, with illumination based amplitude artifacts have been removed [Tarantola, 1984a, Lörtzer and Berkhout, 1989, Castagna, 1993, Luh, 1993]. Accelerated convergence in these settings can in principle be achieved using similar ideas - for some first steps see [Nammour and Symes, 2011].

A fast extended LSM available could be used in a form of optimization based migration velocity analysis based on Born inversion [Liu et al., 2014, Symes, 2015,

[Huang and Symes, 2015](#)]. In fact, inversion (as opposed to migration), with its improved amplitude and especially phase, is actually essential for reliable functioning of such objective velocity analysis algorithms [[Kern and Symes, 1994](#), [Lameloise et al., 2015](#)].

Variable Projection Extended Born Waveform Inversion Enhanced by Preconditioned Least Squares Migration

4.1 Summary

Born waveform inversion is a linearized version of full waveform inversion based on Born modeling and model scale separation, in which both long scale background model and short scale reflectivity are updated to predict the observed seismic data. By allowing the reflectivity to depend on an extra parameter, we get the extended Born waveform inversion (EBWI). And a penalty term that controls the degree of dependence on this extra parameter is added to the EBWI objective function.

EBWI is quadratic on the extended reflectivity. Thus a nested optimization scheme called variable projection method is used: minimize the objective function over the extended reflectivity first in an inner stage (which is closely related to the regularized least squares migration), and then minimize over the background model in an outer stage, which we refer to as the VPE method. A flexibly preconditioned regularized least squares migration scheme solves the inner minimization effectively

and efficiently, and thus generated a better shaped VPE objective function and more accurate gradient than without preconditioning based on numerical examples and thus improve the performance of the VPE method. However, we explain that a theoretical estimation of the gradient computation is hard to get in its current setting and why this is to be expected.

4.2 Introduction

Seismic full waveform inversion (FWI) [Tarantola, 1984a] is a method to get subsurface structures of the earth from seismic data, by solving a data-fitting problem, where wave equations are solved to compute the predicted data. Because of the problem scale, FWI is usually solved by gradient based algorithm. When the initial model is not far away from the true model, FWI is an effective method [Virieux and Operto, 2009]. However, gradient based FWI algorithm usually fails to recover long scale component of the model (parameters to describe the structures of the earth) if a good initial model is unavailable [Gauthier et al., 1986].

Born waveform inversion is a partially linearized version of full waveform inversion. It is defined as a least squares data fitting problem, where the data is predicted using Born modeling, based on scale separation of the model: model is explicitly separated as a sum of long scale background model and short scale reflectivity. Both background model and reflectivity are updated to explain the data. By introducing additional axes to the reflectivity model (background model remains physical), we get an extended Born waveform inversion. With additional degrees of freedom or enlarged domain of search, it is much easier to drive the objective function to a minimum value than the non-extended problem. Thus A DSO penalty term is added to gain some control on the extension [Symes, 2008b].

The resulting extended Born waveform inversion is quadratic in the reflectivity

and thus a nested optimization scheme is employed: namely the variable projection scheme [Golub and Pereyra, 2003], that is for a fixed background model, solve for the reflectivity in an inner stage and results in a background model only objective function [Symes and Carazzone, 1991, Symes, 1993, Kern and Symes, 1994] or the VPE objective function, then update the background model in an outer stage. More recent implementation and numerical results could be found in [Huang and Symes, 2015], where they show numerically that VPE method is more stable and less sensitive to the initial model than its non-extended analogue [van Leeuwen and Mulder, 2009], which is also called reflection waveform inversion by [Xu et al., 2012].

Although VPE method is stable and tends to have large region of convexity in the background model space, it requires the inner problem over the reflectivity (similar to least squares migration (LSM) problem) to be solved accurately enough and thus a lot of iterations are needed. This paper uses a flexibly preconditioned least squares migration scheme (FPCG) [Huang et al., 2016] to accelerate the convergence rate of the inner problem and thus reduce the total computational cost of the inversion. The preconditioner is based on optimal pseudodifferential scaling, proposed in [Nammour and Symes, 2009, Nammour, 2009].

In this paper, we present examples based on Gaussian anomaly model with one reflector below the anomaly that show the inner problem could be solved accurately enough with much less computational cost using FPCG method than the non-preconditioned method. And also we get a better shaped VPE objective function and more accurate gradient computation using the FPCG method than that without preconditioning. We find however that the accuracy of the computed gradient cannot be bounded by the error in the solution of the inner problem, and explain the source of gradient error.

4.3 Theory

Denoted by d the observed data, $m(\mathbf{x})$ the background model (smooth with respect to the wavelength) and $\delta\bar{m}(\mathbf{x}, \mathbf{h})$ the extended reflectivity model, with \mathbf{x} the spatial coordinate and \mathbf{h} the extended coordinate. We use M to denote the physical model space and \bar{M} the extended model space, with M a subspace of \bar{M} . $\bar{F}[m]\delta\bar{m}$ is the extended Born modeling operator at background model m with perturbation $\delta\bar{m}$. If the additional axis is shot record, the extended perturbation means that for different shot, the perturbation is allowed to be different. The extended Born waveform inversion is defined in equation 4.1.

$$J_{\text{EBWI}}[m, \delta\bar{m}] = \frac{1}{2} \|\bar{F}[m]\delta\bar{m} - d\|^2 + \frac{\alpha^2}{2} \|A\delta\bar{m}\|^2. \quad (4.1)$$

Problem 4.1 is a separable least-squares problem since it is quadratic on the reflectivity model. Thus a version of nested optimization scheme: variable projection method is an efficient way to minimize it. The quadratic variable $\delta\bar{m}$ is eliminated in an inner stage by minimizing over the extended model space and results in a background model only objective function, which is referred to as the reduced objective function or VPE objective in this paper and has the following form:

$$J_{\text{VPE}}[m] = \min_{\delta\bar{m} \in \bar{M}} J_{\text{EBWI}}[m, \delta\bar{m}] \quad (4.2)$$

A non-extended version of this objective function has been investigated to update the long scale background model by [van Leeuwen and Mulder, 2009, Xu et al., 2012]. See [Huang and Symes, 2015] for the most recent results of the VPE problem and the comparison with other relevant methods: VPE method is more stable and less sensitive to initial model than its non-extended version.

4.3.1 Computation of Extended Born Modeling Operator

For different model extensions, wave equations that need to be solved to compute the predicted extended Born modeling data are slightly different. We use the second order constant density acoustic wave equation (equation 4.3) in this paper to describe the wave propagation. For shot record model extension, $\bar{F}[m]\delta\bar{m}$ is the solution $\delta u(\mathbf{x}, t; \mathbf{x}_s)$ of the following wave equations 4.3 and 4.4 sampled at receiver positions \mathbf{x}_r , with \mathbf{x}_s the position of source impulse (shot) and $w(t)$ the source wavelet. m is the velocity-squared model and is the background model and $\delta\bar{m}$ is the velocity-squared perturbation, allowed to be different for different source position. The perturbed wave field δu is most accurate when the background model m is smooth and $\delta\bar{m}$ is oscillatory in the wavelength scale [Symes, 2009].

$$\begin{aligned} \left(\frac{\partial^2}{\partial t^2} - m(\mathbf{x})\nabla^2 \right) u(\mathbf{x}, t; \mathbf{x}_s) &= w(t)\delta(\mathbf{x} - \mathbf{x}_s). \\ u(\mathbf{x}, \mathbf{x}_s, t) &= 0, t \ll 0. \end{aligned} \quad (4.3)$$

$$\begin{aligned} \left(\frac{\partial^2}{\partial t^2} - m(\mathbf{x})\nabla^2 \right) \delta u(\mathbf{x}, t; \mathbf{x}_s) &= \delta m(\mathbf{x}; \mathbf{x}_s)\nabla^2 u(\mathbf{x}, t; \mathbf{x}_s) \\ \delta u(\mathbf{x}, \mathbf{x}_s, t) &= 0, t \ll 0. \end{aligned} \quad (4.4)$$

We use regular grid finite-difference scheme to discretize wave equations 4.3 and 4.4. To be convenient, the computational domain in 2D is a rectangle. Time interval is discretized as $t_n = n\Delta t$. The spatial grid size and Δt need to satisfy the CFL condition [Cohen, 2002]. We use second order leapfrog scheme for time discretization and use Dirichlet boundary condition for simplicity, that is $u = 0$ at the four boundaries of the computational domain. In the design of numerical experiments, we only allow the reflected energy on the top boundary. The computational domain is extended such that reflection energies from the rest three boundaries will never be recorded at receivers, and thus mimics absorbing boundary condition on three sides of the

domain. The numerical scheme that propagates wave field from time n to time $n + 1$ by solving equation 4.3 has the following form:

$$u^{n+1} = 2u^n - u^{n-1} + \Delta t^2 m L u^n + \Delta t^2 \omega^n \delta(\mathbf{x} - \mathbf{x}_s) \quad (4.5)$$

where $u^n \approx u(\mathbf{x}, t^n; \mathbf{x}_s)$, L is the discretized Laplacian operator with zero Dirichlet boundary condition and $\omega^n = \omega(t^n)$. By sampling the solution $\{u^n\}$ at receiver positions, we compute the full waveform data. For each shot position \mathbf{x}_s , wave equation 4.3 is solved independently.

The discretized perturbed wave field δu^n solves the following discretized wave equation

$$\delta u^{n+1} = 2\delta u^n - \delta u^{n-1} + \Delta t^2 m L \delta u^n + \Delta t^2 \delta m L u^n,$$

with u^n the solution of the numerical scheme 4.5. For discretizing the Laplacian operator L , we adopt the same zero Dirichlet boundary condition.

The position of sources and receivers are usually dozens of meters below the surface of the earth or the ocean in real acquisition. Thus in the numerical experiment, we put sources 20 meters and receivers 10 meters below the top boundary. This fact together with the zero Dirichlet boundary condition on top (free surface) result in a vertical derivative over the data in both the source side and the receiver side (D_{z_s} and D_{z_r} with z_s and z_r the second coordinate of \mathbf{x}_s and \mathbf{x}_r respectively) as explained in [Hou and Symes, 2015b]. Since these derivatives are applied to the data, which solves wave equations, these derivatives are roughly equivalent to $D_t^2 = \frac{\partial^2}{\partial t^2}$. And thus this numerical scheme and acquisition setting introduce two additional orders to the forward modeling operator and to the Born modeling operator $\bar{F}[m]$ as will be demonstrated in the Numerical Examples section.

The operator $\bar{F}[m]^T \bar{F}[m]$ is a *pseudodifferential operator* (Ψ DO) under assump-

tions: there are no caustics and multi-pathing [Beylkin, 1985, Rakesh, 1988b, Nolan and Symes, 1990, ten Kroode et al., 1998, Stolk, 2000, Stolk and Symes, 2004]. A Ψ DO does not move singularities in its input and has an associated order: to what degree the singularity will be scaled. Theoretically, with absorbing boundary condition, the order of $\bar{F}[m]^T \bar{F}[m]$ is $n - 1$ for n -dimension problems, and we will consider 2D problem from now on, and $\bar{F}[m]^T \bar{F}[m]$ has order 1 in this case. With free surface condition, $\bar{F}[m]^T \bar{F}[m]$ becomes an order 5 Ψ DO: two additional orders in $\bar{F}[m]$ and two additional orders in $\bar{F}[m]^T$. One way to get rid of the influence of free surface condition is to compose the extended Born modeling operator with the square of the integral operator I_t^2 , with

$$I_t = \int_t dt.$$

The integral over t operator I_t is an order -1 operator. By composing with I_t^2 , we add an additional -2 order to the modeling operator, which un-do the action of D_t^2 . This trick gives us an order 1 Ψ DO $\bar{F}[m]^T \bar{F}[m]$, with $\bar{F}[m]$ the original extended Born modeling operator composing with I_t^2 .

4.3.2 Flexibly Preconditioned Least Squares Migration with Regularization

To compute the value of the VPE objective function 4.2 at a given background model m , a regularized least squares migration (RLSM) needs to be solved

$$(\bar{F}[m]^T \bar{F}[m] + \alpha^2 A^T A) \delta \bar{m} = \bar{F}[m]^T d. \quad (4.6)$$

We have taken care of the free surface condition and get an order 1 Ψ DO $\bar{F}[m]^T \bar{F}[m]$.

We want to check the order of the normal operator

$$N[m] = \bar{F}[m]^T \bar{F}[m] + \alpha^2 A^T A.$$

The operator $A^T A$ is a differential operator (a particular case of Ψ DO). For different model extension, $A^T A$ might be different and thus could have different order. We are using shot record domain model extension in this paper, which requires the target solution to be physical, in other words, constant along the extended axis. Thus a natural choice to penalize the non-physicality of the extended reflectivity is $A = \nabla_{\mathbf{x}_s}$, which is an order 1 differential operator and $A^T A$ is an order 2 differential operator. Thus problem 4.6 is an ill-conditioned problem despite the choice of α .

Optimal pseudodifferential scaling method proposed in [Nammour and Symes, 2009, Nammour, 2009] is a choice of preconditioner to improve the convergence rate of algorithm to solve problem 4.6. For each vector an order s Ψ DO applied on, the optimal pseudodifferential scaling algorithm finds an order $-s$ Ψ DO that approximately undoes the application of the order s Ψ DO, by solving an optimization problem.

The sum of an order 1 Ψ DO $\bar{F}[m]^T \bar{F}[m]$ and an order 2 Ψ DO $A^T A$ gives an order 2 Ψ DO. In order for the preconditioning to take into account of both operators, we first change the operator A to

$$A = \nabla_{\mathbf{x}_s} \Lambda^{-\frac{1}{4}},$$

with

$$\Lambda = I - L \tag{4.7}$$

and $L = \omega_x^2 \frac{\partial^2}{\partial x^2} + \omega_z^2 \frac{\partial^2}{\partial z^2} + \omega_{x_s}^2 \frac{\partial^2}{\partial x_s^2}$. Λ is an order 2 differential operator and thus A as a Ψ DO has order $\frac{1}{2}$. The resulting normal operator $N[m]$ for this choice of A has order 1. And then the optimal pseudodifferential scaling algorithm could be used to compute preconditioner to improve the convergence of problem 4.6.

