

RICE UNIVERSITY

**A Nonlinear Differential Semblance Algorithm for Waveform  
Inversion**

by

**Dong Sun**

**Ph.D. Dissertation**

APPROVED, THESIS COMMITTEE:

---

William W. Symes, Noah G. Harding Professor, Chair  
Computational and Applied Mathematics

---

Matthias Heinkenschloss, Professor  
Computational and Applied Mathematics

---

Yin Zhang, Professor  
Computational and Applied Mathematics

---

Colin A. Zelt, Professor  
Earth Science

HOUSTON, TEXAS

OCTOBER 2012



## **Abstract**

# **A Nonlinear Differential Semblance Algorithm for Waveform Inversion**

by

Dong Sun

This thesis proposes a nonlinear differential semblance approach to full waveform inversion as an alternative to standard least squares inversion, which cannot guarantee a reliable solution, because of the existence of many spurious local minima of the objective function for typical data that lacks low-frequency energy. Nonlinear differential semblance optimization combines the ability of full waveform inversion to account for nonlinear physical effects, such as multiple reflections, with the tendency of differential semblance migration velocity analysis to avoid local minima. It borrows the gather-flattening concept from migration velocity analysis, and updates the velocity by flattening primaries-only gathers obtained via nonlinear inversion. I describe a general formulation of this algorithm, its main components and implementation. Numerical experiments show for simple layered models, standard least squares inversion fails, whereas nonlinear differential semblance succeeds in constructing a kinematically correct model and fitting the data rather precisely.

# Contents

<b>Abstract</b>	<b>ii</b>
<b>Acknowledgements</b>	<b>iii</b>
<b>List of Figures</b>	<b>v</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Waveform Inversion . . . . .	4
1.2 Challenges and Modifications of the LS Inversion . . . . .	7
<b>2 Theory and Method</b>	<b>15</b>
2.1 Nonlinear Differential Semblance Optimization: Idea and Formulation	15
2.2 Nonlinear Differential Semblance Optimization for Layered Media . .	24
2.3 Nonlinear Differential Semblance Optimization: Algorithm Flow, Main Components and Implementation . . . . .	33
<b>3 Numerical Experiments</b>	<b>43</b>
3.1 Dam-like Three Layered Model with Absorbing Surface . . . . .	45
3.2 Dam-like Three Layered Model with Free Surface . . . . .	54
3.3 Step-like Three Layered Model with Free Surface . . . . .	63
3.4 Suppressing Multiple Reflections with Least-squares Inversion . . . . .	71
<b>4 Discussion and Conclusion</b>	<b>83</b>

Appendix A	87
Appendix B	91
Bibliography	97

# List of Figures

2.1	Four-Layer Velocity Model and Source Spectrum . . . . .	30
2.2	Comparison of Scans of DS and LS Objectives with Absorbing Surface	31
2.3	Scans of DS Objectives with Free Surface . . . . .	32
3.1	Dam-like Three-layer model . . . . .	45
3.2	Plane-wave Sources . . . . .	46
3.3	Final model, its gather at middle offset, and updated initial model in the 1st DS-iteration . . . . .	49
3.4	Initial model, final model and its gather at middle offset in the 6th DS-iteration . . . . .	50
3.5	Initial model, final model and its gather at middle offset in the 6th DS-iteration . . . . .	51
3.6	1D slices of all initial and final models . . . . .	52
3.7	Data fitting performance after 10 DS-iterations . . . . .	53
3.8	Final model of LS inv starting from homogeneous model . . . . .	54
3.9	Migrated and Inverted bulk-modulus in the 1st DS-iteration . . . . .	57
3.10	Initial and final models in the 3rd DS iteration . . . . .	58
3.11	1D slices of initial models $m_i^k$ ( $k = 0, 1, 2, 3$ ) . . . . .	59
3.12	Data fitting performance after 3 DS-iterations . . . . .	60
3.13	Initial and final models of the LS inversion based on the final model generated by DS inversion . . . . .	61
3.14	Data fitting comparison of two LS inversions starting from different models . . . . .	62
3.15	Step-like Three-layer model . . . . .	63

3.16	Migrated and Inverted bulk-modulus in the 1st DS-iteration . . . . .	65
3.17	Initial and final models in the 3rd DS iteration . . . . .	66
3.18	1D slices of initial models $m_l^k$ ( $k = 0, 1, 2, 3$ ) . . . . .	67
3.19	Data fitting performance after 3 DS-iterations . . . . .	68
3.20	Initial and final models of the LS inversion based on the final model generated by DS inversion . . . . .	69
3.21	Data fitting comparison of two LS inversions starting from different models . . . . .	70
3.22	Four-layer Model . . . . .	72
3.23	Gradient in 1st LS iteration and its gather at the middle offset . . . .	73
3.24	LS solution after 30 LBFGS iterations and its gather at the middle offset	74
3.25	Dipping-layer Model . . . . .	75
3.26	Gradient in 1st LS iteration and its gather at the middle offset . . . .	76
3.27	LS solution after 30 LBFGS iterations and its gather at the middle offset	77
3.28	Dome Model . . . . .	79
3.29	Gradient in 1st LS iteration and its gather at the middle offset . . . .	80
3.30	LS solution after 30 LBFGS iterations and its gather at the middle offset	81
B-1	Source and Data Spectra . . . . .	92
B-2	Four low-pass filters . . . . .	93
B-3	Spectra of data used in each run of inversion . . . . .	94
B-4	Velocity models resulting from each run of inversion . . . . .	95
B-5	Velocity models resulting from the final inversion of the continuum low-frequency inversion procedure . . . . .	95
B-6	Velocity model resulting from the one-inversion procedure . . . . .	96
B-7	Data Fitting Performance . . . . .	96

# Chapter 1

## Introduction

In a reflection seismic experiment, a controlled source (such as dynamites, air guns or seismic vibrators) initiates mechanical vibrations at points on (or near) the surface of the earth, which propagate within the earth; the mechanical response of the earth to the excitation is measured and recorded by receivers on (or near) the surface. A common objective of reflection seismology is to make inferences about physical features (*model*) of subsurface (e.g., velocity distribution, impedance profile, etc.) from data (*seismogram*) recorded on or near the surface.

In general, with reasonably idealized setting <sup>1</sup>, the laws of physics provide the governing equations for computing the data values given a model. This is called the *forward problem*.

The *inverse problem* is to reconstruct the physical properties (model) from a set of measurements (data). Usually, this problem does not have unique solutions, because: (1) the amount of data is finite and cannot carry sufficient information to

---

<sup>1</sup>A common idealized setting in reflection seismology is based on the assumption that the earth is a linearly elastic isotropic body supporting wave propagation governed by acoustic wave equations.



determine the model uniquely (underdetermination), or, (2) the data has more degrees of freedom than those of the desired model (overdetermination) and are inconsistent (because of measurement errors).

Fortunately, it is possible to construct a type of inverse through minimization of an objective function that measures the difference between two points in the data space. Thus, the inversion becomes a model-based data-fitting process that provides a “best fit” solution to the inverse problem, which is often formulated as a PDE-constrained nonlinear least-squares optimization, aiming to minimize the mean squared difference between predicted and observed data via updating the model on which the prediction is based. This nonlinear least squares approach to the inverse problem of reflection seismology has been studied extensively over 30 years. As simulation of acoustic and elastic wave-fields has become more computationally feasible, this approach has enjoyed a robust revival of interest in this decade, under the name full waveform inversion (FWI).

It has been well demonstrated that FWI is capable of reconstructing remarkably detailed models of subsurface structure (Cao et al., 1990; Bunks et al., 1995; Minkoff and Symes, 1997; Plessix et al., 1999; Shin and Min, 2006; Brenders and Pratt, 2007c,d; Vigh and Starr, 2008; Sirgue et al., 2009; Vigh et al., 2010, just to name a few); however, the direct application of FWI in reflection seismology has been strictly restricted by a fundamental obstacle, i.e., its objective functional is very ill-conditioned and has many spurious local minima for typical seismic data which lacks low-frequency energy. Because of this so-called local minima issue, FWI doesn’t work with any descent method (mandatory because of problem size) unless the initial model provides an accurate long scale estimate of the true model, see (Gauthier et al., 1986; Santosa and Symes, 1989; Tarantola et al., 1990; Symes and Carazzone, 1991;

Bunks et al., 1995; Shin and Min, 2006,for example).

The proximal cause of FWI misbehavior is the spectral incompleteness of typical field data, especially the lack of low frequencies. Low-frequency data appear to contain information about the macro trend of the true model. It has been shown that the impedance as a function of vertical travel time in a layered acoustic medium could be reconstructed from the impulse response, which contains all frequency components down to 0 Hz (Bamberger et al., 1979; Symes, 1981; Bube and Burridge, 1983; Symes, 1986; Sacks and Santosa, 1987). For several dimensional problem, numerical examples indicate that impulse responses may determine constant-density acoustic models via least-squares inversion (Bunks et al., 1995; Shin and Min, 2006).

This thesis presents a differential semblance strategy with nonlinear modeling for waveform inversion to recover the missing low-frequency information and address the local minima issue. Among all the attempts tried to address this local minima issue, differential semblance strategy is based on a modified objective functional which may avoid the non-convexity of the least-squares (LS) seismic inversion, hence lead to a well-behaved inversion. Differential semblance optimization based on linearized scattering theory (Born modeling) has been investigated by a number of authors (Symes and Carazzone, 1991; Symes, 1993, 1999; Chauris and Noble, 2001; Mulder and ten Kroode, 2002; Shen et al., 2003, 2005; De Hoop et al., 2005; Albertin et al., 2006; Shen and Symes, 2008). For these approaches, all the nonlinear effects (such as multiple reflections) must be taken off the data so that only the primary reflections remain. Obviously, whenever multiple reflections are non-neglectable, the primaries-only approximation cannot lead to plausible results. The proposed nonlinear differential semblance strategy intends to achieve two main goals: to address the local minima issues associated with the LS inversion as all the other differential semblance variants

do; and, to account in a natural way for nonlinear effects (such as multiple reflections) frequently encountered in actual data.

The thesis reviews my research, which aims to construct, to implement and to evaluate a nonlinear differential semblance algorithm, and at the same time to build some basic components towards forming a general inversion framework (Sun, 2008; Sun and Symes, 2010b,a; Symes et al., 2011). The introduction chapter is intended to provide a historical and scientific base for the work presented and put the thesis into context. The first section presents an overview for waveform inversion via least-squares optimization and its intrinsic difficulties. The second section reviews various strategies aiming to address those impediments preventing successful applications of waveform inversion. As the literature on this topic is extensive, the rest of this chapter only presents a review of selected works and motivates this work.

## 1.1 WAVEFORM INVERSION

Waveform inversion is an important model-based data-fitting approach to reflection seismology. The most familiar objective function for waveform inversion is the least squares functional measuring the mean squared difference between predicted and observed data. It is popular because: (1) it is very simple and corresponds to the maximum likelihood criterion if experimental errors have Gaussian distributions; (2) it does not require picked travel time and can take into account essentially any physics of seismic wave propagation and reconstruct detailed features of subsurface structure.

In the late 70s, Banberger, Chavent, and Lailly applied data fitting inversion to reflection seismology and presented the pioneering work (Banberger et al., 1977, 1979) on the one-dimensional model problem, which illustrates the physical and mathemat-

ical consequences of model space metric definition.

Tarantola and Valette (1982) states a general definition of the nonlinear least squares inversion, which is valid for various kinds of problems (including discrete and continuous, overdetermined and underdetermined, linear and nonlinear problems). Lines and Treitel (1984), Tarantola (1987) and Virieux and Operto (2009) provide excellent overviews of theory of least squares inversion and its applications in exploration geophysics. Here comes an abstract setting for the least-squares inverse problem over a constant density acoustics media: The model space  $\mathbf{M}$  is a set of possible velocity distributions  $m$ , and usually of rather large degrees of freedom on the order of  $10^4 - 10^6$  in 2D, one or more orders of magnitude greater for 3D; the data space  $\mathbf{D}$  consists of samples of reflection response (data) on or near the surface over a time interval;  $d_o$  stands for the observed reflection response (reflection seismogram), which are band-limited in practice for various physical limitations.  $\mathbf{D}$  is regarded as a Hilbert space with norm  $\|\cdot\|$ . The forward map  $F : \mathbf{M} \rightarrow \mathbf{D}$  is a function of the input velocity model  $m$ , denoted by  $F[m]$ , which builds a nonlinear relation between  $\mathbf{M}$  and  $\mathbf{D}$ . The simplest version of data fitting inversion is an Output Least Squares optimization:

$$\min_{m \in \mathbf{M}} J_{LS} := \frac{1}{2} \|F[m] - d_o\|^2.$$

Because the huge orders of magnitude, most attempts to minimize  $J_{LS}$  are to compute the gradient of  $J_{LS}$  with respect to  $m$  and search in the descent direction related to this gradient for an update. The gradient vanishes at a stationary point, which could be a minimum of  $J_{LS}$ . Gauss-Newton and nonlinear conjugate gradient are examples of these kinds of methods. With some version of the  $L^2$  norm in  $M$ , the gradient can be computed through standard adjoint state method and written as  $\nabla_m J_{LS} = DF[m]^T (F[m] - d_o)$ , where  $DF[m]^T$  is the adjoint of the linearized

forward map  $DF[m]$  of  $F$  at the point  $m$ . In Gauss-Newton algorithm, the searching direction can be expressed as  $(-(DF^T DF)^\dagger \nabla_m J_{LS})$ , which is the solution of the linearized least squares problem

$$\min_{\delta m} \frac{1}{2} \|DF[m]\delta m - (d_o - F[m])\|^2.$$

Lailly (1983) applies the adjoint state method to seismic inverse problem and found that  $DF^T$  is equivalent to a migration operator. The linearized inversion can be computed through conjugate gradient method. Tarantola (1984a) discusses solving the linearized problem using iterative algorithms, and showed that the rigorous solution of the linearized seismic inversion can be achieved using the classical methods of migration. As a generalization, Tarantola (1984b) develops a gradient-related iterative approach to solve the nonlinear least-squares inverse problem in the acoustic approximation for seismic reflection data with nonlinear effects (such as multiple reflection). Gauthier et al. (1986) presents the first published exploration of iterative acoustic FWI with a 2-D model and multi-offset data, and brought out some key observations on the applications of this approach, which will be reviewed in the next section.

These kinds of methods are called gradient-related iterative approaches (Nocedal and Wright, 1999), which only use local information of a current iterate  $v$  and yield local convergence. Extensive numerical studies (Cao et al., 1990; Bunks et al., 1995; Minkoff and Symes, 1997; Plessix et al., 1999; Shin and Min, 2006; Brenders and Pratt, 2007b,a; Vigh and Starr, 2008; Sirgue et al., 2009; Vigh et al., 2010) have demonstrated that waveform inversion with gradient-related approaches can reconstruct detailed models of subsurface structure, provided either very low-frequency

components (e.g., less than 1 Hz) or sufficiently good initial model whose kinematics are sufficiently close to those of the data.

On the other hand, there are some attempts to use global optimization methods to minimize  $J_{LS}$  such as genetic (Sen and Stoffa, 1991b) and simulated annealing (Sen and Stoffa, 1991a) methods. These methods use some random search strategies to traverse the model space in order to find the global minimum which corresponds to the smallest objective value. Though global methods don't need a good start model and gradient, they require a great many of evaluations of the objective function (forward problem) before they converge. Considering that a model space in reflection seismology usually has millions or even billions degrees of freedom, global methods are currently infeasible. Accordingly only iterative optimization methods with convergence rates more or less independent of model space dimension are computationally feasible, such as, gradient-related methods.

## 1.2 CHALLENGES AND MODIFICATIONS OF THE LS INVERSION

Though the LS inversion with gradient-related approaches is conceptually attractive and proved feasible, its applications in reflection seismology have been strictly restricted by two major obstacles (Symes, 2007). The first is the computational intensity of wave field modeling and various computation required by the LS inversion, especially in 3D. This computational obstacle is weakening with continuous advances in computer hardware and simulation techniques. For instance, there is an rapid increasing interests in building computational kernels via GPUs to accelerate common computations such as wave propagation and migration, etc.. On the other side, sev-

eral authors demonstrated various strategies, such as phase encoding, source batching, etc. (Krebs et al., 2009; Ali et al., 2009a,b; van Leeuwen and Herrmann, 2012, just to name a few), that significantly reduce the computational cost of waveform inversion.

The second obstacle is more fundamental. The LS objective function is very ill-conditioned and has many spurious local minima for typical seismic data which lacks low-frequency energy. Those local minima will trap any gradient-related iteration. Therefore this inversion doesn't work with any gradient-related optimization method unless the starting velocity model is so accurate that it has the same velocity trend (long scale structure) as the true velocity model. This fact is well observed and discussed in literature. Gauthier et al. (1986) shows that the LS problem is strongly nonlinear and has secondary minima, and concludes that gradient methods fail to converge to the target if the starting model does not contain the long wavelengths of the true model. Also, this paper demonstrates that iterative FWI is good for estimating short scale structure but cannot recover the long scale structure, and is easier to be successful with the presence of transmitted energy than only with the reflected energy. Santosa and Symes (1989) explores in detail the success and limitations of the LS inversion in the context of the layered velocity model. They partly released the obstacle by redefining the least-squares problem to match only the precritical part of the data. But their approach still suffered from the same impediment discussed above. Symes and Carazzone (1992) illustrates the high non-convexity of the LS objective function clearly via a plot of the mean square error over a line segment connecting constant back ground velocity with the reference velocity. I present similar plots in Chapter 2 for both the LS inversion and the proposed method. These plots demonstrate the proposed method is superior to the LS inversion for layered media.