The optimal pseudodifferential scaling problem is defined as: for a given $b \in \bar{M}$, find an order -1 Ψ DO Q , that minimizes

$$\|b - QN[m]b\|^2. \quad (4.8)$$

Please refer to [Nammour, 2009] and [Bao and Symes, 1996] for the implementation and analysis of this problem. In [Nammour, 2009], it is claimed that $Qb \approx N[m]^{-1}b$ asymptotically. Note that the minimizer Q depends on b . Based on problem 4.8 and the flexible preconditioned conjugate gradient method [Notay, 2000, Knyazev and Lashuk, 2006], we propose to use the following flexibly preconditioned conjugate gradient (FPCG) algorithm to solve problem 4.6, which is the normal equation of the following quadratic problem over reflectivity model space

$$\frac{1}{2} \|\hat{F}[m]\delta\bar{m} - \hat{d}\|^2,$$

$$\text{with } \hat{F}[m] = \begin{pmatrix} \bar{F}[m] \\ \alpha A \end{pmatrix} \text{ and } \hat{d} = \begin{pmatrix} d \\ 0 \end{pmatrix}.$$

4.3.3 Gradient Computation of J_{VPE}

The accuracy of Born modeling as a partially linearized modeling depends on the scale separation of the model: background model contains only long scale low frequency components of the model, while the reflectivity contains only short scale high frequency [Symes, 2009]. Thus, we enforce some smoothness in the background model by using a weighted inner product in its space: $\langle m_1, m_2 \rangle_M = \langle m_1, \Lambda m_2 \rangle$, with Λ as the roughening operator defined in equation 4.7 and $\langle \cdot, \cdot \rangle$ the ordinary L^2 inner product. With $\delta\bar{m}$ the solution of 4.6 using Algorithm 2, the gradient of J_{VPE} has the following

Algorithm 2 A flexibly preconditioned conjugate gradient method

Given data \hat{d} and $\delta\bar{m}_0 = 0$, modeling operator \hat{F} , and adjoint operator \hat{F}^T
 Given maximum iterations MaxIter and $i = 0$
 $r_0 = \hat{d}$, $G_0 = \hat{F}^T r_0$
 compute Q_0 by solving equation 4.8 with $b = G_0$
 $g_0 = Q_0 G_0$
 $p_0 = g_0$
while (not converge && $i < \text{MaxIter}$) **do**
 $q_i = \hat{F} p_i$
 $\alpha_i = \frac{\langle g_i, G_i \rangle}{\langle q_i, q_i \rangle}$
 $\delta m_{i+1} = \delta m_i + \alpha_i p_i$
 $r_{i+1} = r_i - \alpha_i q_i$
 $G_{i+1} = \hat{F}^T r_{i+1}$
 compute Q_{i+1} by solving equation 4.8 with $b = G_{i+1}$
 $g_{i+1} = Q_{i+1} G_{i+1}$
 $\beta = \frac{\langle g_{i+1}, G_{i+1} - G_i \rangle}{\langle g_i, G_i \rangle}$
 $p_{i+1} = g_{i+1} + \beta p_i$
 $i = i + 1$
end while

form

$$\nabla J_{\text{VPE}}[m] = \Lambda^{-1} D\bar{F}[m]^T [\delta\bar{m}, \bar{F}[m]\delta\bar{m} - d]. \quad (4.9)$$

The extended Born modeling operator \bar{F} is the first order derivative of the forward modeling operator. Then $D\bar{F}$ is the second order derivative or Hessian of the forward modeling operator, and is a bilinear operator. By fixing one argument, we could define its adjoint $D\bar{F}[m]^T[\delta\bar{m}, \cdot]$, which is usually called tomographic or WEMVA operator [Biondi and Sava, 2004, Biondi and Almomin, 2012]. With the gradient 4.9, we could use a gradient based optimization algorithm to update the background model.

4.4 Numerical Examples

4.4.1 Property of normal operator $\bar{F}[m]^T \bar{F}[m]$

We use constant background model $m = (2\text{km/s})^2$ with one reflector (Figure 4.10b) to numerically study features of operator $\bar{F}[m]$ and $\bar{F}[m]^T \bar{F}[m]$. Since shot record model extension does not change features of Born modeling and migration operators, we use one shot in this test for simplicity. Models are discretized by $10\text{ m} \times 10\text{ m}$ simulation grid. The position of shot is at 20 meter depth and 2000 meter distance. We put 401 receivers at 10 meter depth. The source wavelet $\omega(t)$ is shown in Figure 4.1b.

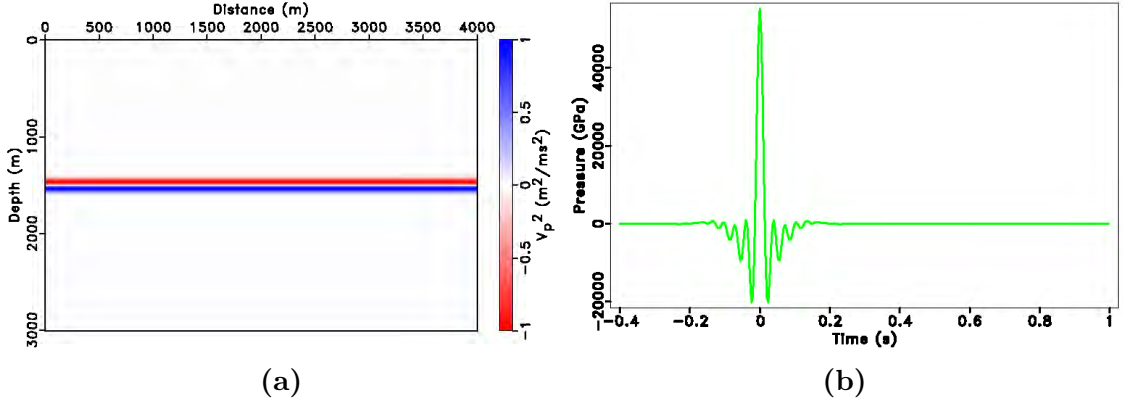


Figure 4.1: Target reflectivity δm (a) and source wavelet with corner frequency 1 Hz, 7 Hz, 28 Hz, 35 Hz (b).

We compute

$$F[m](\delta m + r_k)$$

and

$$F^T[m]F[m](\delta m + r_k)$$

where $F[m]$ is the original Born modeling operator, without composing with I_t^2 and

$$r_k = W \sin(kz).$$

δm is the target reflectivity model shown in Figure 4.10b. W is a vertical windowing operator shown in Figure 4.2. r_k is scaled such that $\|r_k\| = 0.5\|\delta m\|$ and the scale is the same of all k .

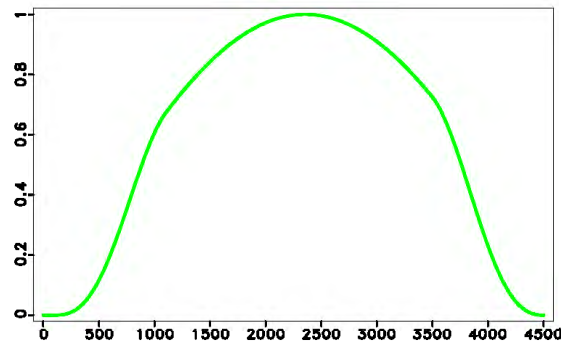


Figure 4.2: Windowing operator W .

We first show outputs of Born operator $F[m]$ and normal operator $N[m] = F[m]^T F[m]$ applied to models $\delta m + r_k$ for $k = 30$ and $k = 70$. Although the L_2 norm of r_k is the same for different k , the frequency components are very different. Figure 4.3b shows the frequency components of r_k with $k = 30$ and the corresponding reflectivity model $\delta m + r_k$ is shown in Figure 4.3a. The Born data and the output of normal operator $N[m]$ to this model are shown in Figure 4.4a and Figure 4.4b. The error r_k in reflectivity model produces a visible error in data, but the main signal is still clear.

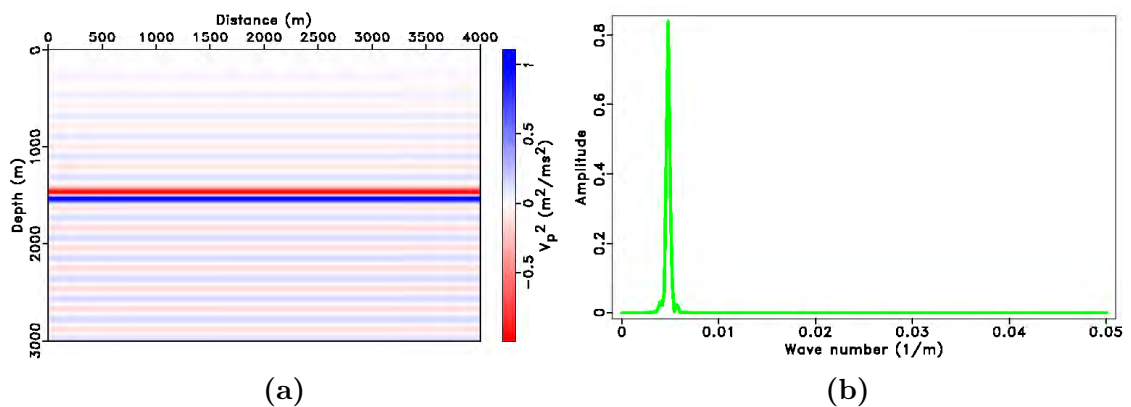


Figure 4.3: Model $\delta m + r_k$ (a) and spectrum of model r_k (b) with $k = 30$.

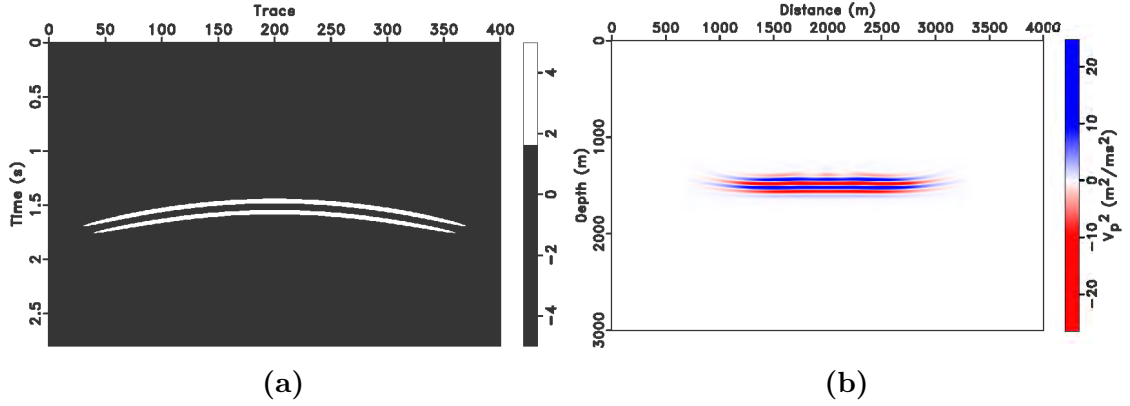


Figure 4.4: With bandpass filtered source wavelet: (a) $F[m](\delta m + r_k)$ and (b) $N[m](\delta m + r_k)$ with $\delta m + r_k$, $k = 30$ shown in Figure 4.3a.

We increase the frequency in r_k by setting $k = 70$ (Figure 4.5b). The corresponding reflectivity model $\delta m + r_k$ is shown in Figure 4.5a. The Born data and the output of $N[m]$ from this model are shown in Figure 4.6a and Figure 4.6b. Although r_k has the same L_2 norm as when $k = 30$, it produces a significant error in data and image. Even the main signal is hard to distinguish.

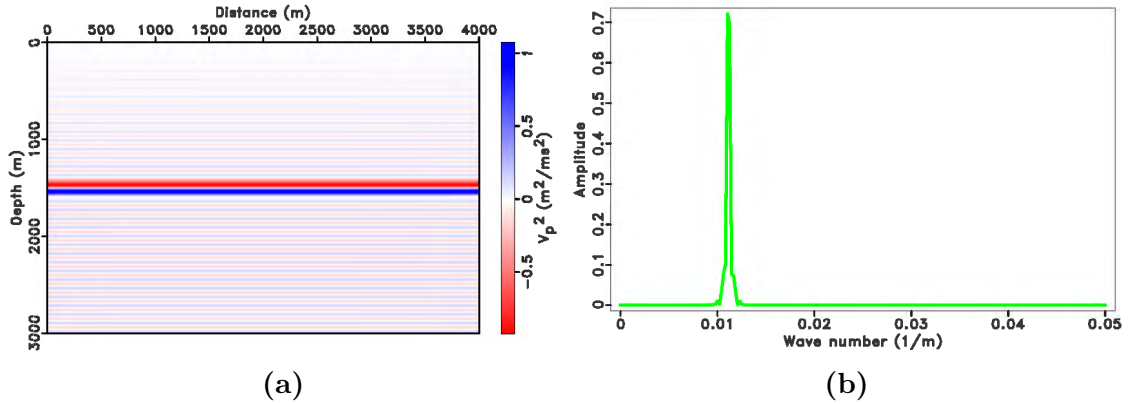


Figure 4.5: Model $\delta m + r_k$ (a) and spectrum of model r_k (b) with $k = 70$.

r_k is considered as the error in the target reflectivity model δm . The L_2 norm of r_k is the same for different k as shown in Figure 4.7. However, after application of the Born modeling operator $F[m]$, the L_2 norm $F[m]r_k$ increases (Figure 4.7a). We fit the curve with power law estimation. The result shows the order is 2.45067 as

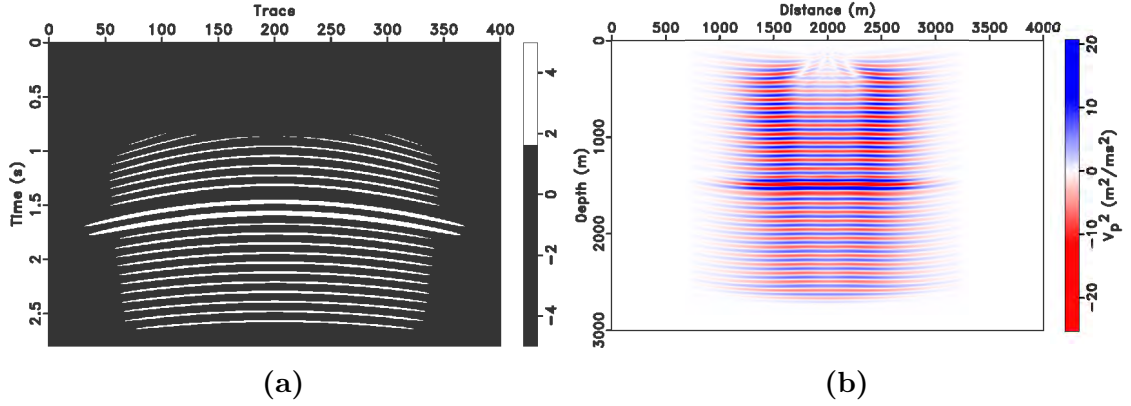


Figure 4.6: With bandpass filtered source wavelet: (a) $F[m](\delta m + r_k)$ and (b) $N[m](\delta m + r_k)$ with $k = 70$.

plotted in Figure 4.7a. Apply the normal operator $N[m]$ to the same sequence of r_k , and compute the relative error $\frac{\|N[m]r_k\|}{\|N[m]\delta m\|}$, we get the curve as shown in Figure 4.7b. The similar power law estimation shows that the order is 4.85. Because we are using free surface and put sources and receivers below the surface, the estimated orders are as expected, despite some fitting errors.