The main factor appears to drive the above behavior of least squares inversion is the band-limitation of typical field data, especially the lack of low frequencies, which leads to the reconstruction ambiguous (Santosa and Symes, 1989). Lots of work has shown that the impedance as a function of vertical travel time in a layered acoustic medium could be reconstructed from the impulse response, which contains all frequency components down to 0 Hz, (Bamberger et al., 1979; Symes, 1981, 1986; Sacks and Santosa, 1987). For several dimensional problem, numerical examples indicate that impulse responses may determine constant-density acoustic models via the LS inversion (Bunks et al., 1995; Shin and Min, 2006). Low-frequency data appear to contain information about the trend of the true model. Only from band-limited reflection data, the nonlinear least squares inversion cannot infer the velocity trend. Appendix B demonstrates this well known fact in 1D.

Many attempts have been tried to deal with the local minima issue associated with the LS inversion.

A number of papers tried to diminish the problem of local minima by a decomposition of the seismic inversion problem by scale. Kolb et al. (1986) suggests a pre-stack continuum inversion algorithm for 1D acoustic medium. This algorithm first recovers the low-frequency trend of the velocity model via inversion of the low-frequency part of the data. Next, a progressive downward determination process is employed to infer the velocity distribution layer by layer. The numerical results demonstrate the efficiency of this continuum inversion process only for data with the very low-frequency components. For 2D pre-stack seismic inversion, Bunks et al. (1995) shows that a multi-scale approach is effective in releasing the difficulty of local minima only for data with much lower frequencies than is normally available in realistic seismic data sets. This kind of approaches inspires me with the continuum low-frequency inver-



sion strategy for data with low-frequency components down to 0 Hz, which leads to a much more efficient approach than the conventional inversion approach. I use this strategy to solve the least-squares subproblem embedded in the proposed algorithm.

Shin and Min (2006) introduces a logarithm objective function to take into account phase and amplitude separately or simultaneously, and then yield three different inversions. Some tests showed that this approach could lead to a better result than the conventional least-squares inversion for some synthetic data with very low-frequencies down to 0.3121 Hz. While the inversion results were not good for data without frequencies below 5 Hz.

Better results were then obtained with the logarithm of the Laplace transform (Shin and Cha, 2008). Numerical experiments show that starting from a rough initial model, the Laplace domain inversion could provide very smooth models that would be good starting models for standard FWI. This approach updates model via minimizing the difference between the DC components of exponentially damped seismogram and predicted data, which are zero for undamped data. Also, the logarithm difference between the DC components of a damped signal and a time-shifted version of itself actually indicates the time shift, which can be seen from a simple derivation as follows: given

$$d(t) = f(t)\chi_0(t),$$

$$u(t) = f(t - \tau)\chi_\tau(t),$$

where  $\chi_a(t)$  is a characteristic function such that

$$\chi_a(t) = \begin{cases} 0, & \text{if } t \leq a \\ 1, & \text{otherwise} \end{cases}.$$

Then

$$\bar{d}(s) = \int_0^{\infty} f(t)e^{-st} dt,$$

$$\bar{u}(s) = \int_{\tau}^{\infty} f(t - \tau)e^{-st} dt = e^{-s\tau} \bar{d}(s),$$

and

$$\left( \ln|\bar{u}(s)| - \ln|\bar{d}(s)| \right)^2 = \left( \ln \left| \frac{\bar{u}(s)}{\bar{d}(s)} \right| \right)^2 = (-s\tau)^2.$$

Give appropriate damping constants and first break times, the damping procedure helps Laplace inversion focus on the early arrivals. But it is tricky to choose appropriate damping constants and pick out the first break times to set up the exponential damping. And, all numerical examples on Laplace inversion seem to suggest the necessity of transmitted energy, which usually only presents in wide-aperture data at very large offsets. This approach shares similar underlying concept of the preconditioning strategies discussed in (Sirgue, 2003) in order to focus the inversion on the early arrivals to mitigate the nonlinearity, which requires wide-aperture data.

The starting model for FWI can also be built by the first arrival traveltimes tomography (FATT), which produces smooth models of the subsurface via nonlinear inversions of first-arrival traveltimes (Nolet, 1987; Hole, 1992; Zelt and Barton, 1998). Brenders and Pratt (2007c,d,e) show successful results for joint FATT and FWI on several blind tests at the oil-exploration scale and at the lithospheric scale, and suggest that very low frequencies and very large offsets are required to gain reliable FWI results. Also, reliable picking of first-arrival times is a difficult task when low-velocity zones exist. Based on the same principle, phase-only inversion estimates subsurface model via minimizing the phase difference of the first arrivals with a frequency-domain waveform-inversion algorithm (Min and Shin, 2006; Ellefsen, 2009). Together with phase-unwrapping strategy, phase-only inversion may release the cycle-skipping re-

striction and improve starting models for standard FWI (Shah et al., 2012b). Moreover, the phase difference of the first arrivals may be used as an indicator of the accuracy of starting models (Shah et al., 2012a). This strategy can be adopted in time-domain and worth of further investigating. For instance, a notorious difficulty needs to be addressed for phase-related strategies is how to handle noisy data.

All the above approaches aim to mitigate the obstacles of FWI by either adopting special strategies to solve LS inversion or relying on the presence of transmitted energy. But none resolve the local minima issue that has been the main impediment to full waveform inversion with reflection.

In contrast, the differential semblance approach is based on a modified least-squares principle which may avoid the non-convexity of the LS inversion, hence lead to a well-behaved inversion. Differential semblance optimization based on linearized scattering theory (Born modeling) has been investigated by a number of authors (Symes and Carazzone, 1991; Symes, 1993; Symes and Versteeg, 1993; Verm and Symes, 2006; Li and Symes, 2007; Symes, 1999; Chauris and Noble, 2001; Mulder and ten Kroode, 2002; Shen et al., 2003, 2005; De Hoop et al., 2005; Albertin et al., 2006; Shen and Symes, 2008). These work suggests that the differential semblance objective is stable against high-frequency data perturbation and essentially monomodal: the only stationary points are physically significant solutions of the waveform inversion problem. Some theoretical evidence exists that a similar algorithm based on (nonlinear) scattering might be feasible, and account in a natural way for nonlinear effects (such as multiple reflection) frequently encountered in actual data (Symes, 1991). This paper describes such a differential semblance strategy with nonlinear modeling for waveform inversion to recover the missing low-frequency information and address the local minima issue. It is an application of the *extended modeling*

concept introduced in Symes (2008).

The next chapter presents the underlying idea and a general formulation of the proposed strategy via the *extended modeling* concept, reviews a specific version of this strategy proposed in my MS thesis for 1D constant-density acoustic model, elaborates the main components and computational flow of this algorithm and briefly describes their implementation.



# Chapter 2

## Theory and Method

This chapter consists of three sections. In the first section, I present the underlying idea and general formulation of a nonlinear differential semblance algorithm. Then, the second section reviews a special case of this algorithm for plane wave propagation in layered media and demonstrates the convexity of the proposed objective via scan experiments. Finally, I elaborate the fundamental components of this algorithm and its computational flow, and briefly describes their implementations.

### **2.1 NONLINEAR DIFFERENTIAL SEMBLANCE OPTIMIZATION: IDEA AND FORMULATION**

In this section, I first review the *extended modeling* concept introduced by Symes (2008), which provides a basis for the nonlinear generalization of migration velocity analysis; then, based on this unifying concept, I formulate waveform inversion as a nonlinear differential semblance optimization problem.

## Extended Modeling

As described in chapter 1, an abstract setting for waveform inversion consists of:

- the model space  $\mathbf{M}$ , a set of possible models of Earth structure;
- the data space  $\mathbf{D}$ , a set of samples  $d$  of reflection response over a time interval;
- the forward map  $F : \mathbf{M} \rightarrow \mathbf{D}$ .

The traditional waveform inversion (least square inversion) is to minimize the mean square data misfit between the forward map output  $F[m]$  and an observed datum  $d_o \in \mathbf{D}$ , i.e.,

$$\min_{m \in \mathbf{M}} J_{LS} := \frac{1}{2} \|F[m] - d_o\|_{\mathbf{D}}^2. \quad (2.1)$$

An extension of model  $F : \mathbf{M} \rightarrow \mathbf{D}$  consists of

- an extended model space  $\overline{\mathbf{M}}$ ,
- an extension operator  $E : \mathbf{M} \rightarrow \overline{\mathbf{M}}$ ,
- an extended modeling operator  $\overline{F} : \overline{\mathbf{M}} \rightarrow \mathbf{D}$ ,

so that  $F[m] = \overline{F}[E[m]]$  for any  $m \in \mathbf{M}$ . The diagram (2.2) gives a simple illustration of this concept:

$$\begin{array}{ccc} \mathbf{M} & \xrightarrow{F} & \mathbf{D} \\ E \downarrow & \nearrow \overline{F} & \\ \overline{\mathbf{M}} & & \end{array} \quad (2.2)$$

Symes (2008) discusses two types of extensions: surface oriented extensions and depth oriented extensions. *For surface-oriented extensions, the extended model simply*

amounts to permitting the coefficients in the wave equation to depend on a surface acquisition parameter, and simulation is independent for each value of the parameter. Thus the computational complexity of the extended modeling operator is no greater than that of the basic modeling operator. For depth oriented extensions, the coefficients in the wave equation are positive definite symmetric operators. The computational complexity of modeling via time stepping is potentially enormous. This work only considers the type of surface oriented extensions. The construction in the next section makes a good example for this type of extensions.

Notice that the extension map  $E$  should be one to one, hence enable one to view the model space  $\mathbf{M}$  as a subset of the extended model space, i.e.,  $E[\mathbf{M}] \subset \overline{\mathbf{M}}$ . Often,  $E[\mathbf{M}]$  is referred to as the “physical models”, since the extended models not belonging to  $E[\mathbf{M}]$  may be in some sense unphysical. This fact will become obvious in the specific application discussed in the next section.

The extended inverse problem then becomes:

$$\text{given } d_o \in \mathbf{D}, \text{ find } \bar{m} \in \overline{\mathbf{M}} \text{ such that } \overline{F}[\bar{m}] \simeq d_o.$$

A solution  $\bar{m}$  is physically meaningful only if  $\bar{m} = E[m]$  for some  $m \in \mathbf{M}$ . In that case  $m$  becomes a solution of the original inverse problem, i.e.,  $F[m] = \overline{F}[E[m]] \simeq d_o$ . That is, *solving the original inverse problem is equivalent to find a solution to the extended inverse problem that belongs to  $E[\mathbf{M}]$ .* (Generally, “ $\simeq$ ” is in the least-squares sense.)

To turn this inversion into an optimization procedure, one needs an objective to measure the extent to which a solution to the extended inverse problem is physically meaningful. Since the range of  $E$  is a linear subspace of  $\overline{\mathbf{M}}$ , any linear operator



vanishing on this subspace gives rise to a quadratic form which can serve as such an objective. An *annihilator* of the range of  $E$  is a map  $A$  from  $\overline{\mathbf{M}}$  to some Hilbert space  $\mathbf{H}$  so that

$$\bar{m} \in E[\mathbf{M}] \iff A\bar{m} = 0.$$

Symes (2008) reviewed and compared different types of annihilators, among which differential semblance is the most appropriate one for building such an objective. This work adopts a differential semblance type of annihilator.

The original waveform inversion (2.1) has the same solution(s) as the the following constrained optimization problem, which however may have much better global behavior than waveform inversion:

$$\begin{aligned} \min_{\bar{m} \in \overline{\mathbf{M}}} \quad & J_A[\bar{m}] := \frac{1}{2} \|A\bar{m}\|_{\overline{\mathbf{M}}}^2 \\ \text{s. t.} \quad & \|\overline{F}[\bar{m}] - d\|_{\mathbf{D}}^2 \approx 0. \end{aligned} \tag{2.3}$$

Note: if there exists a model  $m \in \mathbf{M}$  with  $\|F[m] - d_o\|_{\mathbf{D}} \approx 0$ , then  $\bar{m} = Em$  is a solution to problem (2.3). Conversely, if the objective value of problem (2.3) is near zero, then there exists a model  $m \in \mathbf{M}$  with  $Em \simeq \bar{m}$ , hence  $\|F[m] - d_o\|_{\mathbf{D}} \approx 0$ . That is, the solution  $\bar{m}$  fits the data and is close to the range of  $E$  in the sense that its image under  $A$  is small.

When it comes to solve the problem (2.3), the major issue arise from the very irregular geometry of the feasible model set  $\mathcal{F} = \{\bar{m} \in \overline{\mathbf{M}} : \|\overline{F} - d\|^2 \simeq 0\}$ : how to parametrize the feasible model set  $\mathcal{F}$ ? Since the Lagrangian function of problem (2.3) is just as irregular as the least squares objective, a reparametrization is essential to turn the problem (2.3) into a smooth one amenable to Newton-like methods.

An answer to this critical question comes from the important observation that impulse responses may determine acoustic models via least-squares inversions. In the next part of this section, I present a way to reparametrizing the model space via missing low-frequency data components, and formulate the generalized waveform inversion (2.3) into a nonlinear differential semblance optimization.

## Nonlinear Differential Semblance Optimization

In this work, I consider the following extension of model:

- the model space  $\mathbf{M} := \{m(\mathbf{x})\}$ , a set of possible models of Earth structure;
- an extended model space of models depending on a surface acquisition parameter  $p$ , i.e.,  $\overline{\mathbf{M}} := \overline{m}(\mathbf{x}, p)$ ;
- a surface oriented extension operator  $E : \mathbf{M} \longrightarrow \overline{\mathbf{M}}$ ;
- the data space  $\mathbf{D} := \{d(\mathbf{x}_r, t; p)\}$ , a set of samples of reflection response over a time interval at receiver  $\mathbf{x}_r$  with surface acquisition parameter  $p$ ;
- an extended data space of data with very low-frequency components, i.e.,  $\overline{\mathbf{D}} = \mathbf{D} \oplus \mathbf{D}_l$ , where  $\mathbf{D}_l$  stands for the complimentary low-frequency data space that making up the missing low-frequency band;
- a shrink operator  $\phi : \overline{\mathbf{D}} \longrightarrow \mathbf{D}$ , i.e., band-pass filter;
- the forward map  $F : \mathbf{M} \rightarrow \mathbf{D}$  ;
- an extended modeling map  $\overline{F} : \overline{\mathbf{M}} \rightarrow \mathbf{D}$  defined as  $\overline{F}[\overline{m}](\mathbf{x}, t, p) := F[\overline{m}(\cdot, p)](\mathbf{x}, t, p)$ ;

- a forward map  $F_l : \mathbf{M} \rightarrow \mathbf{D}_l$ , defined by solving the original wave equations with a complementary low frequency source designed to make up the missing low frequency band of the original source;
- an extended modeling map  $\overline{F}_l : \overline{\mathbf{M}} \rightarrow \mathbf{D}_l$  defined as  $\overline{F}_l[\overline{m}](\mathbf{x}, t, p) := F_l[\overline{m}(\cdot, p)](\mathbf{x}, t, p)$ .

This extension obeys the fact that  $F[m] = \overline{F}[E[m]]$ ,  $F_l[m] = \overline{F}_l[E[m]]$ , and  $F[m] = \phi[(\overline{F} + \overline{F}_l)[E[m]]]$  for any  $m \in \mathbf{M}$ . The diagram (2.4) gives a simple illustration of this concept:

$$\begin{array}{ccc}
 \mathbf{M} & \xrightarrow{F} & \mathbf{D} \\
 E \downarrow & & \uparrow \phi \\
 \overline{\mathbf{M}} & \xrightarrow{\overline{F} + \overline{F}_l} & \overline{\mathbf{D}}
 \end{array} \quad . \quad (2.4)$$

In this work, a differential semblance type annihilator  $A : \overline{\mathbf{M}} \rightarrow \overline{\mathbf{M}}$  is defined as  $A\overline{m} := \frac{\partial}{\partial p}\overline{m}$ . Obviously,  $A\overline{m} = 0 \iff \overline{m} \in E[\mathbf{M}]$ .

Based on the solvability of least squares inversion for impulsive response<sup>1</sup>, the solution to the least squares problem ((2.5)) converges to a global minimizer

$$\min_{\overline{m} \in \overline{\mathbf{M}}} E[\overline{m}, d_o + d_l] := \frac{1}{2} \left\{ \left\| (\overline{F} + \overline{F}_l)[\overline{m}] - d_o - d_l \right\|^2 + \alpha^2 \|A\overline{m}\|^2 \right\}, \quad (2.5)$$

i.e., the extension  $(\overline{F} + \overline{F}_l)$  has approximate inverse operator  $\overline{G}$  in least squares sense, such that

$$\overline{G}[(\overline{F} + \overline{F}_l)[\overline{m}^*]] := \operatorname{argmin}_{\overline{m} \in \overline{\mathbf{M}}} E[\overline{m}, (\overline{F} + \overline{F}_l)[\overline{m}^*]] \quad \text{for all } \overline{m}^* \in \overline{\mathbf{M}}.$$

---

<sup>1</sup>Though no theoretical proof exists (except for 1D and layered media (Symes, 1986)), considerable numerical evidence strongly suggests the solvability of the impulsive nonlinear inversion (Bamberger et al., 1979; Symes, 1981; Sacks and Santosa, 1987; Bunks et al., 1995; Shin and Min, 2006).