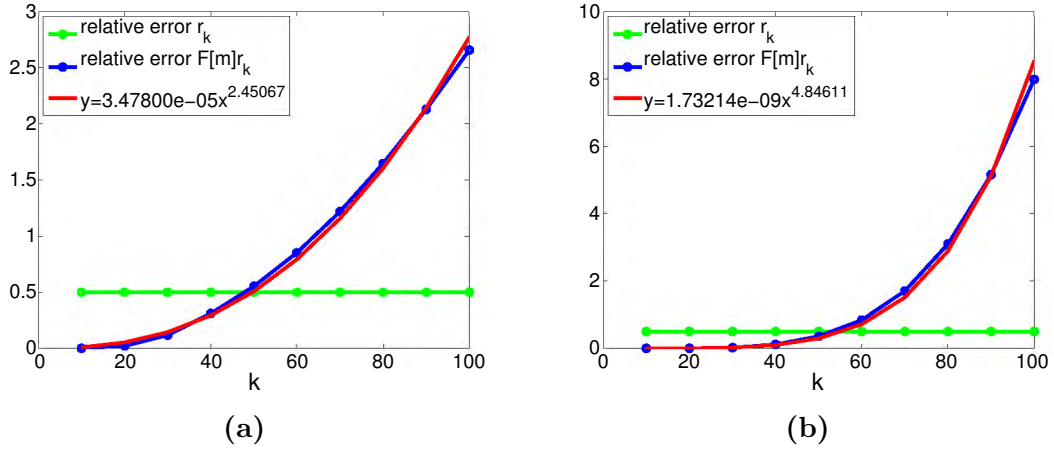


Figure 4.7: With bandpass filtered source wavelet: Green curve in both subfigures is the relative error $\frac{\|r_k\|}{\|\delta m\|}$. (a) the relative error of Born modeling operator $\frac{\|F[m]r_k\|}{\|F[m]\delta m\|}$ and (b) the relative error of normal operator $\frac{\|N[m]r_k\|}{\|N[m]\delta m\|}$ with $r_k = W \sin(kz)$, $m = (2\text{km/s})^2$.

Integrating the source wavelet twice will take care of the two additional orders in Born modeling operator caused by free surface boundary condition. This is equivalent to compose the Born modeling operator $F[m]$ with I_t^2 . Figure 4.8 shows the source wavelet with twice integral over time of the wavelet shown in Figure 4.1b. Same tests as these with bandpass filtered source wavelet are performed. We would like to omit outputs of Born modeling operator and the normal operator with this new source, since these outputs are very similar to these with the original source.

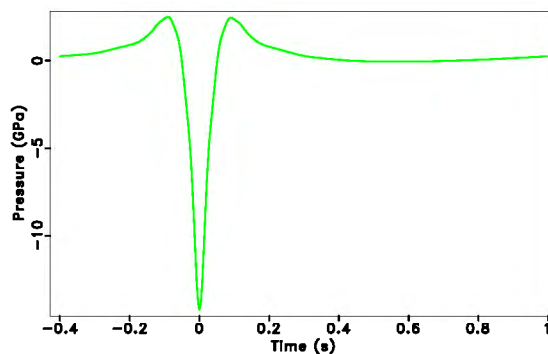


Figure 4.8: Source wavelet after twice integral over time of wavelet in Figure 4.1b.

The same relative error $\frac{\|F[m]r_k\|}{\|F[m]\delta m\|}$ is computed with this integrated source wavelet and shown in Figure 4.9a. We fit the curve with power law estimation. The result shows the order is 0.476 as plotted in Figure 4.9a. Apply the normal operator $N[m]$ to the same sequence of r_k with the same integrated source, and compute the relative error $\frac{\|N[m]r_k\|}{\|N[m]\delta m\|}$, we get the curve as shown in Figure 4.9b. The similar power law estimation shows that the order is 0.928.

As we discussed in the Theory section, the application of free surface condition is equivalent to applying D_{z_s} and D_{z_r} to the data, which is only roughly equivalent to applying D_t^2 , with the approximation closest at high frequency. Thus by composing the Born modeling operator with I_t^2 , the order of the resulting operator will be 0.5 only asymptotically. Thus we exclude the first two points ($k = 10, 20$) when we do the power law fit. Despite these errors, these orders are as expected. And the normal

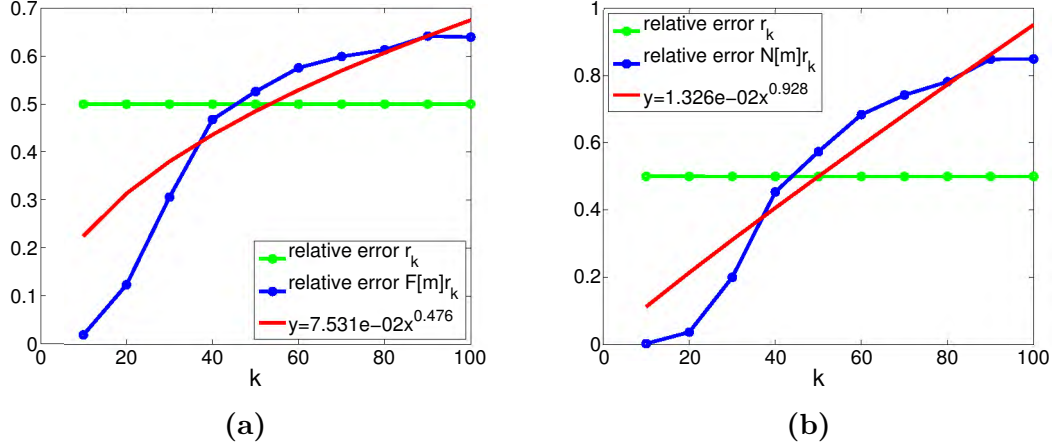


Figure 4.9: With twice integrated source wavelet: Green curve in both subfigures is the relative error $\frac{\|r_k\|}{\|\delta m\|}$. (a) the relative error of Born modeling operator $\frac{\|F[m]r_k\|}{\|F[m]\delta m\|}$ and (b) the relative error of normal operator $\frac{\|N[m]r_k\|}{\|N[m]\delta m\|}$ with $r_k = W \sin(kz)$, $m = (2\text{km/s})^2$.

operator has been modified to be an order 1 Ψ DO as we want.

4.4.2 Flexibly Preconditioned Least Squares Migration with Regularization

In this section, we use fast Gaussian anomaly model (Figure 4.10a) with one reflector (Figure 4.10b) to demonstrate the performance of the flexibly preconditioned least squares migration with regularization.

Wave equations 4.3 and 4.4 are discretized as discussed in section 4.3.1. Isotropic point source, with a bandpass filtered source wavelet (corner frequency 1 Hz, 7 Hz, 28 Hz, 35 Hz, Figure 4.11a) is used. A total of 81 shots are placed at depth 20 m with a spacing of 50 m starting at 0 m on the left. 401 fixed receivers are located at every grid points of depth 10 m. Figure 4.12 shows the computed Born data that is used as the observed data in the following tests (3 of 81 shots are shown here).

In this subsection, we present results of solving the regularized least squares mi-

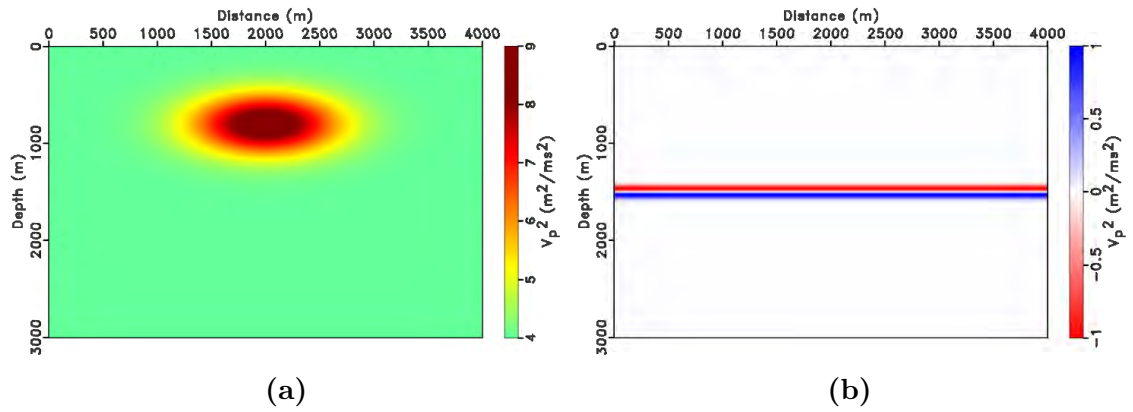


Figure 4.10: (a) target background model and (b) target reflectivity.

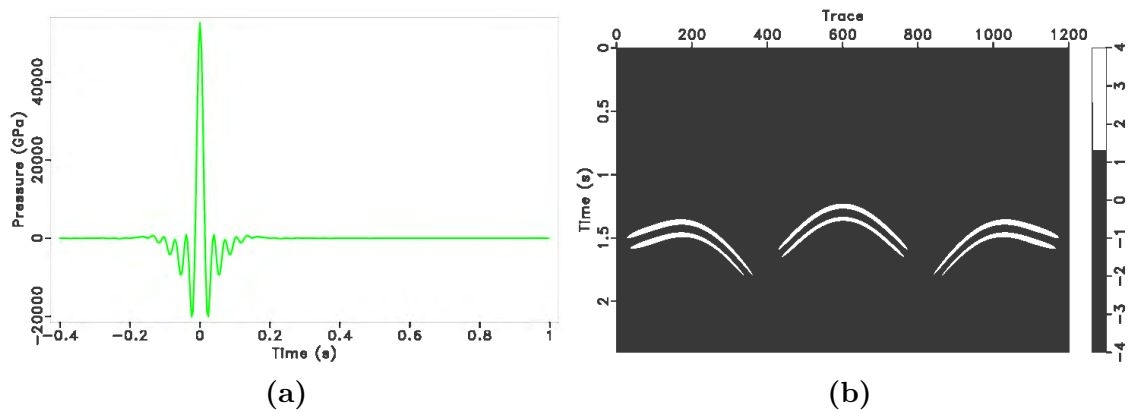


Figure 4.11: (a) source wavelet with corner frequency 1 Hz, 7 Hz, 28 Hz, 35 Hz and (b) 3 of 81 Born shot records.

gration problem 4.6 using Algorithm 2. The first step is to compose the extended Born modeling operator with I_t^2 and get a modified normal operator with order 1. The observed data is preprocessed by applying I_t^2 (shown in Figure 4.12), in order to be consistent with the modeling operator. This is considered as a preconditioning method with preconditioner I_t^4 . We then apply FPCG to this modified problem and compare the performance with Conjugate Gradient (CG) methods applied to this modified problem and the original problem.

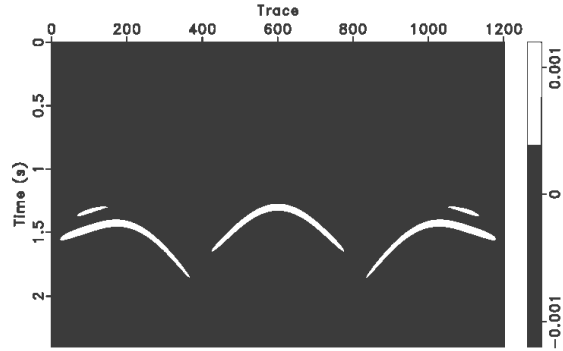


Figure 4.12: Data integrated over t twice.

The value of α is 0.005 and fixed for all tests based on the same trial-and-error as discussed in Chapter 2 and [Huang and Symes, 2015]. We solve the above equation at different background models: at target background model and at the constant background model ($(2 \text{ km/s})^2$). Figure 4.13a shows the normalized residual: the relative error of $\frac{1}{2} \|\bar{F}[m]\delta\bar{m} - d\|^2 + \frac{\alpha^2}{2} \|A\delta\bar{m}\|^2$ compared with the error at initial reflectivity model $\delta\bar{m} = 0$ at the target background model m as shown in Figure 4.10a and Figure 4.13b shows the normalized gradient length of using two algorithms: the Algorithm 2 and its non-preconditioned version. CG with windowing is the conjugate gradient method applied to the original data, with the modeling operator composed with a windowing operator, which smoothly zeros the top 100 m in the reflectivity model to mitigate the acquisition footprint. The horizontal axis is the number of Hessian applications, which is the number of the applications of $\bar{F}[m]^T \bar{F}[m] + \alpha^2 A^T A$.