In the rest of this section, I reformulate waveform inversion as two differential semblance optimization (DSO) problems respectively over  $\mathbf{D}_l$  and  $\mathbf{M}$ .

### Nonlinear DSO over $\mathbf{D}_l$

If we define  $\mathcal{A}_d : \mathbf{D}_l \rightarrow \overline{\mathbf{M}}$  as

$$\mathcal{A}[d_o + d_l] := A \overline{G}[d_o + d_l] \quad \text{for any } d_l \in \mathbf{D}_l,$$

then

$$\begin{aligned} \mathcal{A}_d[d_l] = 0 &\implies \exists m \in \mathbf{M} \text{ s.t. } \overline{G}[d_o + d_l] = E[m] \\ &\implies F[m] = \phi[\overline{F}[E[m]]] = \phi[d_o + d_l] = d_o. \end{aligned}$$

Thus, waveform inversion (2.3) is equivalent to a nonlinear differential semblance optimization (nDSO): finding  $d_l \in \mathbf{D}_l$  to minimize  $\|\mathcal{A}_d[d_o + d_l]\|$ , which can be stated in constrained form as

$$\min_{d_l \in \mathbf{D}_l} \mathcal{J}_{DS}[d_l] := \frac{1}{2} \|A\bar{m}[d_l]\|^2 \quad (2.6)$$

$$\text{s.t. } \bar{m}[d_l] = \operatorname{argmin}_{\bar{m} \in \overline{\mathbf{M}}} \frac{1}{2} \left\{ \|\overline{F}[\bar{m}] + \overline{F}_l[\bar{m}] - d_o - d_l\|^2 + \alpha^2 \|A\bar{m}\|^2 \right\}. \quad (2.7)$$

### Nonlinear DSO over $\mathbf{M}$

If we define  $\mathcal{A} : \mathbf{M} \rightarrow \overline{\mathbf{M}}$  as

$$\mathcal{A}[m_l] := A \overline{G}[F_l[m_l] + d_o] \quad \text{for any } m_l \in \mathbf{M},$$

then

$$\begin{aligned} \mathcal{A}[m_l] = 0 &\implies \exists m \in \mathbf{M} \text{ s.t. } \overline{G}[F_l[m_l] + d_o] = E[m] \\ &\implies F[m] = \phi[(\overline{F} + \overline{F}_l)[E[m]]] = \phi[F_l[m_l] + d_o] = d_o. \end{aligned}$$

Thus, waveform inversion (2.3) is equivalent to a nonlinear differential semblance optimization (nDSO): finding  $m_l \in \mathbf{M}$  to minimize  $\|\mathcal{A}[m_l]\|$ , which can be stated in constrained form as

$$\min_{m_l \in \mathbf{M}} J_{DS}[m_l] := \frac{1}{2} \|A\bar{m}[m_l]\|^2 \quad (2.8)$$

$$\begin{aligned} \text{s.t. } \bar{m}[m_l] = \operatorname{argmin}_{\bar{m} \in \overline{\mathbf{M}}} &\frac{1}{2} \left\{ \|\overline{F}[\bar{m}] + \overline{F}_l[\bar{m}] - d_o - F_l[m_l]\|^2 \right. \\ &\left. + \alpha^2 \|A\bar{m}\|^2 \right\}. \end{aligned} \quad (2.9)$$

As a summary, Form (2.6) is based on the key innovative idea of reparametrizing the control space via artificial low-frequency components that complement the missing low-frequency bands. In the next section, I will demonstrate the convexity of the objective of problem (2.6) via scan tests for plane wave propagation in layered media, which suggests that one may solve (2.6) successfully via gradient-related methods even starting from rough initial guesses. As Form (2.6) puts no constraints on  $d_l$ , in some sense this relaxes the inversion; but, there is no way to ensure low-frequency control  $d_l$  to be consistent with a physical model. To guarantee a physically meaningful solution, Form (2.8) employs extra constraints on add-in low-frequency controls. For Forms (2.6) and (2.8) share most of the main components and the key reparametrizing strategy, we expect to observe similar objective behaviors. The third section presents the gradient derivation and algorithm flow for solving problem (2.8). And, Chapter 3

demonstrate this algorithm with inversion experiments, which confirm the successful applications of gradient-related methods in solving problem (2.8) even starting from rough initial guesses.

## 2.2 NONLINEAR DIFFERENTIAL SEMBLANCE OPTIMIZATION FOR LAYERED MEDIA

This section reviews my master's work (Sun, 2008) on developing the nonlinear differential semblance optimization with Form (2.6) for plane-wave propagation in layered constant-density acoustic model and demonstrating via some primary numerical experiments the smoothness and convexity of the DS objective.

In the layered constant-density acoustic model, the wave field potential  $u(x, z, t)$  ( $x, z \in \mathbb{R}$ ) is governed by the wave equation

$$\begin{aligned} \left( \frac{1}{m^2(z)} \frac{\partial^2}{\partial t^2} - \nabla^2 \right) u(x, z, t) &= w_b(t) \delta(x, z), \\ u(x, z, t) = u_t(x, z, t) &\equiv 0, \quad t < 0, \end{aligned} \tag{2.10}$$

where  $m(z)$  is the acoustic velocity field depending only on the depth  $z$ , and the right-hand side is an isotropic point energy source with the source wavelet  $w_b(t)$ . Notice that  $w_b(t)$  is chosen to be band-limited, as is required by observations of the spectra of seismograms: for various physical limitations, real reflection seismograms don't have Fourier components at very low ( $< \omega_l$  Hz) and very high ( $> \omega_h$  Hz) temporal frequencies <sup>2</sup>.

Regarding the source (i.e.,  $w_b(t)$ ) as known, the pressure field  $\frac{\partial u}{\partial t}$ , hence the seismogram, becomes a function of the acoustic velocity:

$$p[m](x, t) := \frac{\partial u}{\partial t}(x, 0, t), \quad 0 \leq t \leq t_{max}.$$

The goal is to find  $m(z)$  for  $0 \leq z \leq z_{max}$  from the observed seismogram  $p_o$  such that

---

<sup>2</sup>The positive numbers  $\omega_l$  and  $\omega_h$  depend on specific physical settings of real experiments. For example,  $\omega_l = 5, \omega_h = 60$ .

$p[m] \simeq p_o$ .

The first step of constructing the proposed approach is the introduction of the Radon transformed field

$$U(z, p, t) = \int dx u(x, z, t + px), \quad p \in \mathbb{R}.$$

A straightforward calculation shows that the original problem becomes a set of 1-D plane-wave problems

$$\begin{aligned} \left( \frac{1}{v^2(z, p)} \frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial z^2} \right) U(z, p, t) &= w_b(t) \delta(z), \\ U(z, p, t) = U_t(z, p, t) &\equiv 0, \quad t < 0, \end{aligned} \tag{2.11}$$

for suitably small  $p \geq 0$  so that  $mp < 1$ , where the *vertical velocity*

$$v(z, p) := m(z) / \sqrt{1 - m^2(z)p^2}, \quad \text{for } p < p_{max} = 1/m_{max},$$

and  $p$  denotes the ray parameter (slowness).

The plane-wave seismogram is then defined as

$$\begin{aligned} F_{w_b}[m](p, t) &:= \frac{\partial U}{\partial t}(p, 0, t) \\ \text{for } (p, t) \in \mathbf{P} &:= \{(p, t) : |p| \leq p_{max}, 0 \leq t \leq t_{max}\}, \end{aligned} \tag{2.12}$$

which presents a forward map  $F_{w_b} : \mathbf{M} \rightarrow \mathbf{D}$ , where  $\mathbf{D}$  is the *data space*, and the *model space*  $\mathbf{M}$  denotes a set of possible velocity models, incorporating bounds on values and other regularity constraints.

Given the plane-wave seismogram  $d_o \in \mathbf{D}$  (i.e.,  $d_o = U_o$ <sup>3</sup>), this chapter focuses

---

<sup>3</sup> $U_o$  can be computed from  $u_o$  by Radon transform. To focus on the principal algorithm, I leave



on the inverse problem:

$$\begin{aligned} \text{Find} \quad & m(z) \in \mathbf{M} \\ \text{such that} \quad & F_{w_b}[m] \simeq d. \end{aligned} \tag{2.13}$$

Take for the extended model space  $\overline{\mathbf{M}}$  as the set of positive functions  $\bar{m}(z, p)$  of depth  $z$  and slowness  $p$ . The extension map  $E$  simply views a physical velocity  $m(z)$  (positive function of  $z$ ) as a function of  $z$  and  $p$ , i.e. as constant in  $p$ :  $E[m](z, p) \equiv m(z)$ . Then, the corresponding vertical velocity to  $\bar{m}(z, p)$  is

$$v(z, p) := \bar{m}(z, p) / \sqrt{1 - \bar{m}^2(z, p) p^2},$$

and the extended forward map  $\overline{F}_{w_b} : \overline{\mathbf{M}} \rightarrow \mathbf{D}$  is defined as

$$\overline{F}_{w_b}[\bar{m}](p, t) := \frac{\partial U}{\partial t}(p, 0, t) \quad \text{for all } (p, t) \in \mathbf{P}, \tag{2.14}$$

where  $U(z, p, t)$  satisfies (2.11) with  $m$  replaced by  $\bar{m}$ . Hence, the extended modeling operator  $\overline{F}_{w_b}$  satisfies the prerequisite:  $F_{w_b}[m] = \overline{F}_{w_b}[E[m]]$  for any  $m \in \mathbf{M}$ .

Notice that the extension map  $E$  is one-to-one, hence enables one to view the model space  $\mathbf{M}$  as a subset of the extended model space, i.e.,  $E[\mathbf{M}] \subset \overline{\mathbf{M}}$ . Since the extended models will be “unphysical” in the sense that  $\bar{m}(z, p) \in \overline{\mathbf{M}}$  could vary in  $p$ ,  $E[\mathbf{M}]$  consists of the “physical models”.

The extended inverse problem becomes:

$$\text{given } d_o \in \mathbf{D}, \text{ find } \bar{m}(z, p) \in \overline{\mathbf{M}} \text{ such that } \overline{F}_{w_b}[\bar{m}] \simeq d_o.$$

---

out this computation and assume that  $U_o$  is known.

A solution  $\bar{m}$  is physically meaningful only if  $\bar{m} = E[m]$  for some  $m \in \mathbf{M}$ . In that case  $m$  becomes a solution of the original inverse problem (2.13), i.e.  $F_{w_b}[m] = \bar{F}_{w_b}[\bar{m}] \simeq d_o$ . That is, *solving the inverse problem (2.13) is equivalent to find a solution to the extended inverse problem that belongs to  $E[\mathbf{M}]$* . (Generally, “ $\simeq$ ” is in the least-squares sense.)

To turn this inversion into an optimization procedure, one needs an objective to measure the extent to which a solution to the extended inverse problem is physically meaningful. Here, I choose the linear map  $A[\bar{m}] := \frac{\partial \bar{m}}{\partial p}$ , which satisfies the equivalence condition :

$$\bar{m} \in E[\mathbf{M}] \iff A[\bar{m}] = 0 \quad (\text{coherency condition}). \quad (2.15)$$

With the above notations, a differential semblance form of the inverse problem is:

$$\begin{aligned} \min_{\bar{m} \in \bar{\mathbf{M}}} \quad J_A[\bar{m}, d_o] &:= \frac{1}{2} \|A[\bar{m}]\|^2 \\ \text{such that} \quad &\|\bar{F}_{w_b}[\bar{m}] - d_o\|^2 \simeq 0. \end{aligned} \quad (2.16)$$

The major issue arise in formulating any approach to the solution of problem (2.16): *how to parametrize the feasible set  $\mathcal{F} = \left\{ \bar{m} \in \bar{\mathbf{M}} : \|\bar{F}_{w_b} - d_o\|_{\mathbf{D}}^2 \simeq 0 \right\}$ ?*

An answer to this question comes from the solvability of the impulsive inverse problem. Recall that problem (2.10) is reduced to a set of 1D plane-wave problems (2.11) via Radon Transform. The solvability of 1D impulsive inverse problems tells us that with the very-low frequency information, a 1D LS problem is solvable, i.e. the inversion could recover the long-scale structure.

For each fixed slowness  $p$ , given the source time function

$$w_b(t) = \int_{\omega_l < |\omega| \leq \omega_h} d\omega e^{2\pi i \omega t} g(\omega)$$

and the corresponding reflection response

$$d_o(p, t) = \int_{\omega_l < |\omega| \leq \omega_h} d\omega e^{2\pi i \omega t} \eta(p, \omega),$$

take

$$w(t) = w_l(t) + w_b(t),$$

where

$$w_l(t) = \int_{|\omega| \leq \omega_l} d\omega e^{2\pi i \omega t} g(\omega).$$

One can associate a vertical velocity  $v(z, p)$  for each  $p$  with  $\bar{d}(p, t)$  via the 1-D LS inversion, where

$$\bar{d}(p, t) = d_o(p, t) + d_l(p, t)$$

and

$$d_l(p, t) = \int_{|\omega| \leq \omega_l} d\omega e^{2\pi i \omega t} \eta(p, \omega).$$

Then,  $\bar{m}(z, p)$  is computed through  $\bar{m} = v / \sqrt{1 + v^2 p^2}$ .

As a summary, fixing the source wavelet  $w(t)$  with low-frequency components  $w_l(t)$  down to 0 Hz and band-limited data  $d_o(p, t)$ , the extended velocity  $\bar{m}(z, p)$  becomes a function of the very low-frequency data  $d_l(p, t)$  (or  $\eta(p, \omega)$  for  $|\omega| \leq \omega_l$ ).

Recalling Form (2.6), I state the DS problem as:

$$\begin{aligned} \min_{\substack{\eta(p,\omega) \\ (p,\omega) \in \Omega}} \quad & J_{DS} := \frac{1}{2} \|A[\bar{m}]\|^2 \\ \text{such that} \quad & \|\bar{F}_w[\bar{m}](p, t) - d_o(p, t) - d_l[\eta](p, t)\| \simeq 0, \\ & p \in [0, p_{max}] \end{aligned} \tag{2.17}$$

where  $A[\bar{m}] := \frac{\partial \bar{m}}{\partial p}$ ,  $\tilde{\mathbf{D}}_l := \{(p, \omega) : 0 \leq p \leq p_{max}, |\omega| \leq \omega_l\}$ .

Actually,  $J_{DS}$  is continuously differentiable with respect to  $\eta$ ; and, a standard adjoint state derivation leads to a gradient expression for the DS objective:

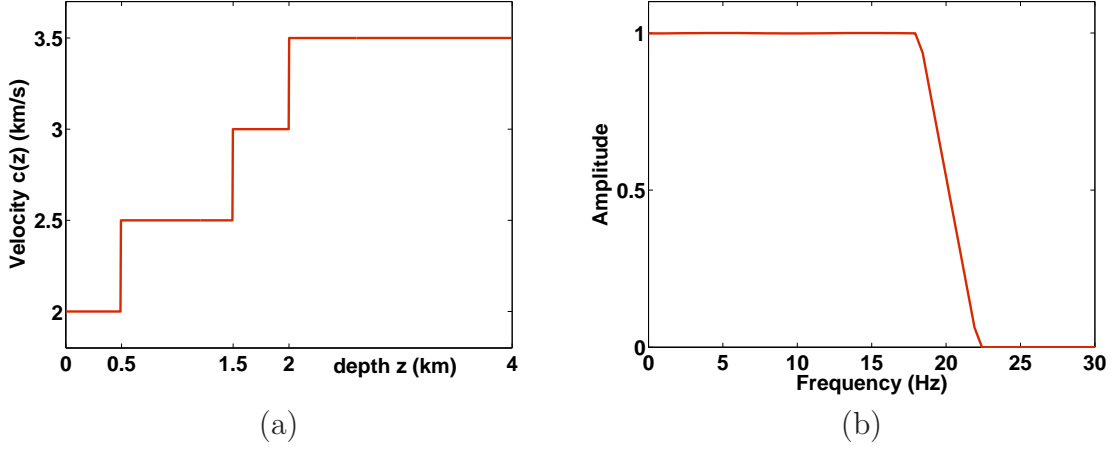
$$\nabla J_{DS} = \phi D\bar{F}_w[\bar{m}] (D\bar{F}_w[\bar{m}]^T D\bar{F}_w[\bar{m}])^\dagger \frac{\partial^2 \bar{m}}{\partial p^2},$$

where  $\phi$  is a projector from data space onto low-frequency data controls,  $D\bar{F}_w$  is Born extended modeling and  $D\bar{F}_w^T$  is its adjoint, computed by the adjoint state method. Please refer to Appendix A for a detailed derivation of this gradient computation.

## Scan Tests

In the rest of this section, I present some primary numerical experiments (“scan” tests) for a four-layer model (Figure 2.1(a)) with the fixed impulsive source time function (plotted in frequency domain in Figure 2.1(b)).

These tests illustrate the smoothness and convexity of the DS function, as well as some issues in its construction which are the subject of future research. The task is to evaluate the proposed DS objective along line segments in the space of low-frequency controls, i.e., compute the DS objective function at a series of data points



**Figure 2.1:** (a) Four-layer velocity model; (b) Normalized source wavelet in frequency domain (frequency 0 to 25  $Hz$ ).

$D_\mu$  (for some  $\mu \in [0, 1.5]$ ) defined by

$$D_\mu = \{(1 - \mu)D_{lpert}(p_i) + \mu D_{obv}(p_i)\}_{i=1}^{N_p}, \quad (2.18)$$

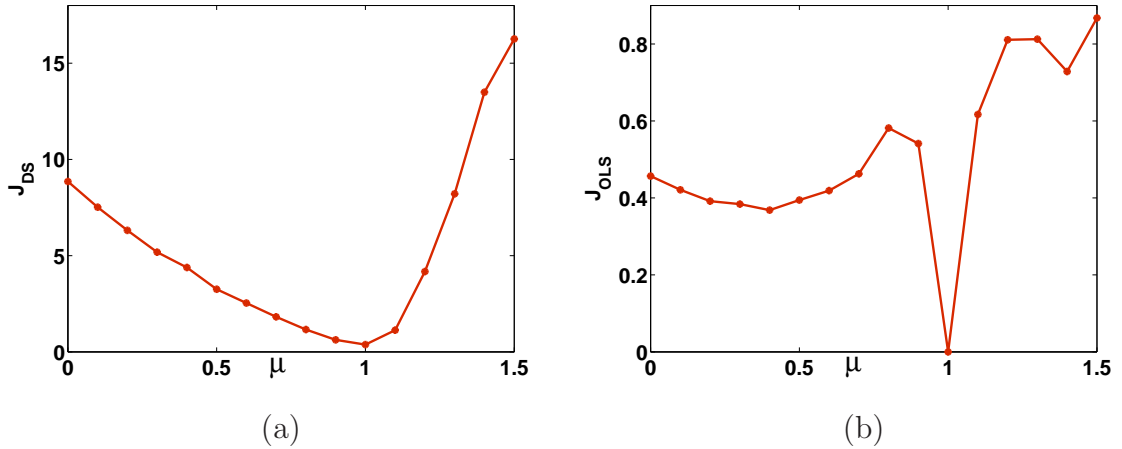
where data  $D_{lpert}(p_i)$  at slowness  $p_i$  ( $i = 1, 2, \dots, N_p$ ) differ from the observed data  $D_{obv}(p_i)$  only by their low-frequency components. The low-frequency components (0 to 5  $Hz$ ) of  $D_{lpert}$  are the corresponding low-frequency components of the seismogram derived from the homogeneous velocity model  $m_{hom}(z) = 2$ .

Choose  $N_p = 50$  and discretize the slowness field  $p$  in the way that  $p^2$  is sampled evenly.

## Experiments with absorbing boundaries

To evaluate the DS objective at a data point, I solve at each slowness the corresponding 1-D least squares problem for  $v(z, p)$ , then compute  $\bar{m}(z, p)$  from  $\bar{m} = v / \sqrt{1 + v^2 p^2}$ , and finally compute the DS objective  $J_{DS} = \frac{1}{2} \left\| \frac{\partial \bar{m}}{\partial p} \right\|^2$ . Please refer to

Appendix B for a detailed description of a frequency continuation solution to the 1-D LS inversions.



**Figure 2.2:** With absorbing surface: (a) 1-D scan through the DS objective; (b) 1-D scan through LS objective

The curve in Figure 2.2(a) interpolates samples of the DS objective at data points  $D_\mu$  defined by (2.18) with  $\mu = 0, 0.1, \dots, 1.5$ . This 1-D slice through the DS objective exhibits the smoothness (at least at the sample scale) and convexity. Also, the minimum is achieved at the data point with correct low-frequency components ( $\mu = 1$ ).

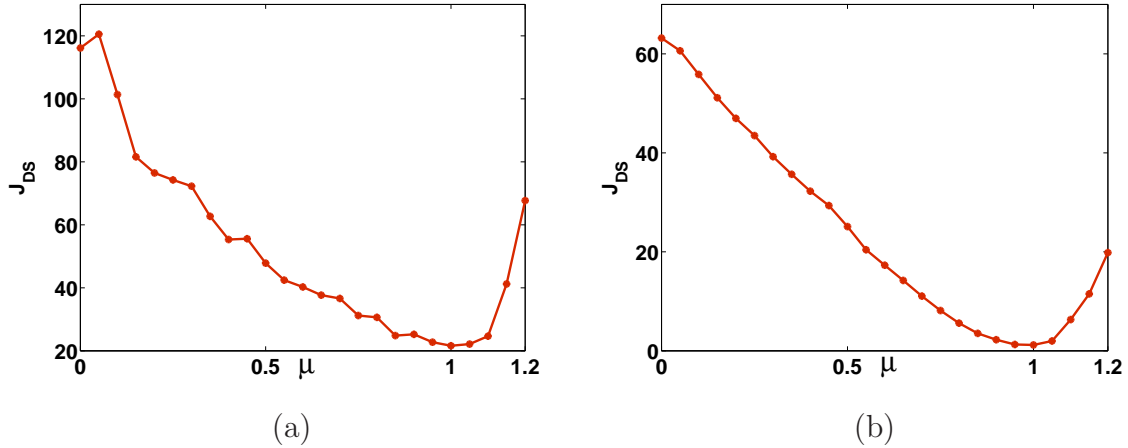
As a contrast, Figure 2.2(b) presents a similar “scan” experiment, which evaluates LS objective function at velocity models  $m_\mu$  chosen as

$$m_\mu(z) = (1 - \mu)m_{hom} + \mu m^*(z)$$

with  $\mu = 0.0, 0.1, \dots, 1.5$ , where  $m_{home}$  and  $m^*$  stand for the homogeneous and true models respectively. This 1-D scan clearly demonstrates the multi-modality of LS objective, which badly jeopardizes the application of gradient-related methods.

### Experiments with free surface

An important objective of the proposed algorithm is to account for nonlinear effects of wave propagation such as multiple reflections. Hence, it is desired to know how this DS objective behaves for problems with free surface, which is an important cause of multiple reflections. In the following test, the free surface boundary condition is adopted.



**Figure 2.3:** With free surface: (a) 1-D scan through the DS objective without post-processing; (b) 1-D scan through the DS objective with  $\bar{m}(z, p)$  smoothed in  $p$

The curves in Figure 2.3(a) samples the DS objective at data points  $D_\mu$  with  $\mu = 0.0, 0.05, \dots, 1.2$ . Though this scan exhibits the convexity near the minimum, it appears to be flat near  $\mu = 1$  and possess some bumps. This adverse behavior may come from the numerical errors accumulated during all the approximating computations. Especially, since the 1D LS inversions are done independently and yield different accuracy, the extended models  $\bar{m}(z, p)$  become inconsistent in slowness, which leads to the noisy behavior of the DS objective.

To improve the behavior of the DS objective, one can adopt a number of strategies to reduce numerical errors, such as using smaller tolerance for 1-D inversions,

choosing different expressions of DS objective, and employing some regularization techniques to smooth  $\bar{m}(z, p)$  in  $p$  and  $z$ , etc.. I have considered some of these strategies. Figure 2.3(b) presents the same scan of the DS objective except that  $\bar{m}(z, p)$  is smoothed in  $p$  via minimizing the Total Variation of  $\bar{m}(z, p)$  with respect to  $p$  at each  $z$ . Now, the scan exhibits the desired smoothness and convexity.

This section briefly reviews the first step of my work towards developing a nonlinear differential semblance approach to waveform inversion. Numerical examples show that this nonlinear DS objective is convex and achieves an extremum at the target model. Thus, gradient-related methods seem promising to solve the proposed DSO problem. In the next section, I start to discuss the algorithm flow and main components of this nonlinear differential semblance strategy based on a general formulation (2.8).

## 2.3 NONLINEAR DIFFERENTIAL SEMBLANCE OPTIMIZATION: ALGORITHM FLOW, MAIN COMPONENTS AND IMPLEMENTATION

Recall that waveform inversion can be formulate as a nonlinear differential semblance optimization problem (2.8), i.e.,

$$\begin{aligned} \min_{m_l \in \mathbf{M}} J_{DS} [m_l] &:= \frac{1}{2} \|A\bar{m}[m_l]\|^2 \\ \text{s. t. } \bar{m}[m_l] &= \operatorname{argmin}_{\bar{m} \in \bar{\mathbf{M}}} \frac{1}{2} \left\{ \left\| \bar{F}[\bar{m}] + \bar{F}_l[\bar{m}] - d_o - F_l[m_l] \right\|^2 \right. \\ &\quad \left. + \alpha^2 \|A\bar{m}\|^2 \right\}. \end{aligned}$$



This section describes the algorithm flow of this approach and elaborates its main components, i.e., the extended sub least squares inversion, the gradient computation for the nonlinear DS objective, and the step-length calculation.

I will consider the above nonlinear DSO problem in inner-product spaces. The numerical solution is the discrete approximation of the continuum problem. Roughly speaking, the inner products involved in this problem are

$$\begin{aligned} \langle m_1, m_2 \rangle_{\mathbf{M}} &:= \int_{\Omega} m_1(\mathbf{x}) m_2(\mathbf{x}) \, \mathbf{d}\mathbf{x}, \\ \langle \bar{m}_1, \bar{m}_2 \rangle_{\bar{\mathbf{M}}} &:= \int_{\Theta} \int_{\Omega} \bar{m}_1(\mathbf{x}, p) \bar{m}_2(\mathbf{x}, p) \, \mathbf{d}\mathbf{x} dp = \int_{\Theta} \langle \bar{m}_1(\cdot, p), \bar{m}_2(\cdot, p) \rangle_{\mathbf{M}} dp, \\ \langle d_1, d_2 \rangle &:= \int_T \int_{\Omega_R} d_1(\mathbf{x}, t) d_2(\mathbf{x}, t) \, \mathbf{d}\mathbf{x} dt \\ \langle \bar{d}_1, \bar{d}_2 \rangle_{\bar{\mathbf{D}}} &:= \int_{\Theta} \int_T \int_{\Omega_R} \bar{d}_1(\mathbf{x}, t, p) \bar{d}_2(\mathbf{x}, t, p) \, \mathbf{d}\mathbf{x} dt dp = \int_{\Theta} \langle \bar{d}_1(\cdot, \cdot, p), \bar{d}_2(\cdot, \cdot, p) \rangle dp, \end{aligned}$$

where  $\Theta$  denotes the set of acquisition parameters chosen to extend the model space,  $\Omega \subset \mathbb{R}^{\dim}$  stands for the spacial region under concern,  $\Omega_R \subset \Omega$  denotes the set of sampling point in space, and  $T$  denotes the recording time range.

Assume all derivatives exist for the following derivation.

## Sub-LS problem

The constraint of problem (2.8) is fulfilled via solving the sub-LS problem for some fixed  $m_i \in \mathbf{M}$ :

$$\min_{\bar{m} \in \bar{\mathbf{M}}} J_{LS}[\bar{m}] := \frac{1}{2} \left\{ \left\| (\bar{F} + \bar{F}_l)[\bar{m}] - \bar{d} \right\|^2 + \alpha^2 \|A\bar{m}\|^2 + \mathcal{R}(\bar{m}) \right\}, \quad (2.19)$$

where  $\bar{d} = d_o + F_l[m_l]$ ,  $\mathcal{R}(m)$  stands for some regularization term (e.g., for Tikohonov term  $\mathcal{R}(\bar{m}) := \beta^2 \|\bar{m}\|^2$ ). Often, (2.19) is ill conditioned and one might have to use regularization term(s) to enforce available prior information, which I omit in the following derivation for simplicity.

Since

$$J_{LS}[\bar{m}] = \frac{1}{2} \int_{\Theta} \langle (F + F_l)[\bar{m}(\cdot, p)] - \bar{d}(\cdot, \cdot, p), (F + F_l)[\bar{m}(\cdot, p)] - \bar{d}(\cdot, \cdot, p) \rangle dp \\ + \frac{1}{2} \alpha^2 \langle A\bar{m}, A\bar{m} \rangle_{\overline{\mathbf{M}}}$$

and

$$(F + F_l)[\bar{m}(\cdot, p) + \delta\bar{m}(\cdot, p)] \simeq (F + F_l)[\bar{m}(\cdot, p)] + D(F + F_l)[\bar{m}(\cdot, p)] \delta\bar{m}(\cdot, p) \\ + \frac{1}{2} D^2(F + F_l)[\bar{m}(\cdot, p)] \cdot (\delta\bar{m}(\cdot, p), \delta\bar{m}(\cdot, p)) \\ = (F + F_l)[\bar{m}(\cdot, p)] + (DF + DF_l)[\bar{m}(\cdot, p)] \delta\bar{m}(\cdot, p) \\ + \frac{1}{2} (D^2F + D^2F_l)[\bar{m}(\cdot, p)] \cdot (\delta\bar{m}(\cdot, p), \delta\bar{m}(\cdot, p)),$$

then

$$DJ_{LS}[\bar{m}] \delta\bar{m} = \int_{\Theta} \langle (DF + DF_l)[\bar{m}(\cdot, p)]^T ((F + F_l)[\bar{m}(\cdot, p)] - \bar{d}(\cdot, \cdot, p)), \delta\bar{m}(\cdot, p) \rangle_{\mathbf{M}} dp \\ + \alpha^2 \langle A^T A\bar{m}, \delta\bar{m} \rangle_{\overline{\mathbf{M}}},$$

(2.20)

i.e.,

$$\begin{aligned} \nabla J_{LS} [\bar{m}] (\cdot, p) &= (DF + DF_l) [\bar{m}(\cdot, p)]^T ((F + F_l) [\bar{m}(\cdot, p)] - \bar{d}(\cdot, \cdot, p)) \\ &\quad + \alpha^2 A^T A \bar{m}(\cdot, p) \end{aligned} \quad (2.21)$$

And, the first order necessary condition for the sub-LS problem (2.19) is:

$$(DF + DF_l) [\bar{m}(\cdot, p)]^T ((F + F_l) [\bar{m}(\cdot, p)] - \bar{d}(\cdot, \cdot, p)) + \alpha^2 A^T A \bar{m}(\cdot, p) = 0, \quad \text{for any } p \in \Theta. \quad (2.22)$$

## Computation of DS Gradient and Initial Step-length

Here, I sketch a brief derivation of the gradient of  $J_{DS}[m_l]$  respect to  $m_l$ .

Define

$$\mathcal{J} [\bar{d}] := \frac{1}{2} \|A \bar{m} [\bar{d}]\|_{\bar{\mathbf{M}}}^2,$$

where  $\bar{m} [\bar{d}]$  solves the following extended least squares problem (2.19) for fixed  $\bar{d} \in \bar{\mathbf{D}}$ , i.e.,

$$\bar{m} [\bar{d}] = \operatorname{argmin}_{\bar{m} \in \bar{\mathbf{M}}} \frac{1}{2} \left\{ \|(\bar{F} + \bar{F}_l) [\bar{m}] - \bar{d}\|^2 + \alpha^2 \|A \bar{m}\|^2 \right\}.$$

Then,

$$J_{DS} [m_l] = \mathcal{J} [d_o + F_l [m_l]], \quad (2.23)$$

and by the chain rule

$$\nabla_{m_i} J_{DS} [m_i] = DF_l [m_i]^T \nabla \mathcal{J}_{\bar{d}} [d_o + F_l [m_i]]. \quad (2.24)$$

Let  $\delta \bar{m} [\bar{d}, \delta \bar{d}] := \bar{m} [\bar{d} + \delta \bar{d}] - \bar{m} [\bar{d}]$ . For simplicity, denote  $\delta \bar{m} [\bar{d}, \delta \bar{d}]$ ,  $\bar{m} [\bar{d}]$  by  $\delta \bar{m}$  and  $\bar{m}$  respectively. Then,

$$\begin{aligned} J [\bar{d} + \delta \bar{d}] - J [\bar{d}] &= \langle A \bar{m}, A \delta \bar{m} \rangle_{\overline{\mathbf{M}}} + \frac{1}{2} \langle A \delta \bar{m}, A \delta \bar{m} \rangle_{\overline{\mathbf{M}}} \\ &= \langle A^T A \bar{m}, \delta \bar{m} \rangle_{\overline{\mathbf{M}}} + \frac{1}{2} \langle A^T A \delta \bar{m}, \delta \bar{m} \rangle_{\overline{\mathbf{M}}} \\ &= \int_{\Theta} \langle A^T A \bar{m}(\cdot, p), \delta \bar{m}(\cdot, p) \rangle_{\mathbf{M}} dp \\ &\quad + \frac{1}{2} \int_{\Theta} \langle A^T A \delta \bar{m}(\cdot, p), \delta \bar{m}(\cdot, p) \rangle_{\mathbf{M}} dp. \end{aligned} \quad (2.25)$$