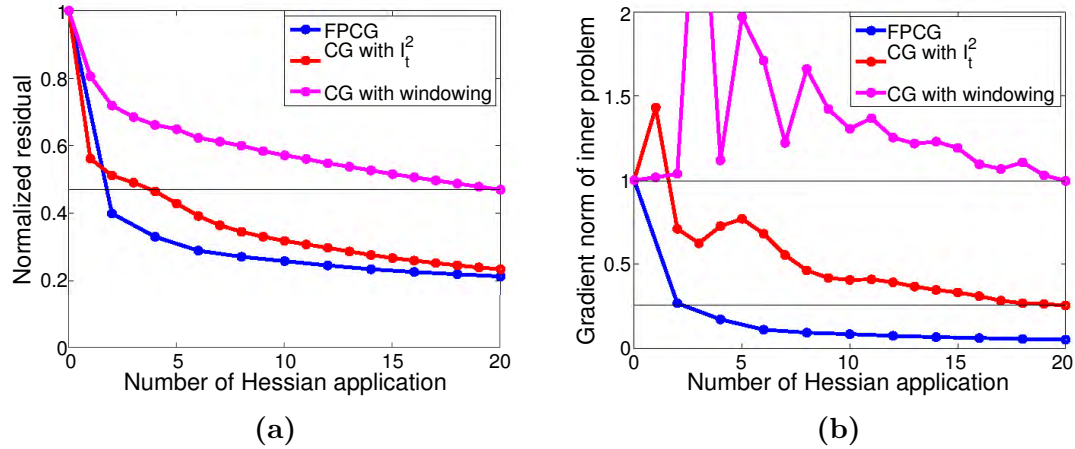


Figure 4.13: At true background model: convergence curves of numerical methods, (a) normalized residual and (b) normalized gradient length.

The first horizontal black line in Figure 4.13a is the position of the error of conjugate gradient method with windowing after 20 applications of Hessian, which only achieves the error level of FPCG method with 2 applications of Hessian. CG with I_t^2 is the method of applying CG to the problem with extended Born modeling operator replaced by $I_t^2 \bar{F}[m]$. For the normalized gradient norm, FPCG with 2 Hessian application achieves the same error level of CG with I_t^2 and a much smaller error level than CG with windowing after 20 Hessian applications.

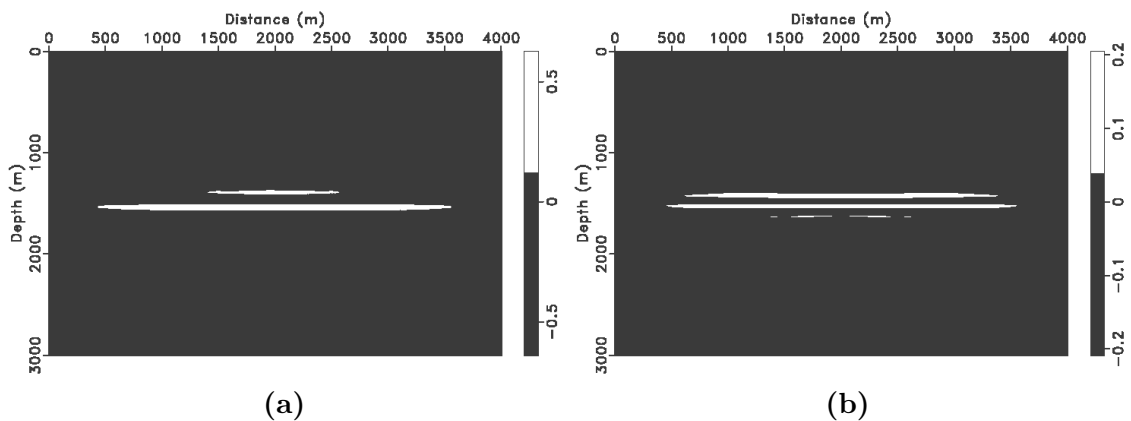


Figure 4.14: At true background model: Stacked image after 20 Hessian applications using FPCG (a) and using CG with windowing (b).

The solution of problem 4.6 is a function of x -horizontal axis, z -vertical axis and

\mathbf{x}_s -shot coordinate. Thus $\delta\bar{m}(z, x, \mathbf{x}_s)$ is a 3D cube. Figure 4.14 shows the solution of problem 4.6 after stacking along the shot coordinate: $\sum_{\mathbf{x}_s} \delta\bar{m}$. Stacked image (reflectivity) after 20 steps of FPCG method (Figure 4.14a) has much less artifacts than that from CG method with windowing (Figure 4.14b).

Figure 4.15 shows 10 panels of common image gathers: $\delta\bar{m}(z, x, \mathbf{x}_s)$ at 10 different x positions. Since m is the target background model, gathers should be flat. FPCG method (Figure 4.15a) produces more clear and less artifacts gathers than CG with windowing method (Figure 4.15b).

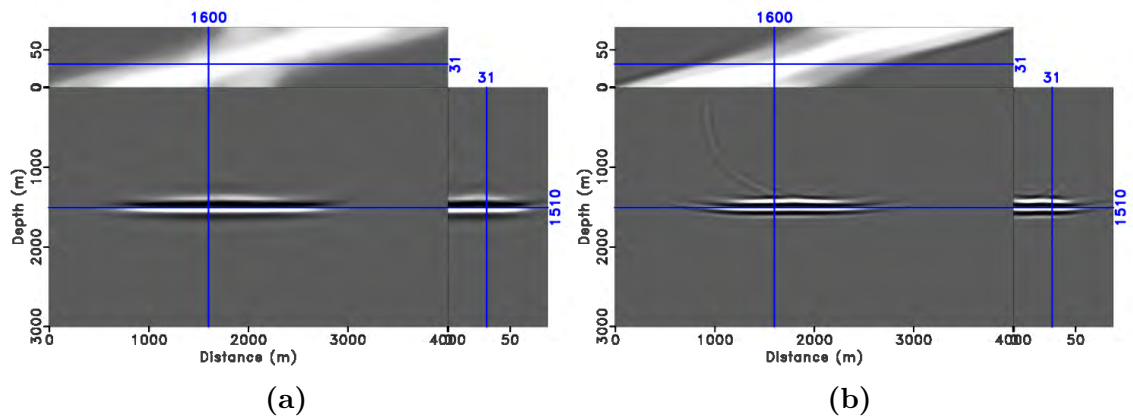


Figure 4.15: At true background model: Common image gathers after 20 Hessian applications using FPCG (a) and using CG with windowing (b).

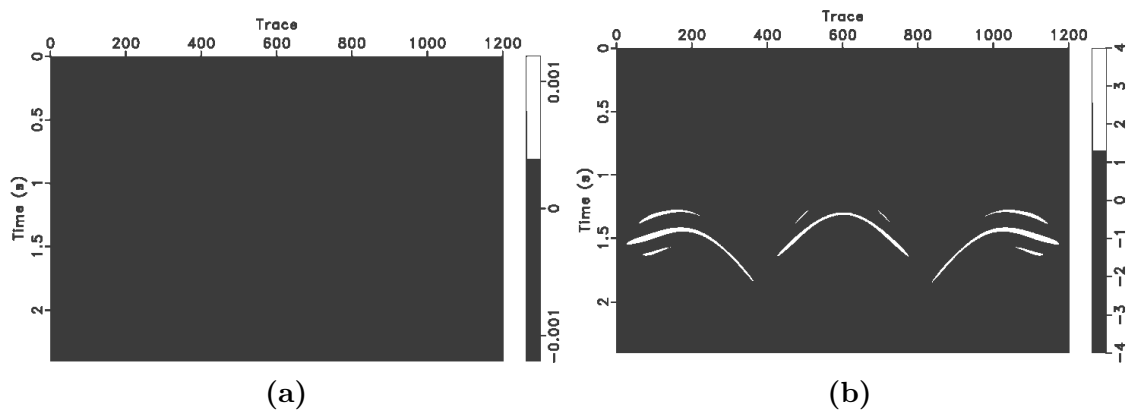


Figure 4.16: At true background model: data residual after 20 Hessian applications using FPCG (a) and using CG with windowing (b).

The data residual is shown in Figure 4.16. Note that we are solving a regularized least squares migration problem. Fitting the data is only part of the objective. FPCG method (Figure 4.16a) is much better than CG with windowing (Figure 4.16b) in the sense that much less residual is left during the optimization process.

Although a good image or reflectivity model is very useful by itself, the most useful information to us is the image gathers, from which we could tell how close is the background model to the target and in turn create methods to improve the background model. Thus we apply the FPCG method to solve the regularized extended least squares migration problem 4.6 at a wrong background model, $(2 \text{ km/s})^2$. The

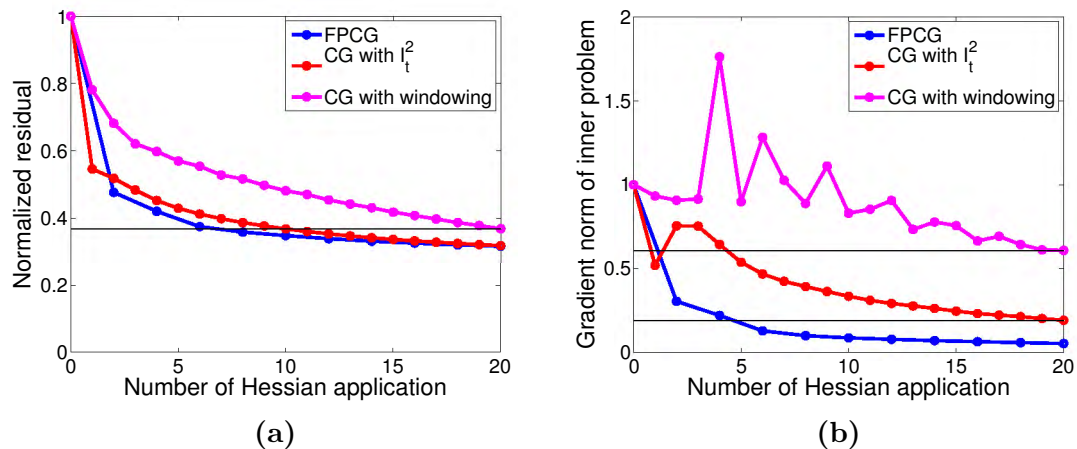


Figure 4.17: At constant background model $((2 \text{ km/s})^2)$: convergence curves of numerical methods, (a) normalized residual and (b) normalized gradient length.

convergence curves are shown in Figure 4.17. To get the same gradient reduction, FPCG method could save at least $\frac{1}{4}$ of the computational cost of CG with I_t^2 algorithm (Figure 4.17b). CG with windowing after 20 applications of Hessian could not reduce the gradient error to the same level of that by FPCG with 2 applications of Hessian. Figure 4.17a shows the residual by these three methods. FPCG reduces the residual very fast at the beginning of the iteration. However, the two methods reached the same error level eventually. This is because the background model is wrong. And thus part of the data cannot be fitted by shot domain model extension

methods no matter which algorithm we are using. 6 applications of Hessian by FPCG fits the data to the same level by 20 applications of Hessian by CG with windowing method.

Figure 4.18 shows the stacked version of the resulting reflectivity model or image. The constant background model is slower than the target background model at the center of the domain. Thus the position of the reflector is higher than the target one at the center. The image from FPCG method (Figure 4.18a) has less side-lobes or artifacts than that from CG with windowing (Figure 4.18b). The same improvement by FPCG method is observed in the common image gathers by comparing Figure 4.19a and Figure 4.19b.

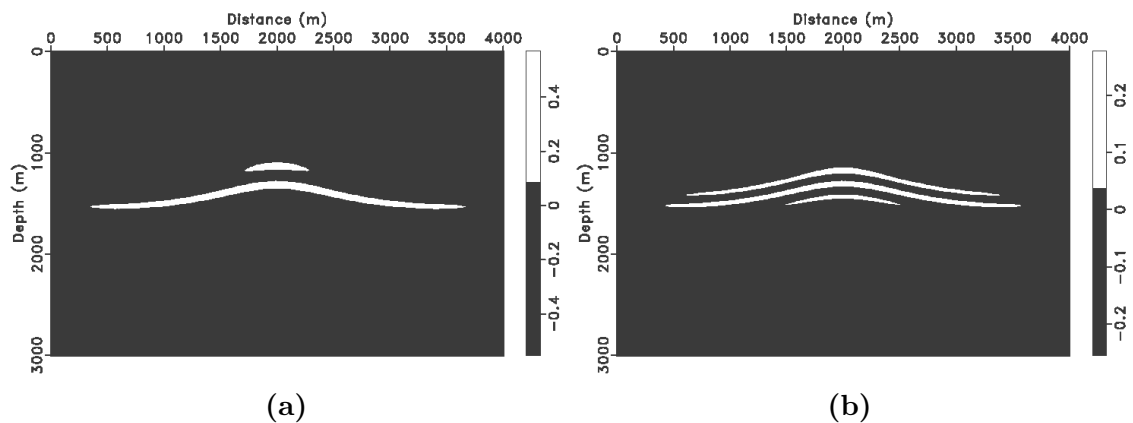


Figure 4.18: At constant background model ($(2 \text{ km/s})^2$): Stacked image after 20 Hessian applications using FPCG (a) and using CG with windowing (b).

Figure 4.20a shows the data residual after 20 applications of Hessian by FPCG method, which has less noise compared with the data residual from CG with windowing method (Figure 4.20b). Both the data residual and the image gather information will be used to compute the gradient for the outer optimization over the background model. Thus a FPCG method is preferred since it creates less noise in the solution and thus the data residual than that by CG method.

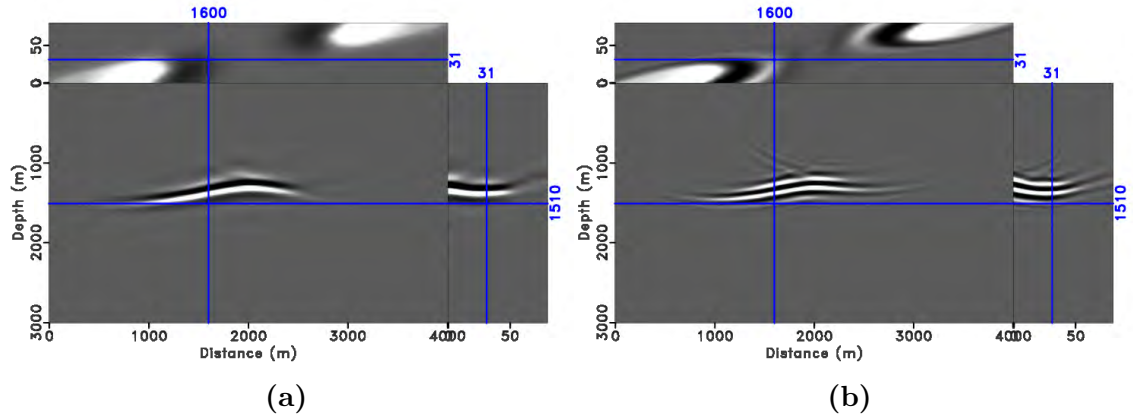


Figure 4.19: At constant background model ($(2 \text{ km/s})^2$): Common image gathers after 20 Hessian applications using FPCG (a) and using CG with windowing (b).