The first order necessary condition (2.22) tells us that

$$(DF + DF_l) [\bar{m}(\cdot, p)]^T ((F + F_l) [\bar{m}(\cdot, p)] - \bar{d}(\cdot, \cdot, p)) + \alpha^2 A^T A \bar{m}(\cdot, p) = 0, \text{ for any } p \in \Theta.$$

which leads to: for any  $p \in \Theta$

$$\begin{aligned} (DF + DF_l) [\bar{m}(\cdot, p) + \delta \bar{m}(\cdot, p)]^T ((F + F_l) [\bar{m}(\cdot, p) + \delta \bar{m}(\cdot, p)] - \bar{d}(\cdot, \cdot, p) - \delta \bar{d}(\cdot, \cdot, p)) \\ + \alpha^2 A^T A \delta \bar{m}(\cdot, p) = 0 \end{aligned}$$

$$\begin{aligned}
\implies 0 &\simeq (DF + DF_l) [\bar{m}(\cdot, p)]^T ((F + F_l) [\bar{m}(\cdot, p)] - \bar{d}(\cdot, \cdot, p)) \\
&\quad + (DF + DF_l) [\bar{m}(\cdot, p)]^T ((DF + DF_l) [\bar{m}(\cdot, p)] \delta\bar{m}(\cdot, p) - \delta\bar{d}(\cdot, \cdot, p)) \\
&\quad + ((D^2F + D^2F_l) [\bar{m}(\cdot, p)] \delta\bar{m})^T ((F + F_l) [\bar{m}(\cdot, p)] - \bar{d}(\cdot, \cdot, p)) \\
&\quad + ((D^2F + D^2F_l) [\bar{m}(\cdot, p)] \delta\bar{m})^T ((DF + DF_l) [\bar{m}(\cdot, p)] \delta\bar{m}(\cdot, p) - \delta\bar{d}(\cdot, \cdot, p)) \\
&\quad + \alpha^2 A^T A \delta\bar{m}(\cdot, p)
\end{aligned}$$

Based on first order perturbation theory and the fact that  $\bar{m}$  solves (2.19), I have

$$H [\bar{m}(\cdot, p)] \delta\bar{m}(\cdot, p) \simeq (DF + DF_l) [\bar{m}(\cdot, p)]^T \delta\bar{d}(\cdot, \cdot, p), \quad (2.26)$$

where

$$H [\bar{m}(\cdot, p)] = (DF + DF_l) [\bar{m}(\cdot, p)]^T (DF + DF_l) [\bar{m}(\cdot, p)] + \alpha^2 A^T A.$$

Hence,

$$\delta\bar{m}(\cdot, p) \simeq H [\bar{m}(\cdot, p)]^{-1} (DF + DF_l) [\bar{m}(\cdot, p)]^T \delta\bar{d}(\cdot, \cdot, p). \quad (2.27)$$

Plugging (2.27) into (2.25) and neglecting higher order perturbations, I have

$$\begin{aligned}
&J [\bar{d} + \delta\bar{d}] - J [\bar{d}] \\
&\simeq \int_{\Theta} \langle (DF + DF_l) [\bar{m}(\cdot, p)] H [\bar{m}(\cdot, p)]^{-1} A^T A \bar{m}(\cdot, p), \delta\bar{d}(\cdot, \cdot, p) \rangle dp.
\end{aligned} \quad (2.28)$$

Then,

$$\nabla J [\bar{d}] (\cdot, \cdot, p) \simeq (DF + DF_l) [\bar{m}(\cdot, p)] H [\bar{m}(\cdot, p)]^{-1} A^T A \bar{m}(\cdot, p). \quad (2.29)$$

Thus, with (2.24), the gradient of the DS objective of problem (2.8) can be approximately written as

$$\nabla_{m_l} J_{DS}[m_l] \simeq DF_l[m_l]^T (D\bar{F} + D\bar{F}_l) [\bar{m}[m_l]] \mathcal{H}[m_l]^{-1} A^T A \bar{m}[m_l], \quad (2.30)$$

where  $\mathcal{H}[m_l] = H[\bar{m}[d_o + F_l[m_l]]]$ , and  $\bar{m}[m_l]$  is a simplified representation of  $\bar{m}[d_o + F_l[m_l]]$ .

Given an updating direction  $\phi$ ,

$$\begin{aligned} J_{DS}[m_l + \xi\phi] &= \frac{1}{2} \langle A \bar{m}[m_l + \xi\phi], A \bar{m}[m_l + \xi\phi] \rangle_{\bar{\mathbf{M}}} \\ &\approx \frac{1}{2} \langle A \bar{m}[m_l] + \xi D_{m_l} \bar{m}[m_l] \phi, A \bar{m}[m_l] + \xi D_{m_l} \bar{m}[m_l] \phi \rangle_{\bar{\mathbf{M}}} \\ &= J_{DS}[m_l] + \xi \langle A D_{m_l} \bar{m}[m_l] \phi, A \bar{m}[m_l] \rangle_{\bar{\mathbf{M}}} + \frac{1}{2} \xi^2 \langle A D_{m_l} \bar{m}[m_l] \phi, A D_{m_l} \bar{m}[m_l] \phi \rangle_{\bar{\mathbf{M}}}. \end{aligned}$$

Let

$$Q(\xi) := J_{DS}[m_l] + \xi \langle \psi, A \bar{m}[m_l] \rangle_{\bar{\mathbf{M}}} + \frac{1}{2} \xi^2 \langle \psi, \psi \rangle_{\bar{\mathbf{M}}}$$

, where

$$\psi = A D_{m_l} \bar{m}[m_l] \phi = A \mathcal{H}[m_l]^{-1} (D\bar{F} + D\bar{F}_l) [\bar{m}[m_l]]^T DF_l[m_l] \phi.$$

The ‘optimal’ step length  $\xi^*$  that minimizes  $Q(\xi)$  is

$$\xi^* = -\frac{\langle \psi, A \bar{m}[m_l] \rangle_{\bar{\mathbf{M}}}}{\langle \psi, \psi \rangle_{\bar{\mathbf{M}}}},$$

which satisfies

$$\frac{dQ}{d\xi}(\xi^*) = 0.$$

In this work,  $\xi^*$  is computed in each DS iteration as the initial step length for updating

$m_l$ .

## Algorithm Flow and Implementation

The proposed nonlinear differential semblance optimization (nDSO) proceeds as:

Nonlinear DSO:

- Initialization: set  $\bar{m}^0(\mathbf{x}, \cdot) = m_l^0(\mathbf{x})$ ,  $\epsilon_1$ ,  $\epsilon_2$ , etc.
- DS Inversion Loop: for  $k = 0, 1, 2, \dots$ 
  1. Solve the extended sub-LS problem (2.19) for  $\bar{m}^k = \bar{m} [m_l^k]$  with  $m_l^k$  as starting model;
  2. Compute  $J_{DS}^k = J_{DS} [\bar{m}^k]$ .  
If  $J_{DS}^k \leq \epsilon_1 J_{DS}^0$ , stop; else, continue;
  3. Perform gradient computation and get an approximate gradient  $\phi_k$  as the updating direction. If  $\|\phi_k\| \leq \epsilon_2 \|\phi_0\|$ , stop; else, continue;
  4. Compute step length  $\beta_k$  and let  $m_l^{k+1} = m_l^k + \beta_k \phi_k$ .

The above procedure mainly consists of two computational blocks:

1. Evaluation of  $J_{DS}[m_l]$ .

This computation requires solution of the extended sub-LS problem (2.19) for  $\bar{m}[m_l]$ , and then computation of the value  $J_{DS}[m_l] = \frac{1}{2} \|A\bar{m}[m_l]\|^2$ , where  $A := \frac{\partial \bar{m}[m_l]}{\partial p}$  in this work. Since this extended sub-LS problem is to fit data  $(d_o + F_l[m_l])$ , which contains the very low-frequency components down to 0 Hz, one can use a homogeneous starting model and adopt the standard frequency continuation strategy to solve the problem via a sequence of LS inversions for

data with increasing frequency bands. Also, one can use  $m_l$  as the starting model and solve this problem with one least-squares inversion. Based on my tests, both approaches lead to kinematically same results, due to the fact that  $F_l[m_l]$  is computed based on  $m_l$  and thus has kinematics consistent with it. For the sake of efficiency, the second approach is adopted in all the numerical experiments shown later.

2. Calculation of the gradient  $\nabla J_{DS}[m_l]$ . Given  $\bar{m}[m_l] = \operatorname{argmin}_{\bar{m} \in \bar{\mathcal{M}}} J_{LS}[\bar{m}]$ , computation of the gradient  $\nabla J_{DS}[m_l]$  requires to:

- (a) compute  $\mathbf{q}^1 = A^T A \bar{m}$ ;
- (b) solve a normal equation  $\mathcal{H}[m_l] \mathbf{q}^2 = \mathbf{q}^1$ ;
- (c) compute  $\nabla J_{DS}[m_l] \approx DF[m_l]^T (D\bar{F} + D\bar{F}_l) [\bar{m}[m_l]] \mathbf{q}^2$ .

Clearly, solving the normal equation

$$\mathcal{H}[m_l] \mathbf{q}^2 = \mathbf{q}^1 \tag{2.31}$$

takes the most amount of computing effort. An efficient preconditioner would be of great value.





# Chapter 3

## Numerical Experiments

This chapter consists of four sections. The first three sections present three numerical experiments to demonstrate nonlinear differential semblance inversion and make comparisons with the traditional least squares approach. The last section presents three extended LS inversions on different models with free surface to elaborate the multiple suppression phenomena during LS inversion.

All the numerical tests in this chapter adopt the following two dimensional acoustic model

$$\frac{1}{\kappa} \frac{\partial u}{\partial t} + \nabla \cdot \mathbf{v} = f(x, z, t), \quad (3.1)$$

$$\frac{1}{b(x, z)} \frac{\partial \mathbf{v}}{\partial t} + \nabla u = 0, \quad (3.2)$$

$$u \equiv 0, \mathbf{v} \equiv \mathbf{0}, \quad t < 0,$$

where  $u(x, z, t)$ ,  $\mathbf{v}(x, z, t)$ ,  $b(x, z)$ ,  $\kappa(x, z)$ ,  $f(x, z, t)$  respectively stand for pressure, particle velocity, buoyancy, bulk modulus, and plane wave source. Let  $m := (\kappa, b)$  denote

the model (parameter vector).

The above equation system defines the forward map

$$F_f[m] := Su, \quad (3.3)$$

where  $S$  is a sampling operator, such as  $Su := \{u(x_r, z_r, t)\}$  in which  $(x_r, z_r)$  denotes the coordinates of selected receivers.

This chapter adopts two types of plane wave source functions  $f_l$  and  $f_h$ , both of which consist of thirty one plane-wave shots (shown in Figure 3.2) whose slownesses are chosen so that their square values are evenly spaced.  $f_h$  is the band-limited source used to define  $F$  (i.e.,  $F[m] := F_{f_h}[m]$ ) and generate the target data, and its spectrum is like a band-pass filter with 5 Hz and 45 Hz respectively as the lower and upper cutoff frequencies.  $f_l$  is the complementary low-frequency source used to define  $F_l$  (i.e.,  $F_l[m] := F_{f_l}[m]$ ) and generate low-frequency data add-in for the extended least squares inversion. The spectrum of  $f_l$  is like a low-pass filter with 6 Hz as the upper cutoff frequency.

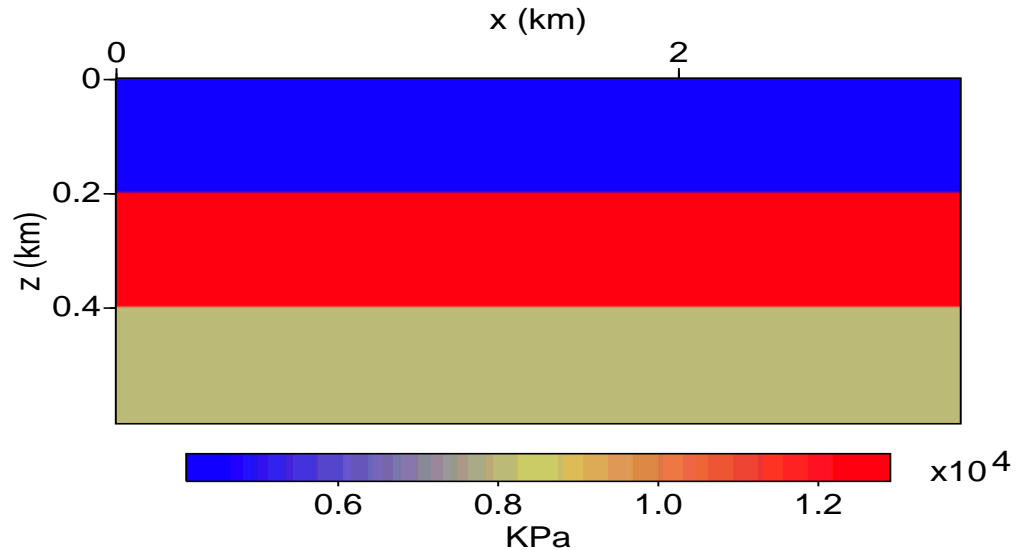
All the numerical experiments in this study were conducted with the inversion package IWAVE++, a collection of C++ classes expressing a wrapper of IWAVE<sup>1</sup>, providing the fundamental computations (simulation, Born simulation and its adjoint action), and accommodating both standard and extended inversions. Please refer to Symes et al. (2011) for a detailed description of the design principles and a concrete implementation of IWAVE++.

---

<sup>1</sup>IWAVE is a software framework for construction of regular grid finite difference and finite element methods for time-dependents partial differential equations.

### 3.1 DAM-LIKE THREE LAYERED MODEL WITH ABSORBING SURFACE

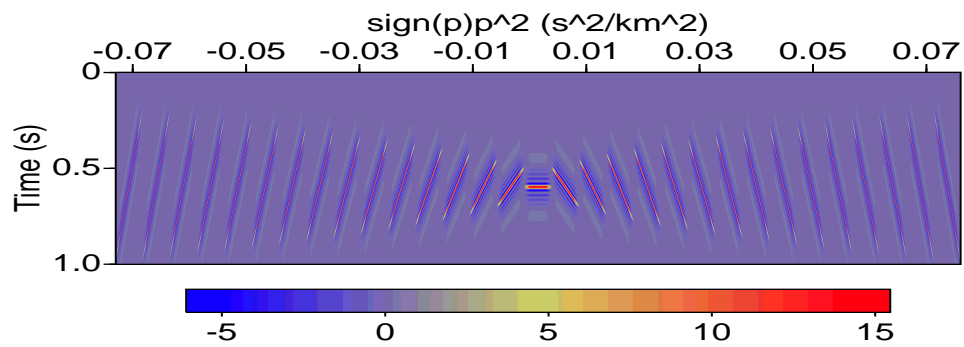
This section shows a nonlinear differential semblance based inversion exercise on a two-dimensional acoustic model with absorbing surface, constant density  $2 \text{ g/cm}^3$  and three layer bulk modulus plotted in Figure 3.1. The acoustic velocities of the three layers are respectively 1.5, 2.5 and 2 km/s. This experiment adopts the two plane wave sources  $f_l$  and  $f_h$  shown in Figure 3.2.



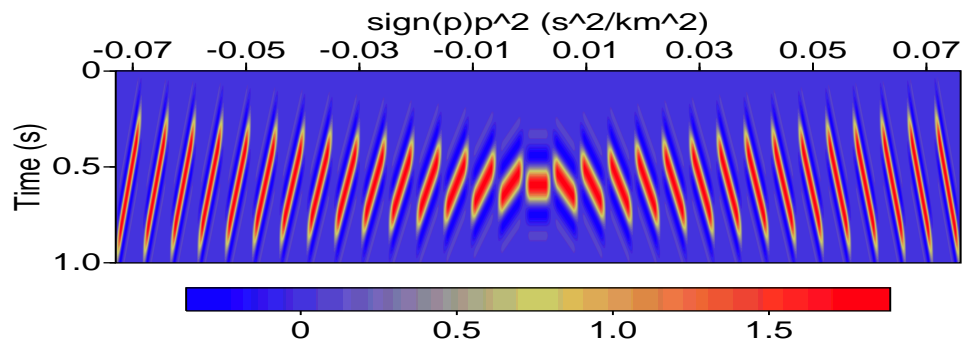
**Figure 3.1:** Three-layer bulk modulus model (4500, 12500, 8000 KPa) with constant density ( $2 \text{ g/cm}^3$ ) – the acoustic velocities: 1.5, 2.5, and 2.0 km/s

This inversion starts from the homogeneous model (4500 KPa) and updates the initial model via 10 differential semblance iterations, each of which contains an extended least-squares inversion via limited memory BFGS (LBFGS) method (Nocedal and Wright, 1999) starting from the newly updated initial model. As described in the previous chapter, in each differential semblance iteration, one can compute the differential semblance gradient via form (2.30). The most expensive and time consuming

Figure 3.2: Band-limited and low-frequency plane wave sources with 31 slowness panels



(a) Band-pass plane-wave source



(b) Low-pass plane-wave source

step of this computation is to solve a normal equation (2.31) to scale  $A^T Am[\bar{m}_i]$  (i.e.,  $\nabla_{\bar{m}} J_{DS}$ ). After computing the gradient, one can adopt a line search strategy to get searching steps and update the initial model, which is expensive because of the necessity of extra evaluations of the differential semblance objective at trial models.