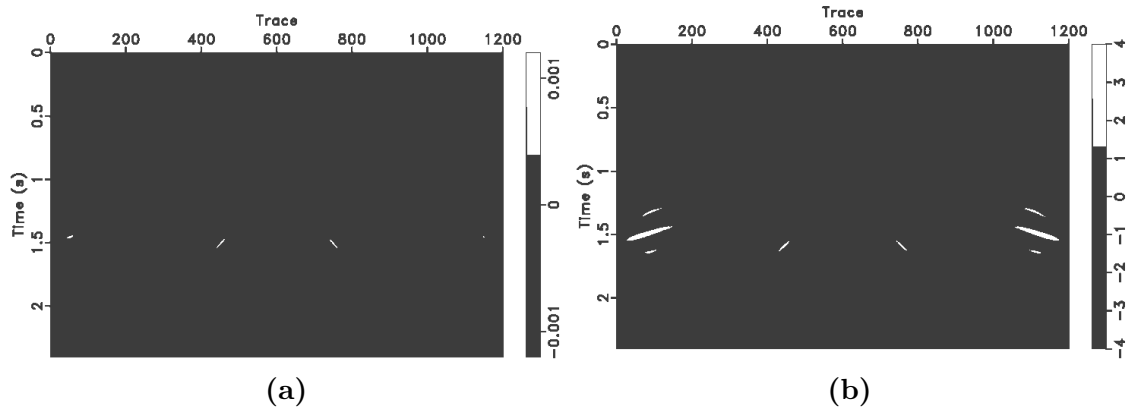


Figure 4.20: At constant background model ($(2 \text{ km/s})^2$): data residual after 20 Hessian applications using FPCG (a) and using CG with windowing (b).

4.4.3 Shape of $J_{\text{VPE}}[m]$ along a line segment

The preconditioning also improves the shape of the VPE objective function. We plot the value of $J_{\text{VPE}}[m]$ along a line segment in the background model space

$$m = \sigma m_{\text{target}} + (1 - \sigma)m_0,$$

with m_{target} shown in Figure 4.10a and $m_0 = (2\text{km/s})^2$. $\sigma = 1$ in Figure 4.21 is the position of target background model and $\sigma = 0$ is the position of constant background model m_0 .

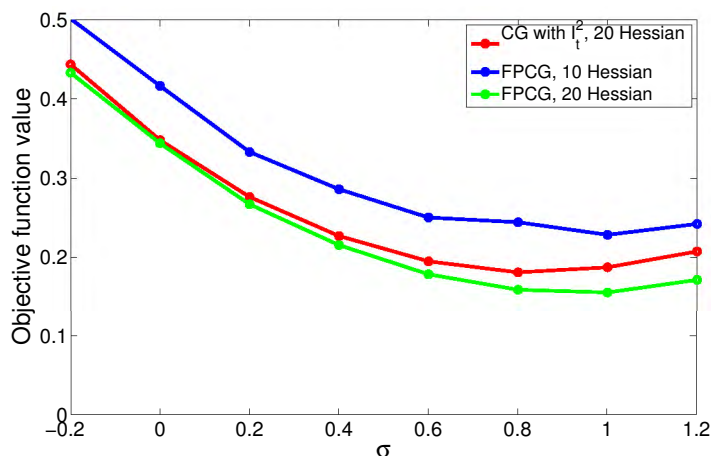


Figure 4.21: Value of VPE objective by different numerical methods and different number of inner iterations along line segment $m = \sigma m_{\text{target}} + (1 - \sigma)m_0$ with m_{target} shown in Figure 4.10a and $m_0 = (2\text{km/s})^2$ with $\sigma = 0$ the position of target background model.

We apply different number of iterations and different methods to solve the inner problem to evaluate the objective function $J_{\text{VPE}}[m]$. As shown in Figure 4.21, when using CG method with I_t^2 , even after 20 applications of Hessian, there is still a model ($\sigma = 0.8$) that gives smaller objective function value than the target model. With FPCG method to solve the inner problem, after 10 applications of Hessian, the target model produces the smallest objective function value among all these models we

tested. We do not show the same test using CG with windowing method, because it gives a very bad shape of the objective function (model that gives the smallest objective function value is far from the target one).

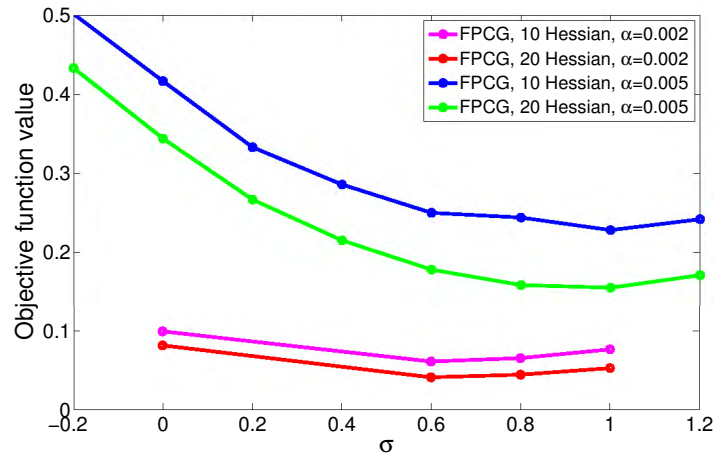


Figure 4.22: Value of VPE objective by FPCG method with different penalty weight α and different number of iterations along line segment $m = \sigma m_{\text{target}} + (1 - \sigma)m_0$ with m_{target} shown in Figure 4.10a and $m_0 = (2\text{km/s})^2$ with $\sigma = 0$ the position of target background model.

However, the shape of the objective function depends also on the value of the penalty weight α . As shown in Figure 4.22, with small α , the objective function is flatter than with large α . And there are models around the target model that give smaller values of objective function ($\sigma = 0.6$). This indicates that when the inverted background model is close to the target, we need to increase α . How to determine α at each iteration of the background update is out of the range of this paper and we will not discuss here.

This test shows that preconditioning has the ability to produce a better shaped objective function and could potentially improve the performance of the optimization over the background model.

4.4.4 Gradient of VPE objective function

In this section, we compute the gradient of $J_{\text{VPE}}[m]$ at $m = (2 \text{ km/s})^2$. The solution $\delta\bar{m}$ to the regularized LSM and the data residual as shown in Figure 4.20 are inputs to the WEMVA operator $D\bar{F}[m]^T$ to compute the gradient defined by equation 4.9.

Figure 4.23a shows the output of $D\bar{F}[m]^T[\delta\bar{m}, \bar{F}[m]\delta\bar{m} - d]$ with $\delta\bar{m}$ computed by FPCG method. Figure 4.23b shows the same output with $\delta\bar{m}$ computed by CG with windowing method. Then we apply Λ^{-1} to Figure 4.23 and get the gradient as shown in Figure 4.24.

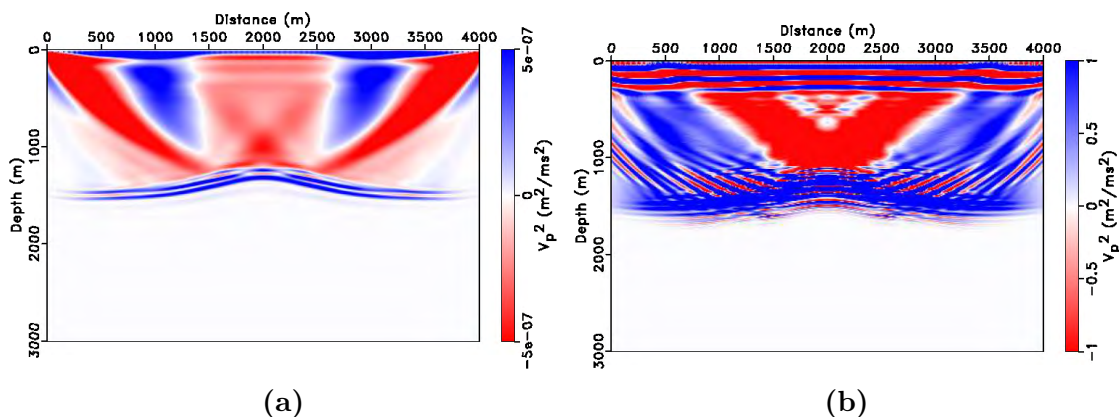


Figure 4.23: At constant background model ($m = (2 \text{ km/s})^2$): output of $D\bar{F}[m]^T[\delta\bar{m}, \bar{F}[m]\delta\bar{m} - d]$ with $\delta\bar{m}$ the solution of problem 4.6 after 20 Hessian applications using FPCG (a) and using CG with windowing (b).

The implementation is explained in Appendix A. We set the top 100 m to zero and use odd boundary condition on the top of the domain and use even boundary condition on the rest three boundaries for discrete Fourier transform. We choose $\omega_x = \omega_z = 100$ in this example. The anomaly in the middle of the domain has been successfully indicated in the gradient, although there are unwanted noise (blue color) around the anomaly (red color)(Figure 4.24).

We also compute a gradient accuracy test as we did in Chapter 2. We compute the relative error of the following two quantities for $h = 0.2$, with the one on the right

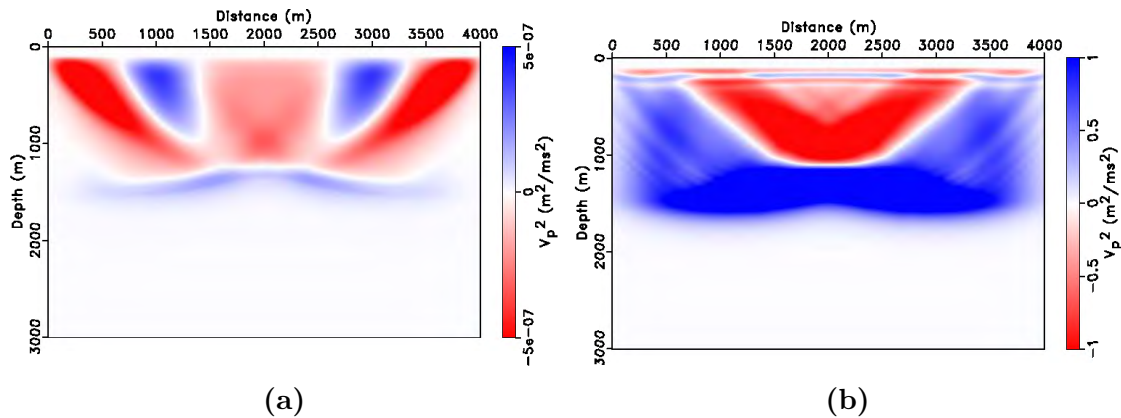


Figure 4.24: At constant background model ($m = (2 \text{ km/s})^2$): gradient computed by equation 4.9 with $\delta\bar{m}$ the solution of problem 4.6 after 20 Hessian applications using FPCG (a) and using CG with windowing (b).

Numerical methods	dm1			dm2		
h	0.1	0.05	0.025	0.1	0.05	0.025
FPCG	0.1108	0.0525	0.0643	0.1715	0.0814	0.1346
CG with I_t^2	0.1131	0.1147	0.1105	0.0467	0.0579	0.0633
CG with windowing	0.9890	1.0088	1.0673	1.3502	1.1384	0.6748

Table 4.1: Gradient test at constant background model $m = (2\text{km/s})^2$ for different dm and different numerical methods with 20 applications of LSM Hessian.

hand side as the value of reference

$$\frac{J_{\text{VPE}}[m + h * dm] - J_{\text{VPE}}[m - h * dm]}{2h} \approx \langle \nabla J_{\text{VPE}}[m], dm \rangle_s, \quad (4.10)$$

at the background model $m = (2 \text{ km/s})^2$. We compute the relative error for different directions (different dm): the anomaly in the model (denoted by dm1 in Table 4.1) and the anomaly shifted 1000 m to the left (denoted by dm2 in Table 4.1). Relative errors by several methods in Table 4.1 tells us that preconditioning helps to improve the accuracy of the gradient computation, at least in directions we test. At direction dm1, FPCG works better than CG with I_t^2 as the preconditioner. At direction dm2, CG with I_t^2 produces more accurate gradient than FPCG. Both FPCG and CG with I_t^2 produce more accurate gradient than CG method without preconditioning.

Preconditioning can significantly improve the convergence rate of the inner problem and produce high quality reflectivity and can also improve the gradient computation based on our numerical test. However, we could not estimate the error in the gradient computation theoretically at this moment, which we will discuss the reason in the next subsection.

4.4.5 Property of $DF[m]$ operator

Although preconditioning improves the behavior of the objective function and makes its computation more efficient, and gives a computationally more accurate gradient than without precondition, we still could not give an error estimate to the gradient based on the error in solving the inner problem. We explain the reason by studying the property of $DF[m]$ operator, that is $DF[m]$ operator is of one order higher than $F[m]$ operator [Symes, 2015].

We compute

$$DF[m][dm, \delta m + r_k]$$

with $m = (2\text{km/s})^2$, dm the fast Gaussian anomaly shown in Figure 4.26 and $r_k = W \sin(kz)$, as we did for the Born modeling operator and normal operator in the Property of normal operator subsection. δm is the target reflectivity model shown in Figure 4.10b. We follow the same one shot acquisition setting as before. W is a vertical windowing operator shown in Figure 4.2. We perform the test for both bandpass filtered source and the twice integral of it.

The $DF[m][dm, \delta m + r_k]$ data with different source wavelets are shown in Figure 4.27a and Figure 4.27b. The error r_k in reflectivity model produces a visible error in data, but the main signal is still clear.