To reduce the computational cost, this experiment makes use of

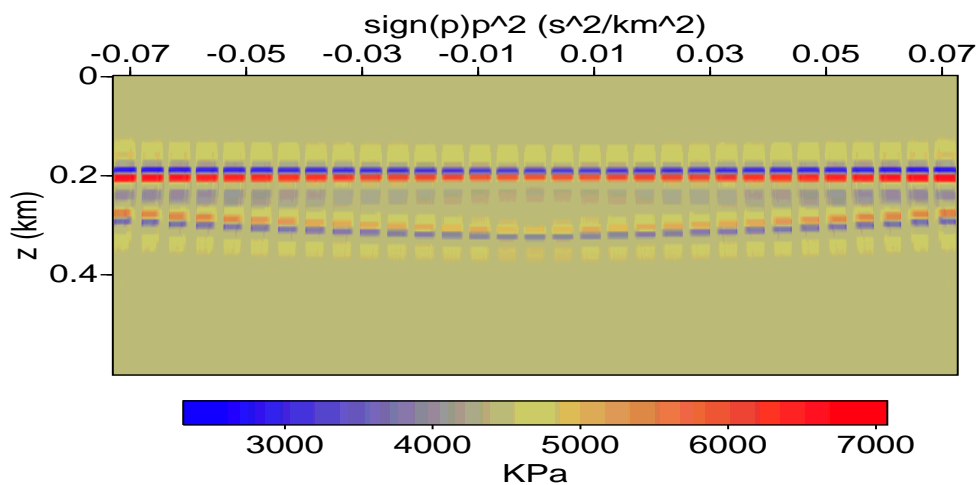
$$g = -DF[m_i]^T (D\bar{F} + D\bar{F}_l) [\bar{m}[m_i]] A^T A \bar{m}[m_i]$$

as the searching direction and adopts fixed searching step 0.5 in each differential semblance iteration instead of performing line search along  $-\nabla_{m_i} J_{DS}$ . Note that this kind of simplifications is just an initial effort towards reducing the computational cost; to use this kind of simplifications for other models, one needs to make further theoretical or numerical justifications to check if each searching direction is descent.

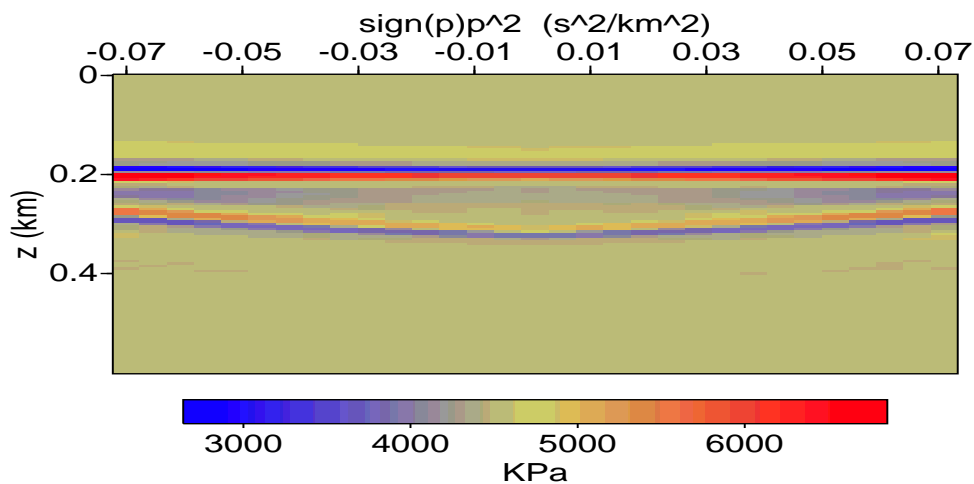
Figure 3.3 shows the final bulk-modulus model computed in the 1st differential semblance iteration, the corresponding model-gather at the middle offset (1500(m)), and the updated initial model  $m_i^2$  that will be used in the next differential semblance iteration. Figure 3.4 plot the initial model  $m_i^6$  used in the 6th differential semblance iteration, the final bulk-modulus model computed in this iteration and the corresponding model-gather at the middle offset. Figure 3.5 consists of the same set of plots for the 10th differential semblance iteration. As shown, this differential semblance inversion recovers the long scale structure and flattens model gathers. Figure 3.6 plots the vertical slices of all the initial and final models at the middle offset (1500(m)). Figures 3.7 presents the target data, final predicted data and data residual after 10 differential semblance iterations. As a comparison, Figure 3.8 shows the solution to a standard least squares inversion starting from the homogeneous model.

As shown, in this experiment the proposed approach yields both a small data residual and a good estimate of the long scale structure, which could not be achieved at the same time via standard least squares inversion.

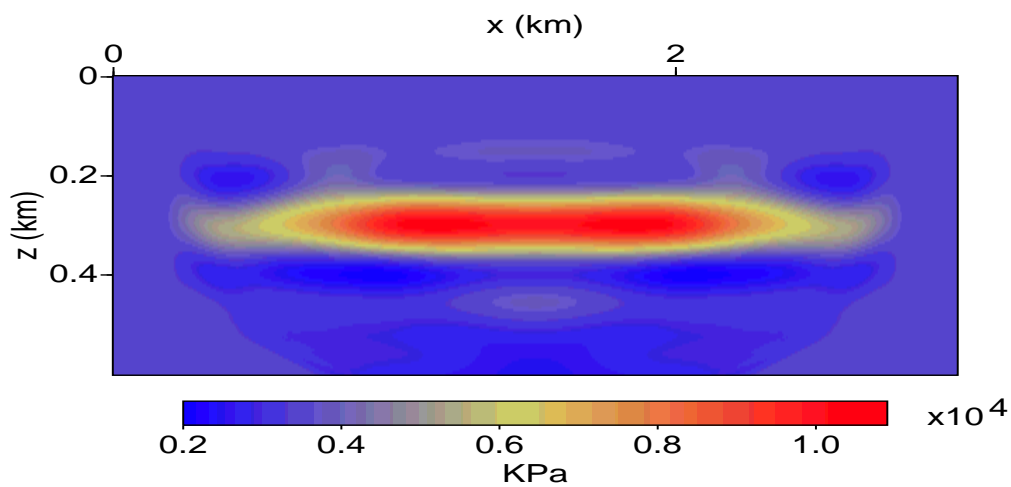
**Figure 3.3:** Final model, its gather at middle offset, and updated initial model in the 1st DS-iteration



(a) Final model computed in the 1st DS-iteration



(b) Gather of final model at the middle offset (1500(m)) in the 1st DS-iteration



(c) Updated initial model in the 1st DS-iteration



**Figure 3.4:** Initial model, final model and its gather at middle offset in the 6th DS-iteration

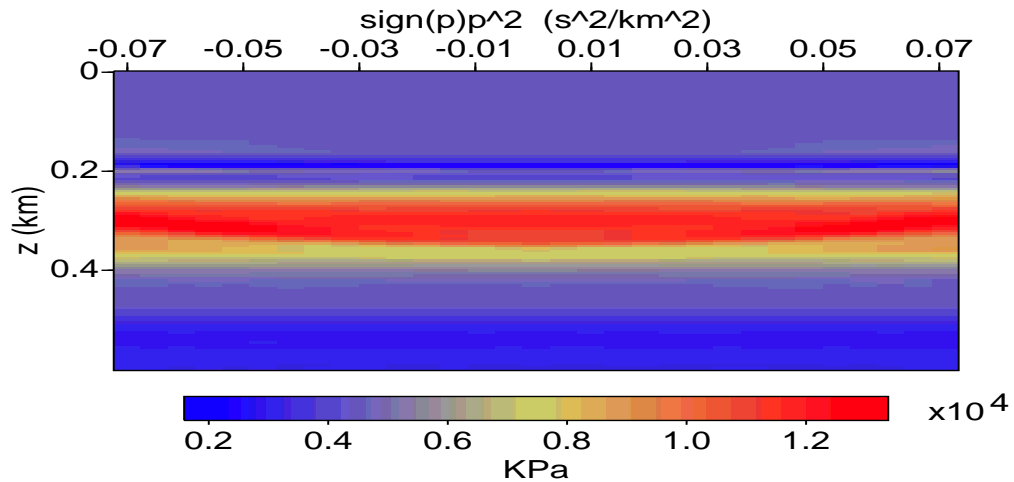
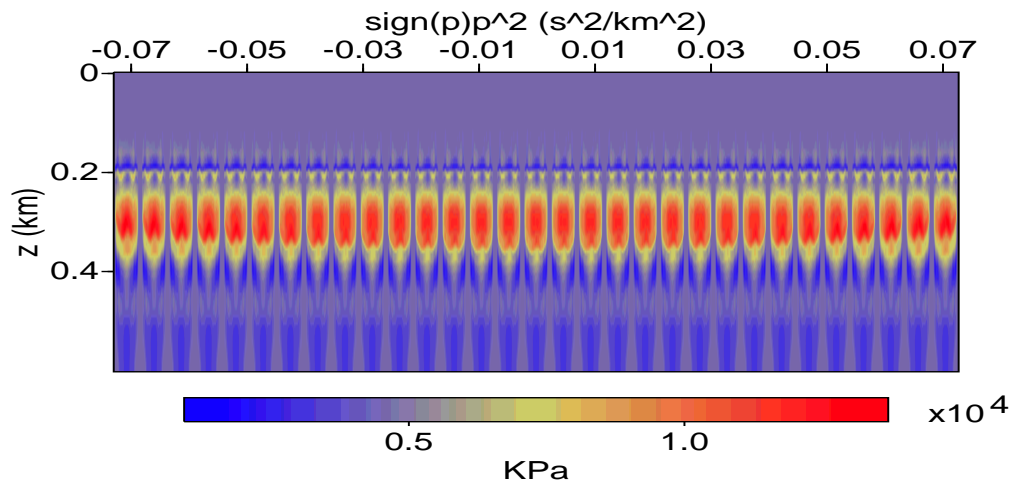
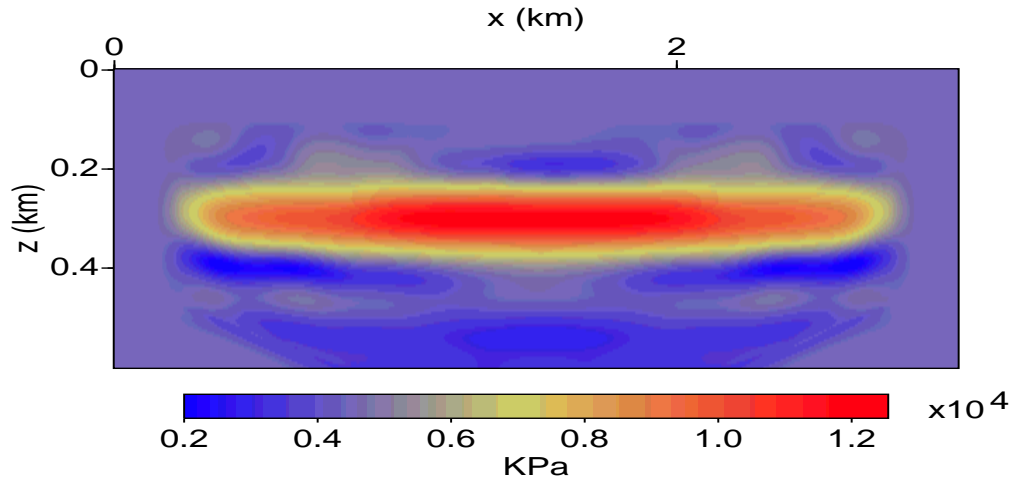


Figure 3.5: Initial model, final model and its gather at middle offset in the 6th DS-iteration<sup>51</sup>

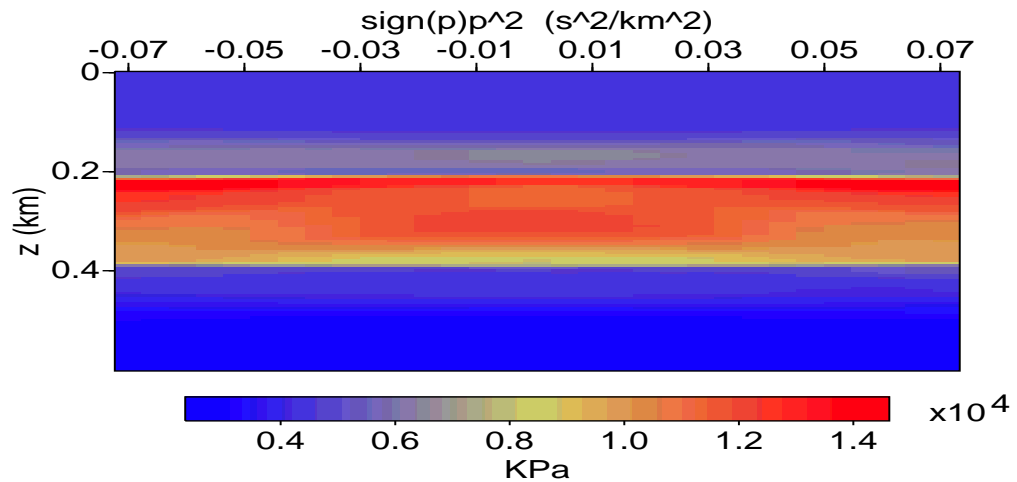
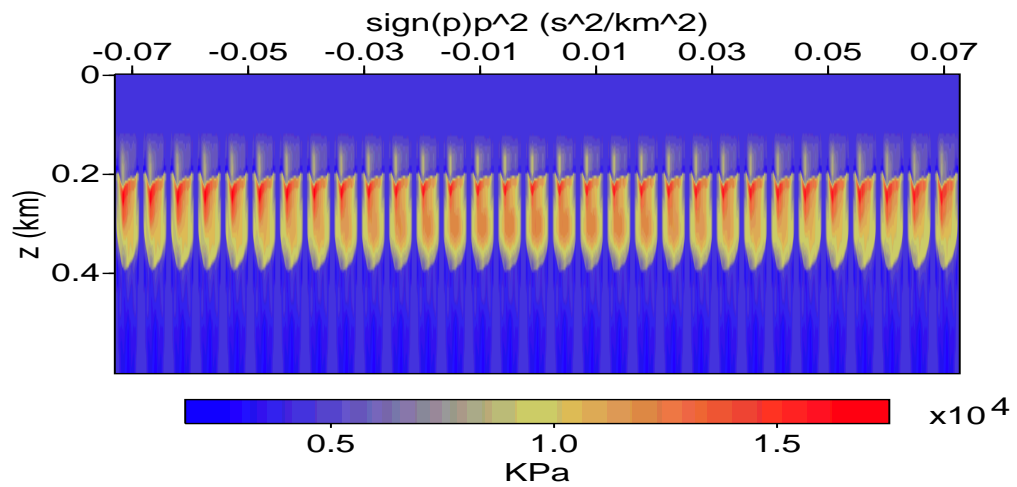
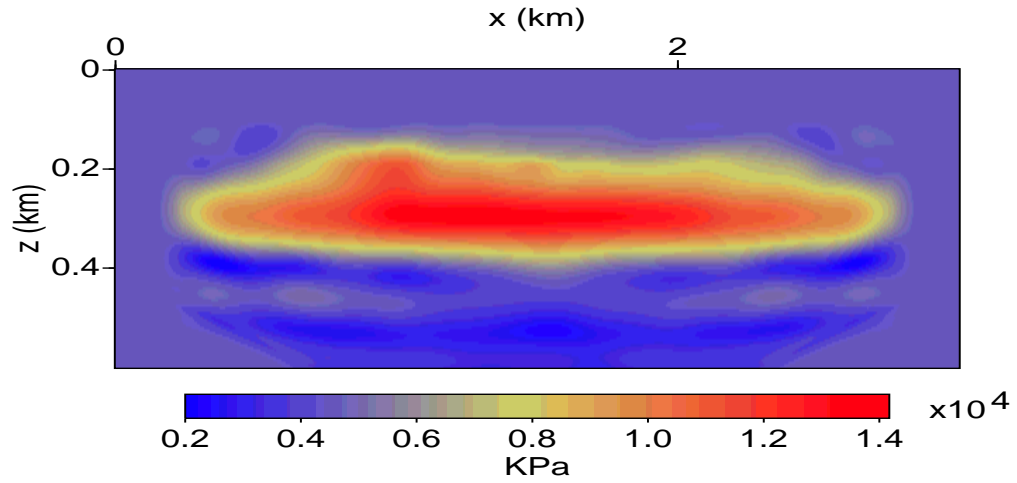
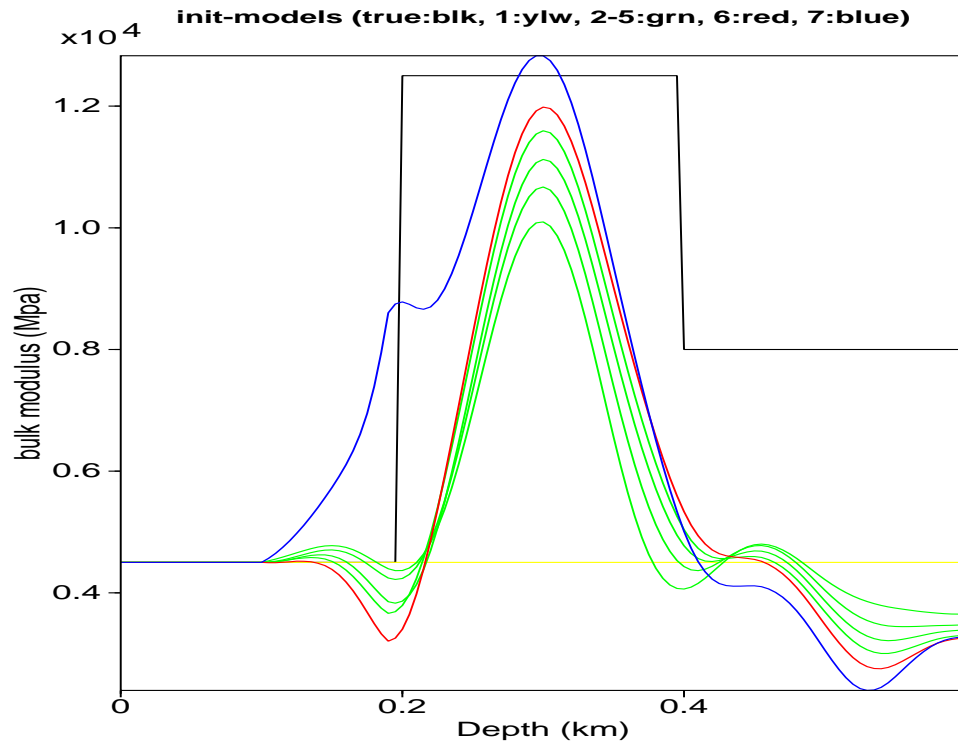
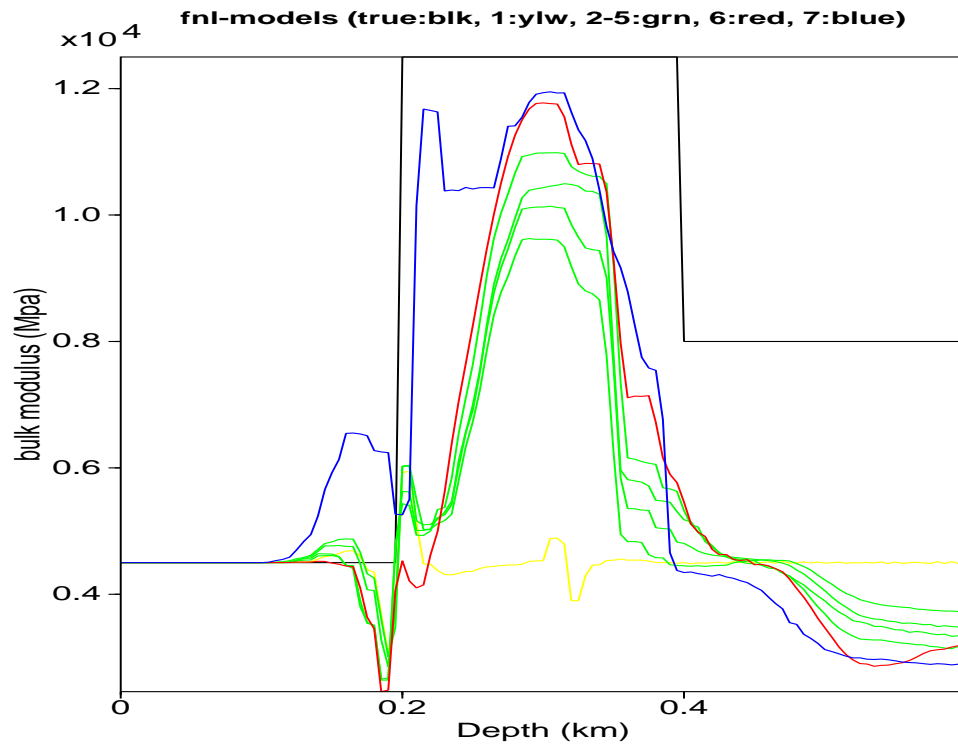


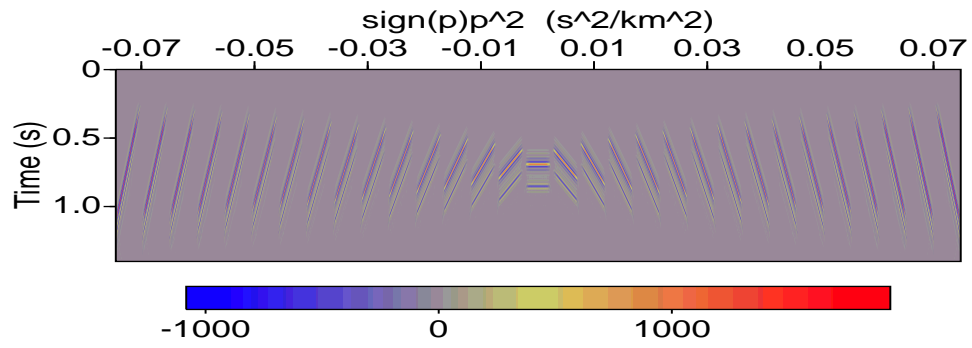
Figure 3.6: Vertical slices of initial and final models at  $x = 1500$ (m)

(a) 1D slices of initial models

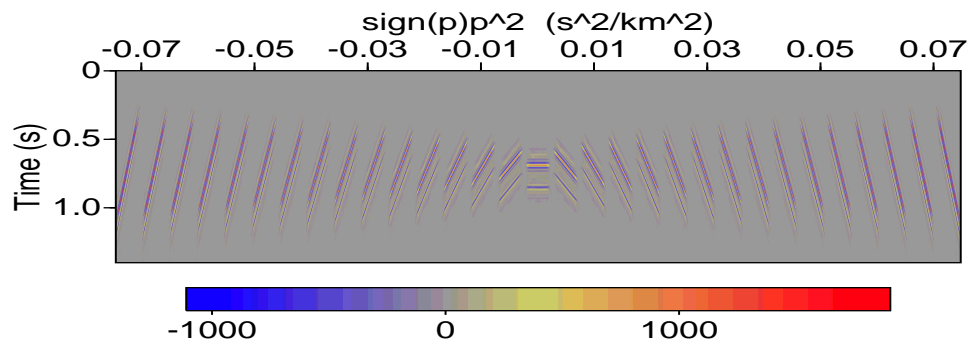


(b) 1D slices of final models

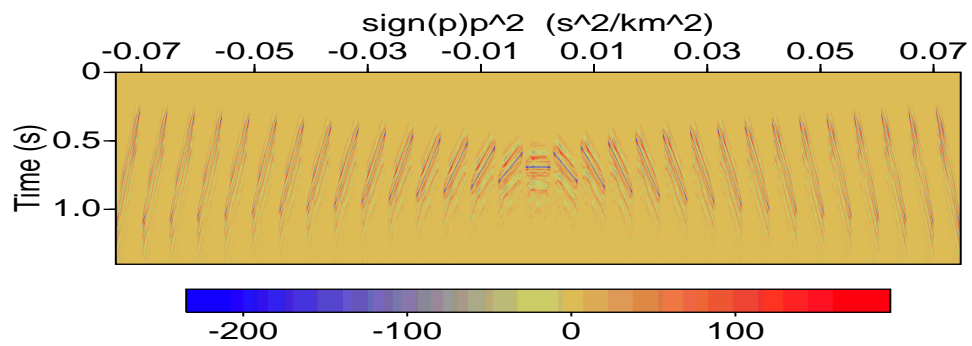
Figure 3.7: Target data, final predicted data and data residual in the 10th DS-iteration



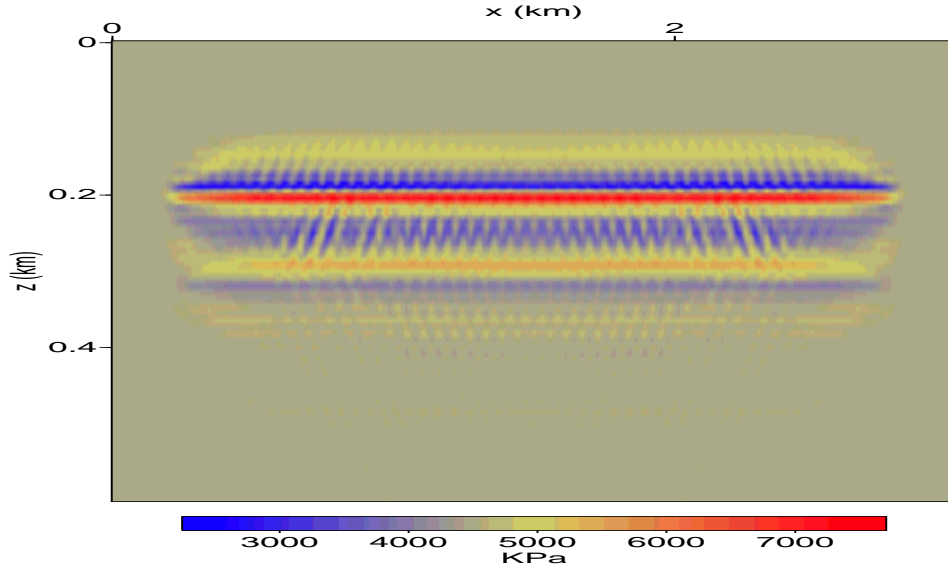
(a) Target data



(b) Final predicted data in DS-iteration 10



(c) Final data residual in DS-iteration 10



**Figure 3.8:** Final model of LS inv starting from homogeneous model after 30 LBFGS iterations

## 3.2 DAM-LIKE THREE LAYERED MODEL WITH FREE SURFACE

This section shows a nonlinear differential semblance based inversion with the same problem setups as used in the previous section except that free surface is adopted in this experiment. This inversion starts also from the homogeneous model (4500 KPa) and updates the initial model via three nonlinear differential semblance iterations, each of which mainly consists of an extended least squares inversion starting from the newly updated initial model, and the computation of updating direction and step-length.

Figure 3.9 plots the gradient computed in the 1st extended least squares iteration and the solution to the first extended least squares inversion after 60 LBFGS iterations, which exhibits clearly moveouts due to the wrong kinematic information provided by the current initial model  $m_l^0$ . Also, this figure demonstrates that least

squares inversion somehow is capable of suppressing multiple reflections, which make it possible to measure primaries-only moveouts but multiple ghosts. Later, I will present more numerical experiments to demonstrate that it is in fact general that multiple reflections are suppressed via least squares inversion.

Figure 3.10 plots the initial model  $m_l^2$  computed from the 2nd differential semblance iteration, the solution  $\bar{m}[m_l^2]$  to the extended least squares inversion in the 3rd differential semblance iteration starting from  $m_l^2$ , and a gather of this solution at the middle surface location  $x = 1.5\text{ km}$ . As one could see, long-scale model update is achieved and the model gather is almost flat.

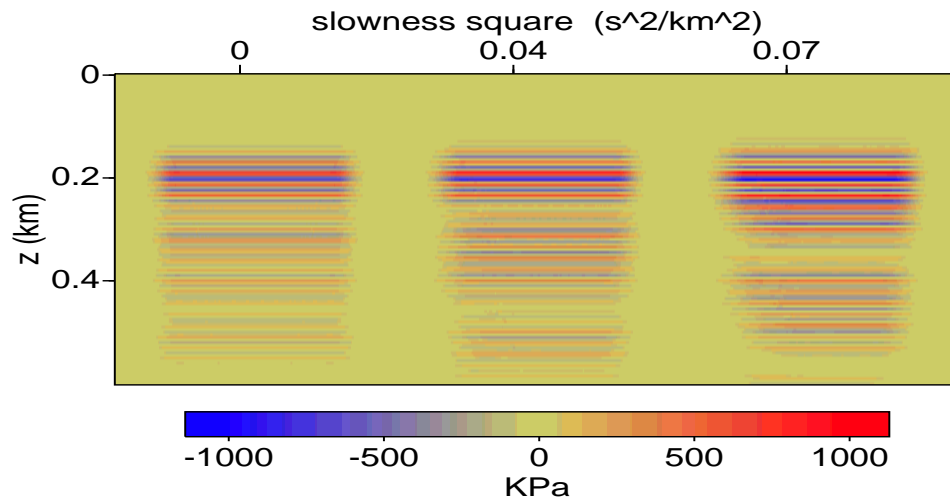
Figure 3.11 shows the middle slices at  $x = 1.5\text{ km}$  of each  $m_l$  computed during the first three nonlinear differential semblance iterations. Figure 3.12 demonstrates the data fitting performance of the extended least squares inversion in the 3rd differential semblance iteration, whose relative root mean square error is about 15%. As shown, the big chunk of anomaly has been recovered after three nonlinear differential semblance iterations. Even with the occurrence of multiple reflections due to free surface, the proposed approach yields a good estimate of the long scale structure as reducing the data misfit and flattening the model gathers, which could not be achieved via either standard least squares inversion or differential semblance variants based on Born assumption.

Figure 3.13(a) is the stack of  $\bar{m}[m_l^2]$ , which could be viewed as a improved starting model for standard FWI. To demonstrate this point, I did a standard FWI with band-limited source  $f_h$  and starting from the stacked model shown in Figure 3.13(a). After 153 LBFGS iterations, this inversion stopped due to the satisfaction of one of our minimization criteria, i.e., the relative gradient reduction is less than 1%. Figure 3.13(b) shows the solution to this inversion, whose the relative root mean

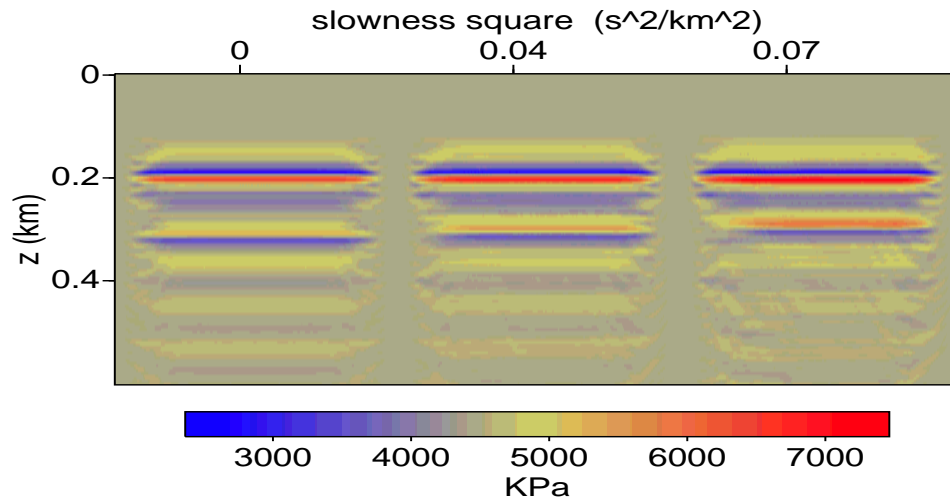
square error is about 6%. As a comparison, I run another FWI starting from the homogeneous model (4500 KPa). Figure 3.13(c) shows its solution after 30 LBFGS iterations. Comparing to the previous one, this inversion only reduced the root mean square residual to about 27% and hung there. Figure 3.14 presents a comparison of their data fitting performance, where only the panels with slowness 0 are plotted for the sake of clarity. Clearly, since the first FWI starts from a more accurate model, it outperforms the second one in all aspects.

With the occurrence of multiple reflections due to free surface, this experiment poses much more difficulties than the one shown in the previous section does. As a result, in this experiment I have to compute the approximate gradient  $\nabla_{m_l} J$  via solving the normal equation. The more complex the model, the more accurate solution to the normal equation is necessary to guarantee the decency of searching direction. It is an important future topic to investigate how the inexact solutions to the extended least-squares inversion and normal equation affect the whole minimization procedure. Until achieving at least a rough estimate of the accuracies one needs for solving the extended least-squares inversion and the normal equation, one has to take the cost of overly solve the two sub problems, which are so expensive that prevents the application of this algorithm in more complex models.