Similar output of $DF[m][dm, \delta m + r_k]$ for $k = 70$ are shown in Figure 4.28a and Figure 4.28b. Although r_k has the same L_2 norm as when $k = 30$, it produces a

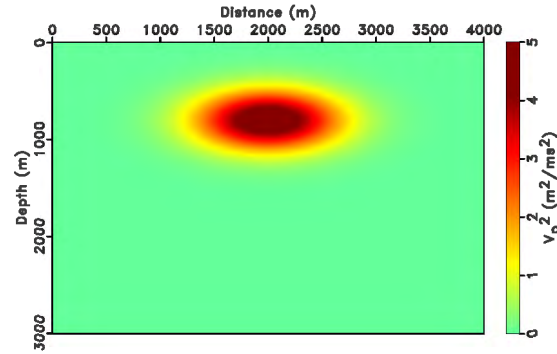


Figure 4.25

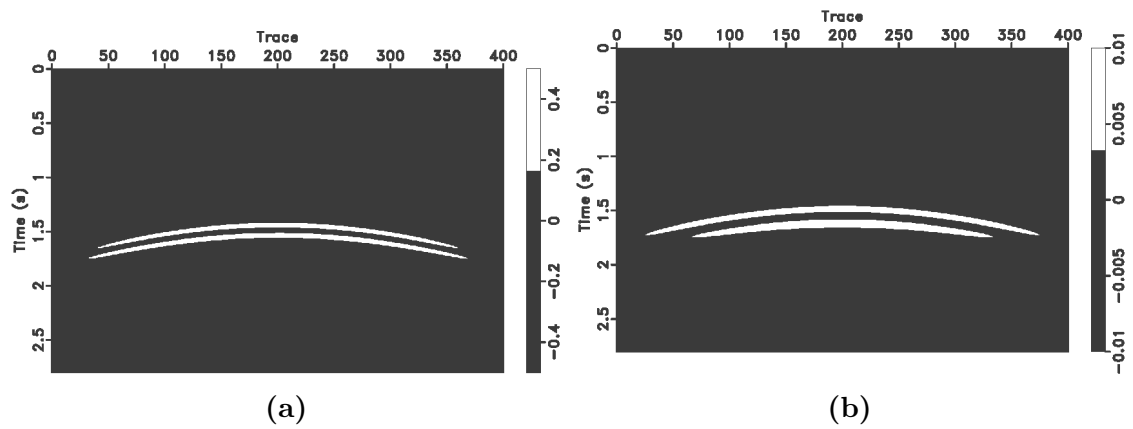
Figure 4.26: Fast Gaussian anomaly dm .

Figure 4.27: Output of $DF[m][dm, \delta m + r_k]$ with $k = 30$, $m = (2\text{km/s})^2$ and δm shown in Figure 4.10b with bandpass filtered source wavelet (a) and source wavelet integrated over time twice (b).

significant error in data. Even the main signal is hard to distinguish.

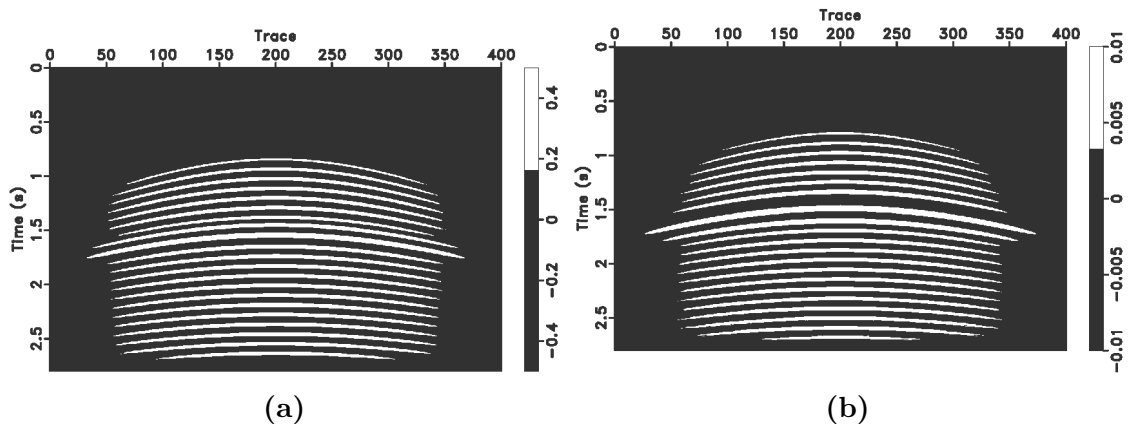


Figure 4.28: Output of $DF[m][dm, \delta m + r_k]$ with $k = 70$, $m = (2\text{km/s})^2$ and δm shown in Figure 4.10b with bandpass filtered source wavelet (a) and source wavelet integrated over time twice (b).

The relative L_2 error of $DF[m][dm, r_k]$ over $DF[m][dm, \delta m]$ for different k is shown in Figure 4.29. Theoretically, the derivative of Born modeling operator is of order 3.5 if using original source wavelet, and is of order 1.5 if composing with I_t^2 operator. The power law estimation shows that the order is 3.45 for original problem and is 1.43 if composing with I_t^2 .

The quotient of L_2 norms of $DF[m][dm, r_k]$ and $F[m]r_k$ as a function of k is shown in Figure 4.30, which shows that operator $DF[m]$ is of one order higher than $F[m]$, as explained in [Symes, 2014, ten Kroode, 2014, Symes, 2015]. Despite some errors in fitting the curve, the estimated orders are as expected.

The most important information Figure 4.30 reveals is that a small L_2 error in data misfit (achieved by solving preconditioned regularized least squares migration) does not necessarily imply a small L_2 error in the output of $DF[m][dm, \cdot]$, and thus does not imply a small error in the gradient computation, which depends on $DF^T[m]$.

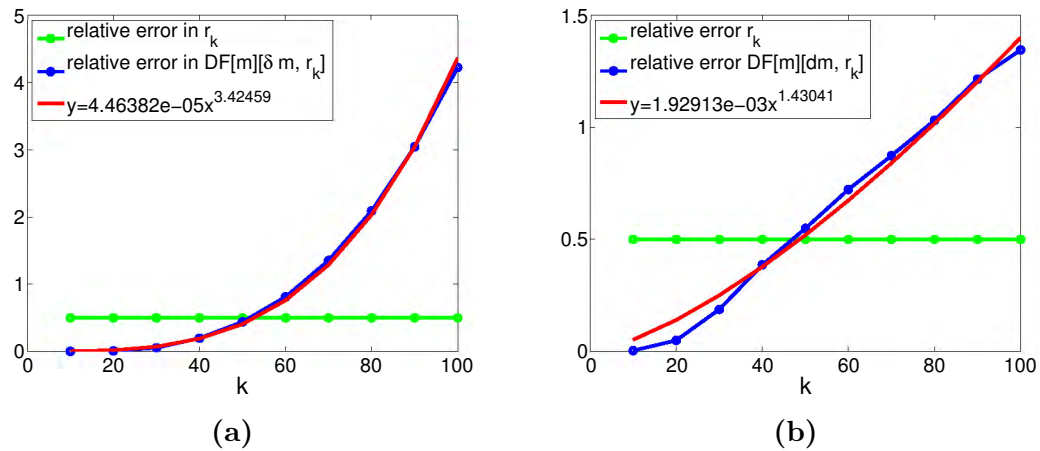


Figure 4.29: Green curve in both subfigures is the relative error $\frac{\|r_k\|}{\|\delta m\|}$. the relative error of $DF[m]$ operator $\frac{\|DF[m][dm, r_k]\|}{\|DF[m][dm, \delta m]\|}$ with $r_k = W \sin(kz)$, $m = (2\text{km/s})^2$ and dm the fast Gaussian anomaly shown in Figure 4.26 when using (a) bandpass filtered source wavelet (Figure 4.1b) and (b) twice integral of bandpass filtered source wavelet (Figure 4.8).

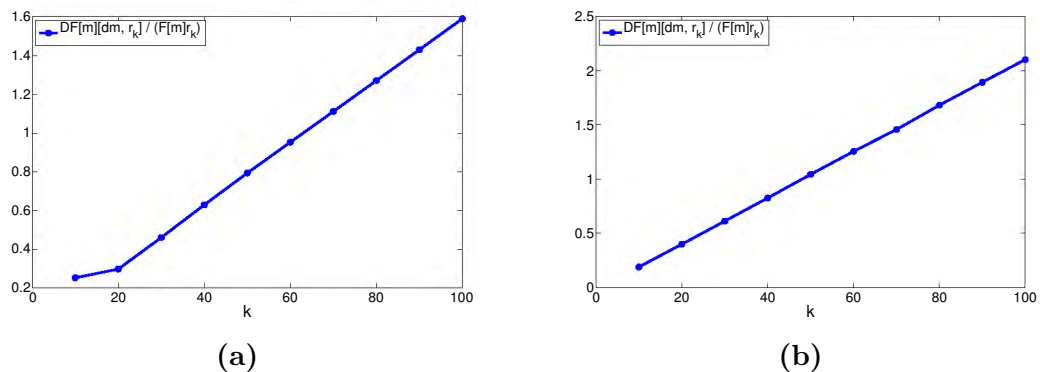


Figure 4.30: Quotient of L_2 norms of $DF[m][dm, r_k]$ and $F[m]r_k$, when using (a) bandpass filtered source wavelet (Figure 4.1b) and (b) twice integral of bandpass filtered source wavelet (Figure 4.8).

4.5 Discussion

We discussed a nested optimization scheme to solve the extended Born waveform inversion: namely the variable projection method (VPE method) and proposed a way to modify the DSO term to have the same order with the least squares Hessian. And then the optimal pseudodifferential scaling could be used to accelerate the convergence of the inner problem (FPCG algorithm) and thus reduce the over all computational cost of the VPE method.

Numerical discretization of the wave equation influences the order of the normal operator $\bar{F}[m]^T \bar{F}[m]$, as shown by numerical examples. Simple preconditioning by I_t^2 improves the condition number of the problem with free surface boundary condition and gives an order 1 normal operator.

Numerical results based on the fast Gaussian anomaly model indicates that the optimal pseudodifferential preconditioning significantly improved the accuracy and quality of the solution to the regularized LSM problem (inner problem) and generated a better shaped VPE objective function and a more accurate gradient.

However, the theoretical relation between the error in solving the inner problem and the error in the gradient computation has not been established. The reason is that the small L_2 error in the data misfit does not imply a small error in the WEMVA operator. In other words, a small error in the inverted reflectivity model could produce large error in the gradient, because the WEMVA operator ($D\bar{F}[m]^T$) is of one order higher than the RTM operator ($\bar{F}[m]^T$). How to give an theoretical estimate on the error in the gradient of VPE method is still unsolved.

Discussion

This thesis investigated methods to modify full waveform inversion to recover low frequency components of the model (background model) from seismic data, which is usually hard for FWI. We replaced the forward modeling operator in FWI with Born modeling operator based on the scale separation of the model: the model is explicitly separated as background model and reflectivity and results in a Born waveform inversion method. BWI is quadratic in reflectivity and non-linear in background model, and thus a variable projection method could be applied: minimize over the reflectivity model in an inner stage and result in an objective function that is of background model only; and then minimize over the background model in an outer stage. Variable projection method improves the iterative BWI method (as shown in Chapter 2 and [[van Leeuwen and Mulder, 2009](#), [Xu et al., 2012](#)]) to some extent. Further improvement is achieved by model extension: allow the reflectivity to depend on an extra parameter (shot coordinate in this thesis). And then add a weighted DSO penalty term to control the degree of non-physicality in the extended reflectivity. Examples in Chapter 2 shown that both model extension and variable projection are necessary for a successful Born waveform inversion.

Another conclusion we draw from Chapter 2 is that the inner problem needs

to be solved accurately enough for two reasons: first, there are models that give smaller objective function value than the target model if the inner problem is not solved accurately enough, and the more accurate the solution of the inner problem, the closer the minimum to the target model; second, the accuracy of the gradient computation relies on the accuracy of the inner problem solution. This observation motivate us to propose a method that could solve the inner problem efficiently.

We proposed a flexibly preconditioned conjugate gradient algorithm (FPCG) to improve the convergence rate of least squares migration in shot record domain (very similar to the inner problem of VPE method, by dropping the DSO term) in Chapter 3. The preconditioner is the optimal pseudo-differential scaling proposed by [Symes, 2008a, Nammour and Symes, 2009, Nammour, 2009] based on the feature of the least squares Hessian: it is a pseudo-differential operator. Numerical examples based on constant density acoustics in Chapter 3 shown that the proposed method could achieve a smaller data misfit and gradient error with a fixed computational cost than without preconditioning, and produced high quality image with more balanced amplitude and less artifacts than that produced by the non-preconditioned algorithm.

Although we used acoustic constant density wave equation in the analysis and examples, the foundations do not rely on acoustic constant density and could be extended to complex physics, such as variable density and elasticity. The resulted balanced amplitude image has potential to improve the amplitude-versus-offset (AVO) analysis [Tarantola, 1984a, Lörtzer and Berkhout, 1989, Castagna, 1993, Luh, 1993] and other velocity analysis methods, such as inversion velocity analysis [Liu et al., 2014, Symes, 2015] and VPE method discussed in Chapter 2.

In Chapter 4, we first discussed how to modify the inner problem (regularized least squares migration in shot record domain) of VPE method to use the FPCG method proposed in Chapter 3. Numerical results shown that the FPCG method

improved the convergence rate of the inner optimization problem and generated a better shaped VPE objective function and more accurate gradient computation of VPE objective with less computational cost than without preconditioning, and may potentially improve the overall convergence of VPE method. However, we still cannot get a theoretical bound of the error in gradient gradient computation. We explained the reason by analyzing the behavior of $DF[m]$ operator: it has one more order than the Born operator $F[m]$. Small L_2 errors in the reflectivity may result in large error in the gradient. Thus the error in the gradient (depends on $DF[m]^T$ in equation 4.9) is not controlled by the error in solving the inner problem (equation 4.6). Any attempts to use high frequencies to invert for the model will face this problem.

Power of Laplace operator Λ and implementation

Assume we consider the problem in 2-dimensional space. We follow the analysis in [Kern and Symes, 1994]. The anisotropic Laplace operator on $m(\mathbf{x})$ with $\mathbf{x} = (x, z)$ is defined as

$$Lm = \left(\omega_x^2 \frac{\partial^2}{\partial x^2} + \omega_z^2 \frac{\partial^2}{\partial z^2} \right) m.$$

The discretized version of operator L

$$(Lm)_{i,j} = \omega_x^2 \frac{m_{i+1,j} - 2m_{i,j} + m_{i-1,j}}{dx^2} + \omega_z^2 \frac{m_{i,j+1} - 2m_{i,j} + m_{i,j-1}}{dz^2}.$$

could be diagonalized by the discrete Fourier operator \mathcal{F} , which is unitary. That is there exists an diagonal operator Σ such that

$$\mathcal{F}^T \Sigma \mathcal{F} = L.$$

L is a uniform elliptic operator and $-L$ is a nonnegative operator. Thus the square root of $I - L$ exists and is invertible, with I being the identity operator. Define

$$\Lambda = I - L.$$

Then $\Lambda = \mathcal{F}^T(I - \Sigma)\mathcal{F}$ and $I - \Sigma$ has diagonal approximately $1 + \omega_x^2 k^2 dx^2 + \omega_z^2 l^2 dz^2$ for small wave number kdx, ldz (number of waves that exist over the distance dx for x direction). It is obvious that Λ is self-adjoint. We use FFTW software package [Frigo and Johnson, 2005,] in our program to compute the Fourier transform and its adjoint (inverse).

APPENDIX B

Computation of the gradient of J_{VPE}

For a background model m , the directional derivative of J_{VPE} for a given perturbation dm is

$$DJ_{\text{VPE}}[m]dm = D_m J_{\text{EBWI}}[m, \delta\bar{m}]dm + D_{\delta\bar{m}} J_{\text{EBWI}}[m, \delta\bar{m}]D_m \delta\bar{m}dm$$

The second term in the above equation is

$$\begin{aligned} D_{\delta\bar{m}} J_{\text{EBWI}}[m, \delta\bar{m}]D_m \delta\bar{m}dm &= \langle \bar{F}[m]D_m \delta\bar{m}dm, \bar{F}[m]\delta\bar{m} - d \rangle + \alpha^2 \langle AD_m \delta\bar{m}dm, A\delta\bar{m} \rangle \\ &= \langle D_m \delta\bar{m}dm, \bar{F}[m]^T (\bar{F}[m]\delta\bar{m} - d) \rangle + \alpha^2 \langle D_m \delta\bar{m}dm, A^T A\delta\bar{m} \rangle \\ &= \langle D_m \delta\bar{m}dm, (\bar{F}[m]^T \bar{F}[m] - \alpha^2 A^T A)\delta\bar{m} - \bar{F}^T[m]d \rangle \end{aligned}$$

Since $\delta\bar{m}$ satisfies the normal equation (2.9), the second term inside the angle brackets vanishes. Thus only the first term will contribute to the final gradient of J_{VPE} . Thus

$$DJ_{\text{VPE}}[m]dm = \langle D\bar{F}[m][\delta\bar{m}, dm], \bar{F}[m]\delta\bar{m} - d \rangle$$

Since $\bar{F}[m] = D\mathcal{F}[m]$, we have $D\bar{F}[m] = D^2\mathcal{F}[m]$, which is the Hessian of the extended forward operator $\mathcal{F}[m]$. It takes two arguments and is a bilinear operator.

Then by fixing one argument, we could define its adjoint operator $D\bar{F}[m]^T$

$$\begin{aligned} DJ_{\text{VPE}}[m]dm &= \langle dm, D\bar{F}[m]^T[\delta\bar{m}, \bar{F}[m]\delta\bar{m} - d] \rangle \\ &= \langle dm, \Lambda^{-1}D\bar{F}[m]^T[\delta\bar{m}, \bar{F}[m]\delta\bar{m} - d] \rangle_{\bar{M}} \end{aligned}$$

Thus we get the gradient

$$\nabla J_{\text{VPE}}[dm] = \Lambda^{-1}D\bar{F}[m]^T[\delta\bar{m}, \bar{F}[m]\delta\bar{m} - d].$$

References

- [Al Yahya, 1989] Al Yahya, K., 1989, Velocity analysis by iterative profile migration: *Geophysics*, **54**, 718–729. [2.4](#)
- [Aoki and Schuster, 2009] Aoki, N., and G. Schuster, 2009, Fast least-squares migration with a deblurring filter: *Geophysics*, **74**, WCA83–WCA93. [1](#), [3.2](#)
- [Bao and Symes, 1996] Bao, G., and W. Symes, 1996, Computation of pseudo-differential operators: *SIAM J. Sci. Comput.*, **17**, 416–429. [3.2](#), [3.3.1](#), [3.3.1](#), [3.3.1](#), [4.3.2](#)
- [Baysal et al., 1983] Baysal, E., D. D. Kosloff, and J. W. C. Sherwood, 1983, Reverse time migration: *Geophysics*, **48**, 1514–1524. [3.2](#)
- [Beylkin, 1985] Beylkin, G., 1985, Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized radon transform: *Journal of Mathematical Physics*, **26**, 99–108. [3.2](#), [3.3](#), [3.3.1](#), [4.3.1](#)
- [Beylkin and Burridge, 1990] Beylkin, G., and R. Burridge, 1990, Linearized inverse scattering problem of acoustics and elasticity: *Wave Motion*, **12**, 15–22. [3.5](#)
- [Biondi and Almomin, 2012] Biondi, B., and A. Almomin, 2012, Tomographic full waveform inversion (TFWI) by combining full waveform inversion with wave-equation migration velocity analysis: 82nd Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, SI9.5. [2.2](#), [2.3.3](#), [4.3.3](#)

-
- [Biondi and Almomin, 2014] ———, 2014, Simultaneous inversion of full data bandwidth by tomographic full-waveform inversion: *Geophysics*, **79**, WA129–WA140. [1](#), [2.2](#)
- [Biondi and Sava, 2004] Biondi, B., and P. Sava, 2004, Wave-equation migration velocity analysis - I: Theory, and II: Subsalt imaging examples: *Geophysics*, **52**, 593–623. [2.3.3](#), [4.3.3](#)
- [Bourgeois et al., 1989] Bourgeois, A., B. Jiang, and P. Lailly, 1989, Linearized inversion: A significant step beyond pre-stack migration: *Geophysical Journal International*, **99**, 435–445. [1](#), [3.2](#)
- [Bourgeois et al., 1991] Bourgeois, A., P. Lailly, and R. Versteeg, 1991, The Marmousi model, *in* *The Marmousi Experience: IFP/Technip*. [2.2](#), [2.4](#)
- [Boyd and Vandenberghe, 2004] Boyd, S., and L. Vandenberghe, 2004, *Convex Optimization*: Cambridge University Press. [3.3](#)
- [Brandsberg-Dahl et al., 2003] Brandsberg-Dahl, S., M. V. De Hoop, and B. Ursin, 2003, Focusing in dip and ava compensation on scattering-angle/azimuth common image gathers: *Geophysics*, **68**, 232–254. [2.3.2](#)
- [Burrige et al., 1998] Burrige, R., M. V. De Hoop, D. Miller, and C. Spencer, 1998, Multiparameter inversion in anisotropic elastic media: *Geophysical Journal International*, **134**, 757–777. [3.5](#)
- [Castagna, 1993] Castagna, J. P., 1993, AVO analysis - Tutorial and review, *in* *Offset Dependent Reflectivity - Theory and Practice of AVO Analysis*: SEG, 1, 3–36. [3.5](#), [5](#)
- [Chauris and Noble, 2001] Chauris, H., and M. Noble, 2001, Two-dimensional velocity macro model estimation from seismic reflection data by local differential

- semblance optimization: applications synthetic and real data sets: *Geophysical Journal International*, **144**, 14–26. [1](#), [2.2](#)
- [Chauris et al., 1998] Chauris, H., M. Noble, and P. Podvin, 1998, Testing the behaviour of differential semblance for velocity optimization: Presented at the 60th Conference EAGE, European Association for Geoscientists and Engineers. (Expanded Abstract). [2.2](#)
- [Chauris and Plessix, 2013] Chauris, H., and R.-E. Plessix, 2013, Differential waveform inversion - a way to cope with multiples?: Presented at the 75th Conference EAGE, European Association for Geoscientists and Engineers. (Expanded Abstract). [2.2](#)
- [Chavent and Jacewitz, 1995] Chavent, G., and C. A. Jacewitz, 1995, Determination of background velocities by multiple migration fitting: *Geophysics*, **60**, 476–490. [2.2](#)
- [Chavent and Plessix, 1999] Chavent, G., and R. Plessix, 1999, An optimal true-amplitude least-squares prestack depth-migration operator: *Geophysics*, **64**, 508–515. [1](#), [3.2](#)
- [Claerbout and Nichols, 1994] Claerbout, J. F., and D. Nichols, 1994, Spectral preconditioning: Technical Report 82, Stanford Exploration Project, Stanford University, Stanford, California, USA. [3.2](#)
- [Clément and Chavent, 1993] Clément, F., and G. Chavent, 1993, Waveform inversion through MBTT formulation, *in* *Mathematical and Numerical Aspects of Wave Propagation*: Society for Industrial and Applied Mathematics. [2.2](#)
- [Cohen, 2002] Cohen, G. C., 2002, Higher order numerical methods for transient wave equations: Springer. [4.3.1](#)

-
- [De Hoop and Bleistein, 1997] De Hoop, M. V., and N. Bleistein, 1997, Generalized Radon transform inversions for reflectivity in anisotropic elastic media: Inverse Problems, **16**, 669–690. [3.5](#)
- [Demanet et al., 2012] Demanet, L., P. Létourneau, N. Boumal, H. Calandra, J. Chiu, and S. Snelson, 2012, Matrix probing: a randomized preconditioner for the wave-equation Hessian: Applied and Computational Harmonic Analysis, **32**, 55–168. [3.2](#)
- [Dutta et al., 2014] Dutta, G., Y. Huang, W. Dai, X. Wang, and G. T. Schuster, 2014, Making the most out of the least (squares migration): 84th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 4405–4410. [1](#), [3.2](#)
- [Ehinger and Lailly, 1993] Ehinger, A., and P. Lailly, 1993, Prestack imaging by coupled linearized inversion. [2.3.3](#)
- [Frigo and Johnson, 2005] Frigo, M., and S. G. Johnson, 2005, The design and implementation of fftw3: Invited paper, Special Issue on Program Generation, Optimization, and Platform, IEEE Press, 216–231. [A](#)
- [Gauthier et al., 1986] Gauthier, O., A. Tarantola, and J. Virieux, 1986, Two-dimensional nonlinear inversion of seismic waveforms: Geophysics, **51**, 1387–1403. [2.2](#), [4.2](#)
- [Gockenbach et al., 1995] Gockenbach, M., W. Symes, and R. Tapia, 1995, The dual regularization approach to seismic velocity inversion: Inverse Problems, **11**, 501–531. [2.3.2](#)
- [Golub and Pereyra, 2003] Golub, G., and V. Pereyra, 2003, Separable nonlinear least squares: the variable projection method and its applications: Inverse Problems, **19**, R1–R26. [2.2](#), [2.3.3](#), [4.2](#)

-
- [Golub and van Loan, 1996] Golub, G. H., and C. F. van Loan, 1996, Matrix computations, 3rd ed.: Johns Hopkins University Press. [3.3.2](#)
- [Guitton and Kaelin, 2006] Guitton, A., and B. Kaelin, 2006, Least squares attenuation of reverse time migration artifacts: 76th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, SPMI 1.6. [3.2](#)
- [Herrmann et al., 2008] Herrmann, F., P. Moghaddam, and C. Stolk, 2008, Sparsity- and continuity-promoting seismic image recovery with curvelet frames: Applied and Computational Harmonic Analysis, **24**, 150–173. [3.2](#)
- [Herrmann et al., 2009] Herrmann, F. J., C. R. Brown, and Y. A. Erlangga, 2009, Curvelet-based migration preconditioning and scaling: Geophysics, **74**, A41–A46. [3.2](#)
- [Hou and Symes, 2015a] Hou, J., and W. Symes, 2015a, Accelerating extended least squares migration with weighted conjugate gradient iteration: 85th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 4243–4248. [3.2](#)
- [Hou and Symes, 2015b] ———, 2015b, An approximate inverse to the extended Born modeling operator: Geophysics, **80(6)**, R331–R349. [3.2](#), [3.4](#), [4.3.1](#)
- [Huang et al., 2016] Huang, Y., R. Nammour, and W. W. Symes, 2016, Flexibly preconditioned extended least squares migration in shot record domain: Technical report, The Rice Inversion Project, Rice University. (Submitted to Geophysics). [4.2](#)
- [Huang and Symes, 2015] Huang, Y., and W. W. Symes, 2015, Born waveform inversion via variable projection and shot record model extension: 85rd Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 1326–1331. [3.5](#), [4.2](#), [4.3](#), [4.4.2](#)

-
- [Jervis et al., 1996] Jervis, M., M. Sen, and P. Stoffa, 1996, Prestack migration velocity estimation using nonlinear methods: *J. Geophysics*, **60**, 138–150. [1](#), [2.2](#)
- [Kern and Symes, 1994] Kern, M., and W. Symes, 1994, Inversion of reflection seismograms by differential semblance analysis: Algorithm structure and synthetic examples: *Geophysical Prospecting*, **99**, 565–614. [1](#), [2.2](#), [3.3](#), [3.5](#), [4.2](#), [A](#)
- [Knyazev and Lashuk, 2006] Knyazev, A. V., and I. Lashuk, 2006, Steepest descent and conjugate gradient methods with variable preconditioning: *SIAM J. Matrix Anal. & Appl.*, **29(4)**, 1267–1280. [1](#), [3.2](#), [3.3.2](#), [4.3.2](#)
- [Lameloise et al., 2014] Lameloise, C.-A., H. Chauris, and M. Noble, 2014, Improving the gradient of the image-domain objective function using quantitative migration for a more robust migration velocity analysis: *Geophysical Prospecting*, **62**, (in press). [2.2](#)
- [Lameloise et al., 2015] ———, 2015, Improving the gradient of the image-domain objective function using quantitative migration for a more robust migration velocity analysis: *Geophysical prospecting*, **63**, 391–404. [3.2](#), [3.5](#)
- [Liu et al., 2014] Liu, Y., W. Symes, and Z. Li, 2014, Inversion velocity analysis via differential semblance optimization: Presented at the 76th Annual Meeting, Extended Abstracts, European Association of Geoscientists and Engineers. [3.5](#), [5](#)
- [Loewenthal and Mufti, 1983] Loewenthal, D., and I. Mufti, 1983, Reversed time migration in spatial frequency domain: *Geophysics*, **48**, 627–635. [3.2](#)
- [Lörtzer and Berkhout, 1989] Lörtzer, G. J. M., and A. J. Berkhout, 1989, Linear AVO inversion of multicomponent seismic data: 59th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 967–972. [3.5](#), [5](#)
- [Luh, 1993] Luh, P. C., 1993, Wavelet attenuation and AVO, *in* Offset Dependent Reflectivity - Theory and Practice of AVO Analysis: SEG, 3, 190–198. [3.5](#), [5](#)

-
- [Luo and Hale, 2014] Luo, S., and D. Hale, 2014, Least-squares migration in the presence of velocity errors: *Geophysics*, **79**, S153–S161. [1](#), [3.2](#)
- [Luo and Sava, 2011] Luo, S., and P. Sava, 2011, A deconvolution-based objective function for wave-equation inversion: 81th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 2788–2792. [2.2](#)
- [Métévier et al., 2015] Métevier, L., R. Brossier, and J. Virieux, 2015, Combining asymptotic linearized inversion and full waveform inversion: *Geophysical Journal International*, **201**, 1682–1703. [3.2](#)
- [Moore, 1920] Moore, E. H., 1920, On the reciprocal of the general algebraic matrix: *Bulletin of the American Mathematical Society*, **26**, 394–395. [3.3.1](#)
- [Mora, 1987] Mora, P., 1987, Nonlinear two-dimensional elastic inversion of multioffset seismic data: *Geophysics*, **52**, 1211–1228. [2.2](#)
- [Mulder and ten Kroode, 2002] Mulder, W., and F. ten Kroode, 2002, Automatic velocity analysis by differential semblance optimization: *Geophysics*, **67**, 1184–1191. [1](#), [2.2](#)
- [Mulder and Plessix, 2004] Mulder, W. A., and R.-E. Plessix, 2004, A comparison between one-way and two-way wave equation migration: *Geophysics*, **69**, 1491–1504. [3.2](#)
- [Nammour, 2009] Nammour, R., 2009, Approximate inverse scattering using pseudodifferential scaling: Technical Report 09-09, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, USA. [1](#), [3.2](#), [3.3.1](#), [3.3.1](#), [3.3.1](#), [3.3.1](#), [4.2](#), [4.3.2](#), [4.3.2](#), [5](#)
- [Nammour and Symes, 2009] Nammour, R., and W. Symes, 2009, Approximate constant density acoustic inverse scattering using dip-dependent scaling: 80th Annual

-
- International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, SI 3.6. [1](#), [2.5](#), [3.2](#), [3.3.1](#), [3.3.1](#), [3.3.1](#), [4.2](#), [4.3.2](#), [5](#)
- [Nammour and Symes, 2011] Nammour, R., and W. W. Symes, 2011, Multiparameter inversion: Cramer’s rule for pseudodifferential operators: International Journal of Geophysics, **2011**, 780291 (12 pp). (DOI:10.1155/2011/780291). [3.5](#)
- [Nemeth et al., 1999] Nemeth, T., C. Wu, and G. Schuster, 1999, Least-squares migration of incomplete reflection data: Geophysics, **64**, 208–221. [1](#), [3.2](#)
- [Nocedal and Wright, 1999] Nocedal, J., and S. Wright, 1999, Numerical Optimization: Springer Verlag. [2.4](#), [3.3.2](#), [3.3.2](#)
- [Nolan and Symes, 1996] Nolan, C. J., and W. Symes, 1996, Imaging and coherency in complex structure: 66th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 359–363. [3.3.1](#), [4.3.1](#)
- [Notay, 2000] Notay, Y., 2000, Flexible conjugate gradients: SIAM J. Sci. Comput., **22**(4), 1444–1460. [1](#), [3.2](#), [3.3.2](#), [4.3.2](#)
- [Pan et al., 2014] Pan, W., K. Innanen, and M. G., 2014, A comparison of different scaling methods for least-squares migration/inversion: 76th Annual International Conference and Exhibition, Expanded Abstract, European Association for Geoscientists and Engineers, WE G103 14. [3.2](#)
- [Plessix, 2000] Plessix, R.-E., 2000, Automatic cross-well tomography: an application of the differential semblance optimization to two real examples: Geophysical Prospecting, **48**, 937–951. [2.2](#)
- [Plessix, 2009] ———, 2009, Three dimensional frequency-domain full waveform inversion with an iterative solver: Geophysics, **74**, WCC149–WCC163. [1](#)

-
- [Plessix et al., 1999] Plessix, R.-E., Y.-H. de Roeck, and G. Chavent, 1999, Waveform inversion of reflection seismic data for kinematic parameters by local optimization: *SIAM Journal on Scientific Computation*, **20**, 1033–1052. [2.2](#)
- [Plessix et al., 2000] Plessix, R.-E., W. Mulder, and F. ten Kroode, 2000, Automatic cross-well tomography by semblance and differential semblance optimization: theory and gradient computation: *Geophysical Prospecting*, **48**, 913–935. [2.2](#)
- [Rakesh, 1988a] Rakesh, 1988a, A linearized inverse problem for the wave equation: *Communications on Partial Differential Equations*, **13**, 573–601. [2.3.3](#)
- [Rakesh, 1988b] ———, 1988b, A linearized inverse problem for the wave equation: *Communications on Partial Differential Equations*, **13**, 573–601. [3.2](#), [3.3.1](#), [3.3.1](#), [4.3.1](#)
- [Rickett and Sava, 2002] Rickett, J., and P. Sava, 2002, Offset and angle-domain common image-point gathers for shot profile migration: *Geophysics*, **67**, 883–889. [2.3.2](#)
- [Rickett, 2003] Rickett, J. E., 2003, Illumination-based normalization for wave-equation depth migration: *Geophysics*, **68**, 1371–1379. [3.2](#)
- [Ronen and Claerbout, 1985] Ronen, J., and J. F. Claerbout, 1985, Surface-consistent residual statics estimation by stack-power maximization: *Geophysics*, **50**, 2759–2767. [2.2](#)
- [Shen, 2004] Shen, P., 2004, Wave equation migration velocity analysis by differential semblance optimization: PhD thesis, Rice University. [1](#)
- [Shen, 2012] ———, 2012, An RTM based automatic migration velocity analysis in image domain: 82nd Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, SVE2.2. [1](#)
- [Shen and Symes, 2008] Shen, P., and W. Symes, 2008, Automatic velocity analysis via shot profile migration: *Geophysics*, **73**, VE49–60. [2.2](#)

-
- [Shin et al., 2001] Shin, C., S. Jang, and D.-J. Min, 2001, Improved amplitude preservation for prestack depth migration by inverse scattering theory: *Geophysical Prospecting*, **49**, 592–606. [3.2](#)
- [Stolk, 2000] Stolk, C. C., 2000, Microlocal analysis of a seismic linearized inverse problem: *Wave Motion*, **32**, 267–290. [3.2](#), [3.3.1](#), [3.3.1](#), [4.3.1](#)
- [Stolk et al., 2009a] Stolk, C. C., M. V. de Hoop, and T. P. M. Op 't Root, 2009a, Reverse time migration inversion from single-shot data: 79th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 2995–2999. [2.5](#), [3.2](#)
- [Stolk et al., 2009b] Stolk, C. C., M. V. de Hoop, and W. Symes, 2009b, Kinematics of shot-geophone migration: *Geophysics*, **74**, WCA18–WCA34. [3.2](#)
- [Stolk and Symes, 2003] Stolk, C. C., and W. Symes, 2003, Smooth objective functionals for seismic velocity inversion: *Inverse Problems*, **19**, 73–89. [2.2](#)
- [Stolk and Symes, 2004] ———, 2004, Kinematic artifacts in prestack depth migration: *Geophysics*, **69**, 562–575. [3.2](#), [3.3.1](#), [4.3.1](#)
- [Sun and Symes, 2012] Sun, D., and W. Symes, 2012, Waveform inversion via nonlinear differential semblance optimization: 82nd Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, SI3.3. [1](#), [2.2](#)
- [Symes, 1986] Symes, W., 1986, Stability and instability results for inverse problems in several-dimensional wave propagation, *in* Proc. 7th International Conference on Computing Methods in Applied Science and Engineering: North-Holland. [1](#), [2.2](#)
- [Symes, 1990] ———, 1990, Non-interactive estimation of the marmousi velocity model by differential semblance optimization: Technical Report 90-36, Department of Mathematical Sciences, Rice University, Houston, TX. [1](#), [2.2](#)

- [Symes, 1993] ———, 1993, A differential semblance criterion for inversion of multi-offset seismic reflection data: *Journal of Geophysical Research*, **98**, 2061–2073. [1](#), [2.2](#), [4.2](#)
- [Symes, 1998] ———, 1998, *Mathematics of reflection seismology*. (available at <http://www.trip.caam.rice.edu>). [2.3.3](#)
- [Symes, 2008a] ———, 2008a, Approximate linearized inversion by optimal scaling of prestack depth migration: *Geophysics*, **73**, R23–35. [3.2](#), [5](#)
- [Symes, 2008b] ———, 2008b, Migration velocity analysis and waveform inversion: *Geophysical Prospecting*, **56**, 765–790. [2.2](#), [3.3](#), [4.2](#)
- [Symes, 2009] ———, 2009, The seismic reflection inverse problem: *Inverse Problems*, **25**, 123008:1–24. [2.2](#), [2.3.1](#), [4.3.1](#), [4.3.3](#)
- [Symes, 2014] ———, 2014, *Seismic inverse problems: recent developments in theory and practice*: Presented at the Inverse Problems - from Theory to Application, Proceedings, Institute of Physics. [4.4.5](#)
- [Symes, 2015] ———, 2015, *Algorithmic aspects of extended waveform inversion*: Presented at the 76th Annual International Conference and Exhibition, Expanded Abstract, European Association for Geoscientists and Engineers. [3.5](#), [4.4.5](#), [4.4.5](#), [5](#)
- [Symes and Carazzone, 1991] Symes, W., and J. J. Carazzone, 1991, Velocity inversion by differential semblance optimization: *Geophysics*, **56**, 654–663. [1](#), [2.2](#), [4.2](#)
- [Symes and Santosa, 1988] Symes, W., and F. Santosa, 1988, Computation of the Newton Hessian for least-squares solution of inverse problems in reflection seismology: *Inverse Problems*, **4**, 211–233. [2.4](#)

-
- [Tang, 2009] Tang, Y., 2009, Target-oriented wave-equation least-squares migration/inversion with phase-encoded hessian: *Geophysics*, **74**, WCA95–WCA107. [1](#), [2.5](#), [3.2](#)
- [Tarantola, 1984a] Tarantola, A., 1984a, Inversion of seismic reflection data in the acoustic approximation: *Geophysics*, **49**, 1259–1266. [1](#), [2.2](#), [3.5](#), [4.2](#), [5](#)
- [Tarantola, 1984b] ———, 1984b, Linearized inversion of seismic reflection data: *Geophysical Prospecting*, **32**, 998–1015. [3.3](#)
- [Taylor, 1981] Taylor, M., 1981, *Pseudodifferential Operators*: Princeton University Press. [3.3.1](#), [3.3.1](#)
- [ten Kroode, 2014] ten Kroode, F., 2014, A Lie group associated to seismic velocity estimation: Presented at the Inverse Problems - from Theory to Application, Proceedings, Institute of Physics. [4.4.5](#)
- [ten Kroode et al., 1998] ten Kroode, F., D. Smit, and A. Verdel, 1998, A microlocal analysis of migration: *Wave Motion*, **28**, 149–172. [3.2](#), [3.3.1](#), [4.3.1](#)
- [Valenciano et al., 2015] Valenciano, A., S. Lu, N. Chemingui, and J. Yang, 2015, High resolution imaging by wave equation reflectivity inversion: Presented at the 77th Annual International Conference and Exhibition, Expanded Abstract, European Association for Geoscientists and Engineers. [1](#), [3.2](#)
- [van Leeuwen and Herrmann, 2013] van Leeuwen, T., and F. J. Herrmann, 2013, Mitigating local minima in full-waveform inversion by expanding the search space: *Geophysical Journal International*, **195**, 661–667. [2.2](#)
- [van Leeuwen and Mulder, 2009] van Leeuwen, T., and W. A. Mulder, 2009, A variable projection method for waveform inversion: Presented at the 71st Conference EAGE, European Association for Geoscientists and Engineers. (Expanded Abstract). [2.2](#), [2.3.3](#), [4.2](#), [4.3](#), [5](#)

- [Versteeg and Grau, 1991] Versteeg, R., and G. Grau, 1991, Practical aspects of inversion: The Marmousi experience: Proceedings of the EAEG, The Hague. [3.4.3](#)
- [Vigh et al., 2013] Vigh, D., K. Jiao, W. Huang, N. Moldoveanu, and J. Kapoor, 2013, Long-offset-aided Full-waveform Inversion: 75th Annual Meeting, Expanded Abstracts, European Association of Geoscientists and Engineers, 76th Annual Meeting. [2.2](#)
- [Virieux and Operto, 2009] Virieux, J., and S. Operto, 2009, An overview of full waveform inversion in exploration geophysics: *Geophysics*, **74**, WCC127–WCC152. [1](#), [2.2](#), [4.2](#)
- [Warner and Guasch, 2014a] Warner, M., and L. Guasch, 2014a, Adaptive waveform inversion-fwi without cycle skipping-theory: 75th Annual International Conference and Exhibition, Expanded Abstracts, European Association for Geoscientists and Engineers, We E106 13. ([document](#)), [1.1](#)
- [Warner and Guasch, 2014b] ———, 2014b, Adaptive waveform inversion: Theory: 84th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 1089–1093. [2.2](#)
- [Weibull and Arntsen, 2014] Weibull, W., and B. Arntsen, 2014, Anisotropic migration velocity analysis using reverse time migration: *Geophysics*, **79**, R13–R25. [2.2](#)
- [Xu et al., 2012] Xu, S., D. Wang, F. Chen, Y. Zhang, and G. Lambare, 2012, Full waveform inversion for reflected seismic data: Presented at the 74th Conference EAGE, European Association for Geoscientists and Engineers, Expanded Abstracts. [2.2](#), [2.3.3](#), [4.2](#), [4.3](#), [5](#)
- [Zhang et al., 2003] Zhang, Y., G. Zhang, and N. Bleistein, 2003, True amplitude wave equation migration arising from true amplitude one-way wave equations: *Inverse Problems*, **19**, 1113–1138. [3.2](#)