**Figure 3.9:** Migrated and Inverted bulk-modulus in the 1st DS-iteration



(a) 1st LS gradient in 1st extended LS inversion



(b) Inverted bulk-modulus in 1st DS-iteration



Figure 3.10: Initial and final models in the 3rd DS iteration

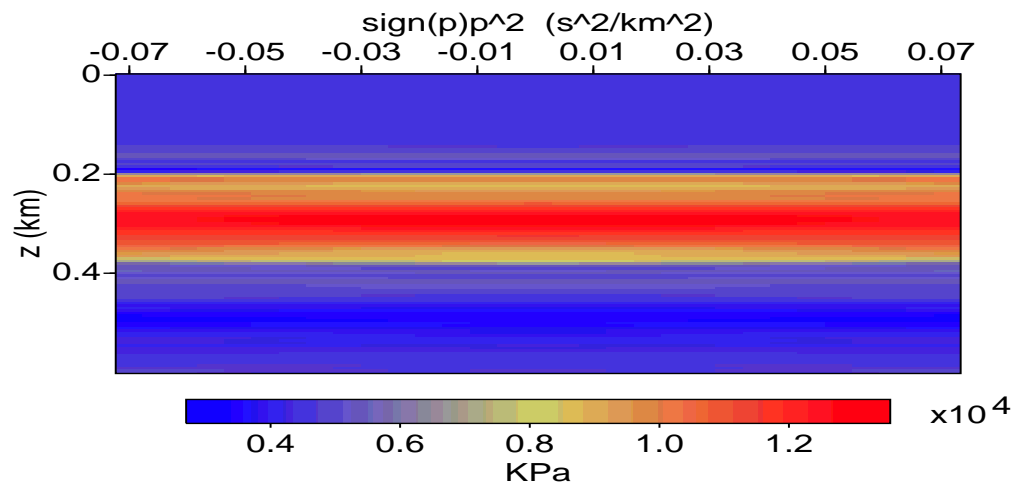
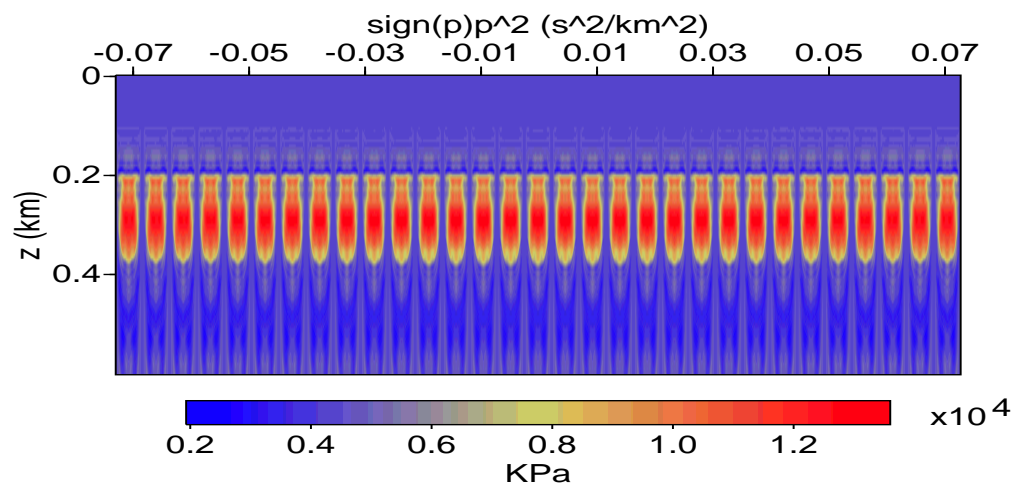
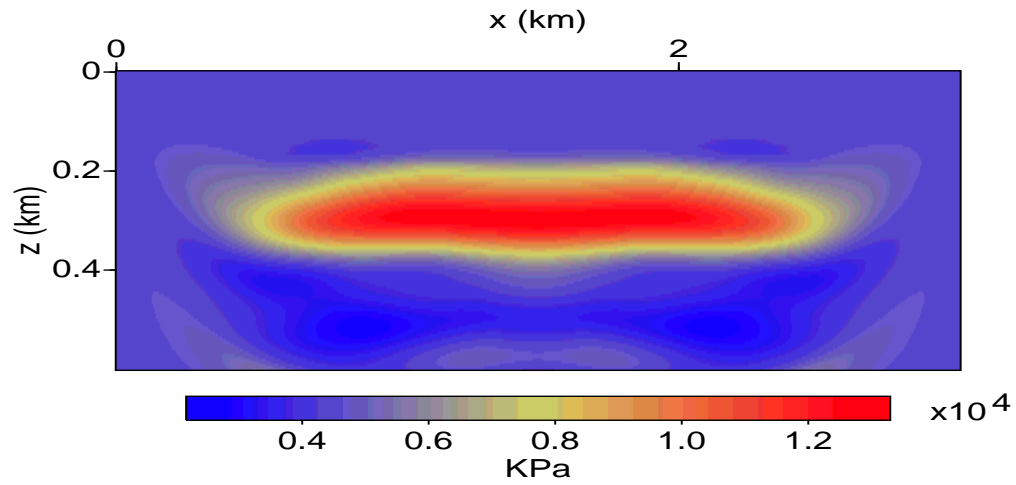
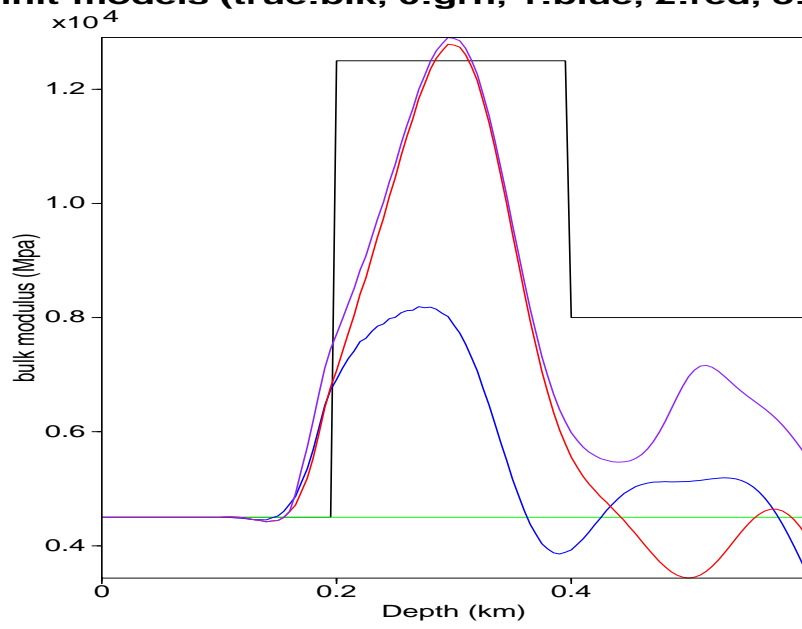
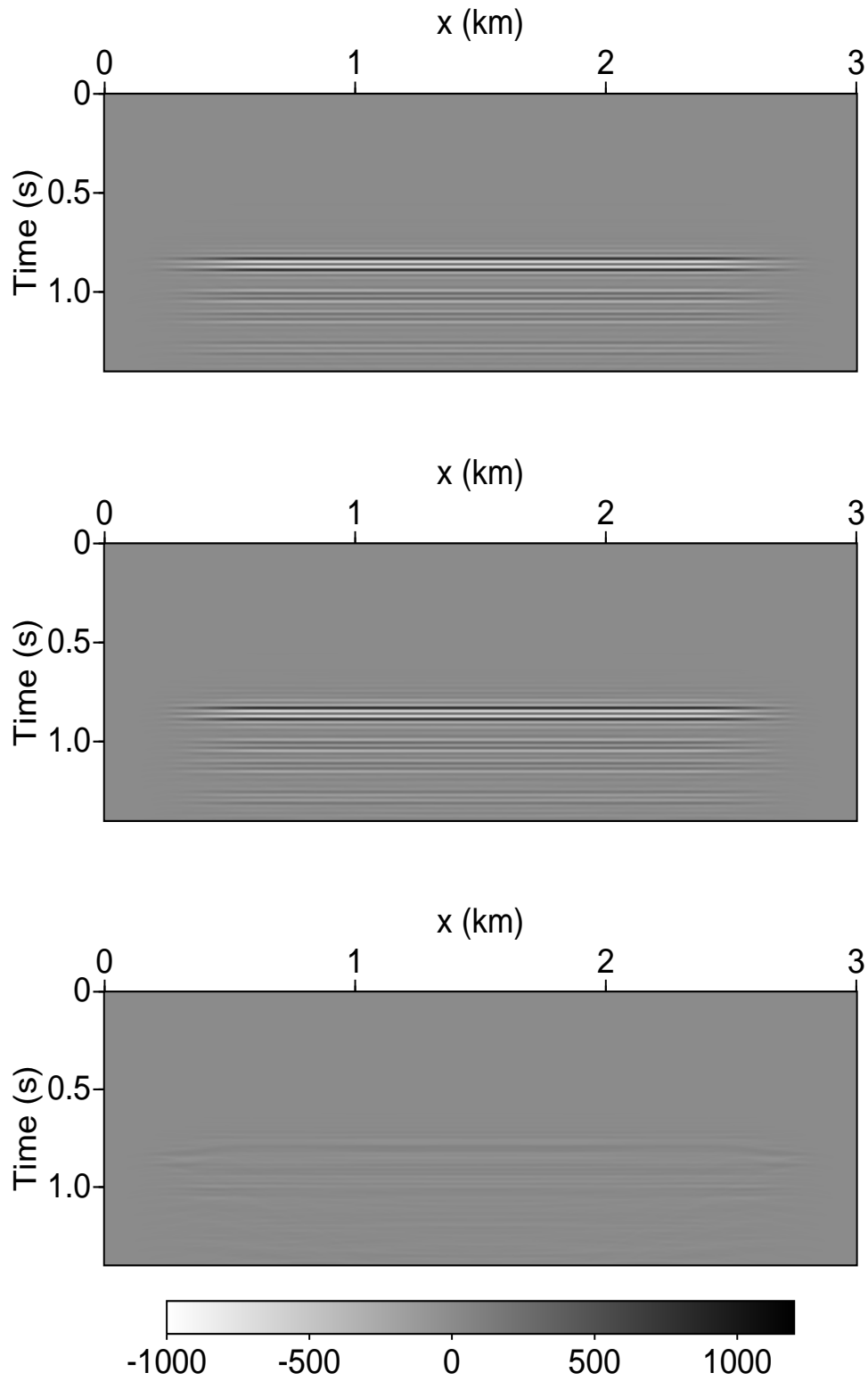


Figure 3.11: Middle slices of  $m_i^k$  ( $k = 0, 1, 2, 3$ )

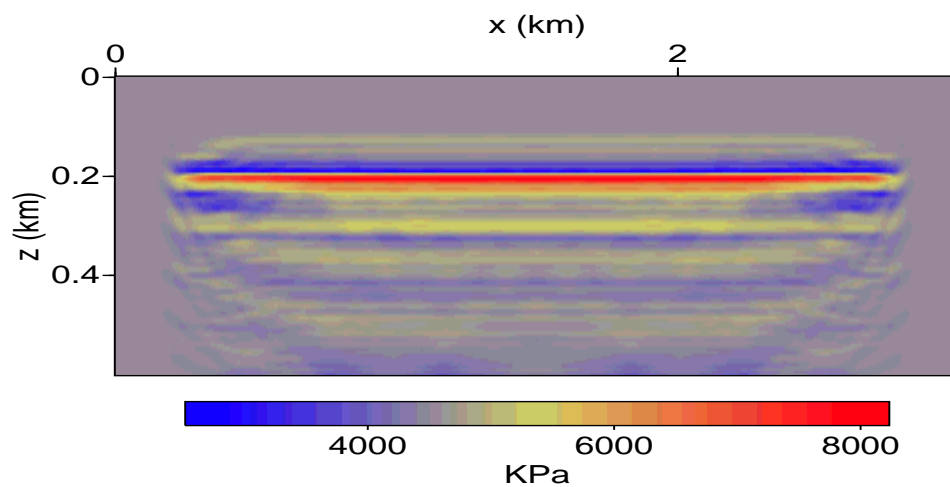
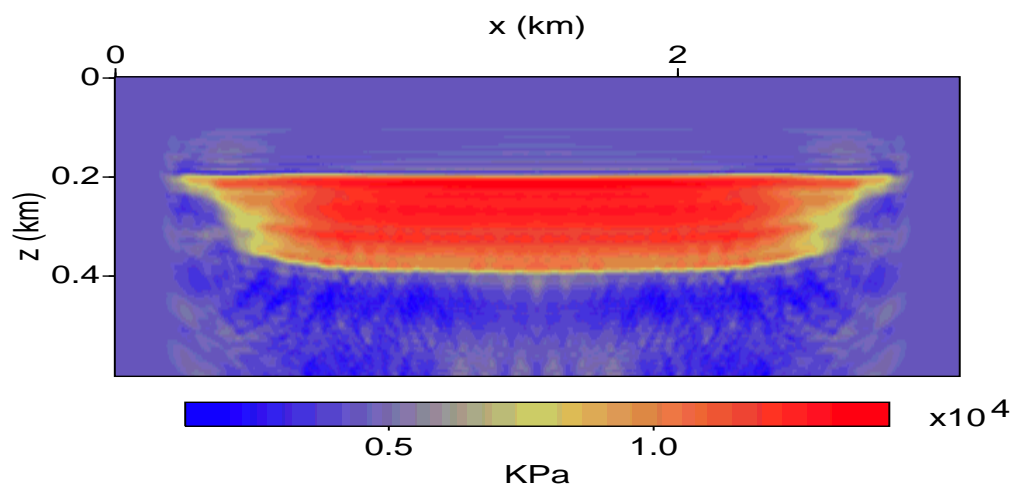
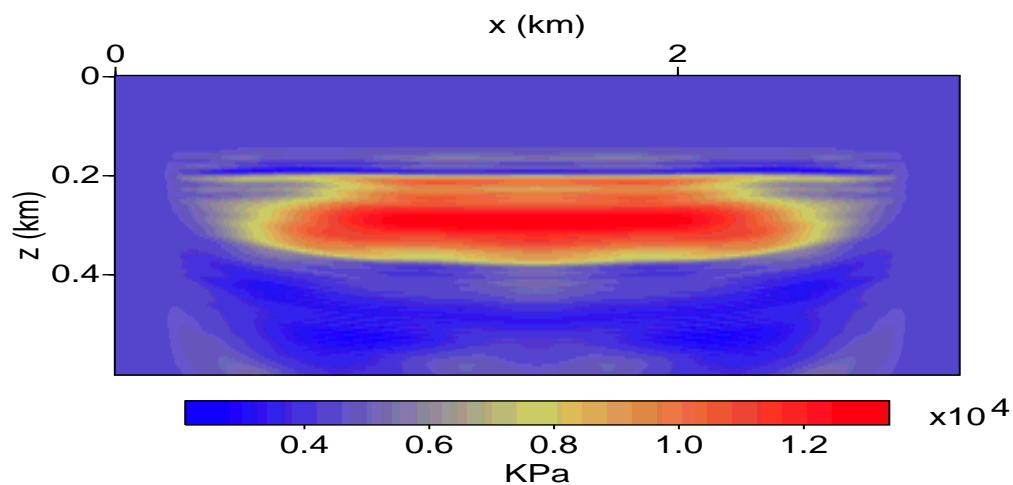
init-models (true:blk, 0:grn, 1:blue, 2:red, 3:purp)



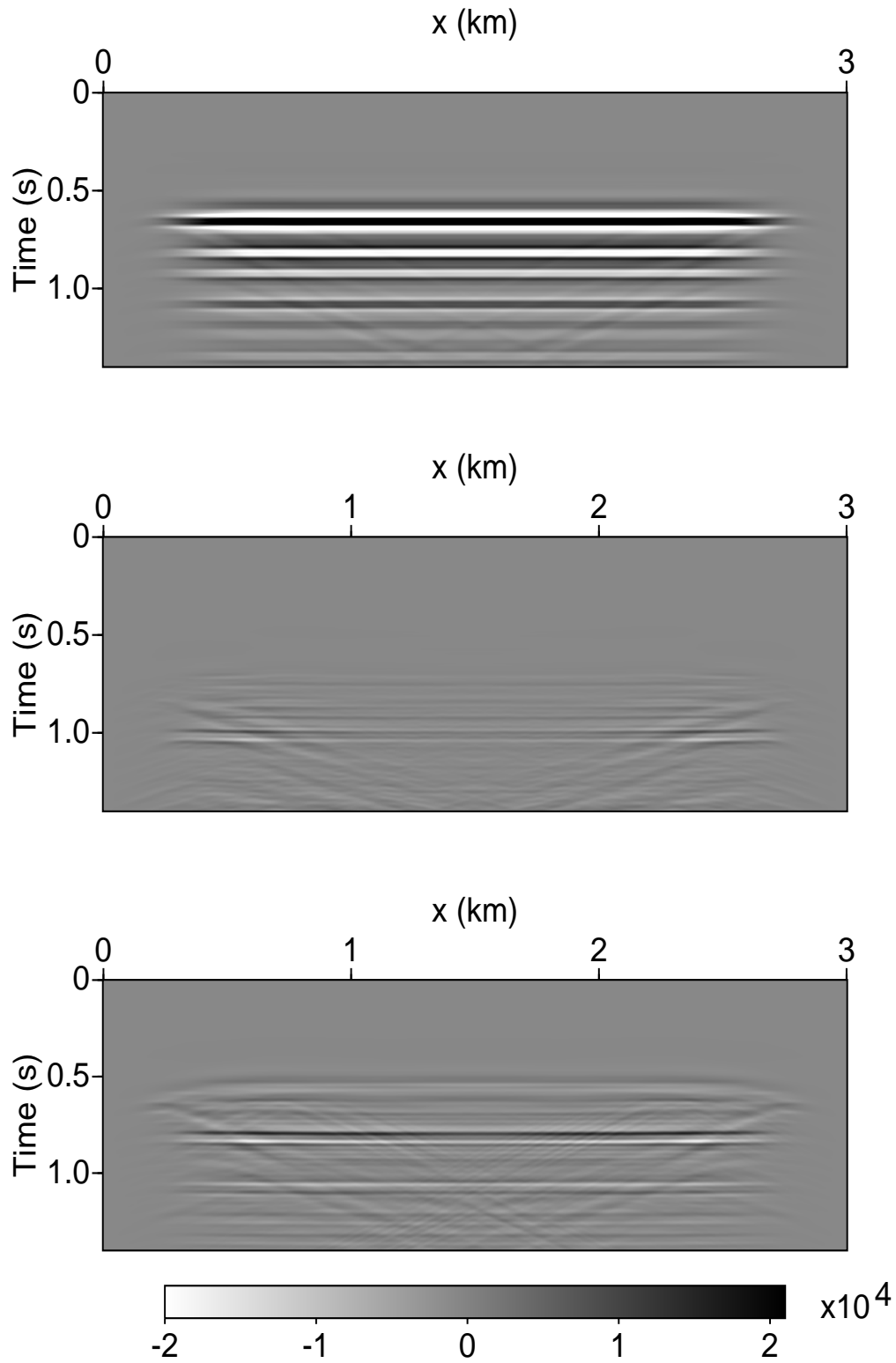
**Figure 3.12:** Top, target data; Middle, predicted data from  $\bar{m}[m_l^2]$ ; Bottom, data residual (0 slowness panels)



**Figure 3.13:** Initial and final models of the LS inversion based on the final model generated by DS inversion

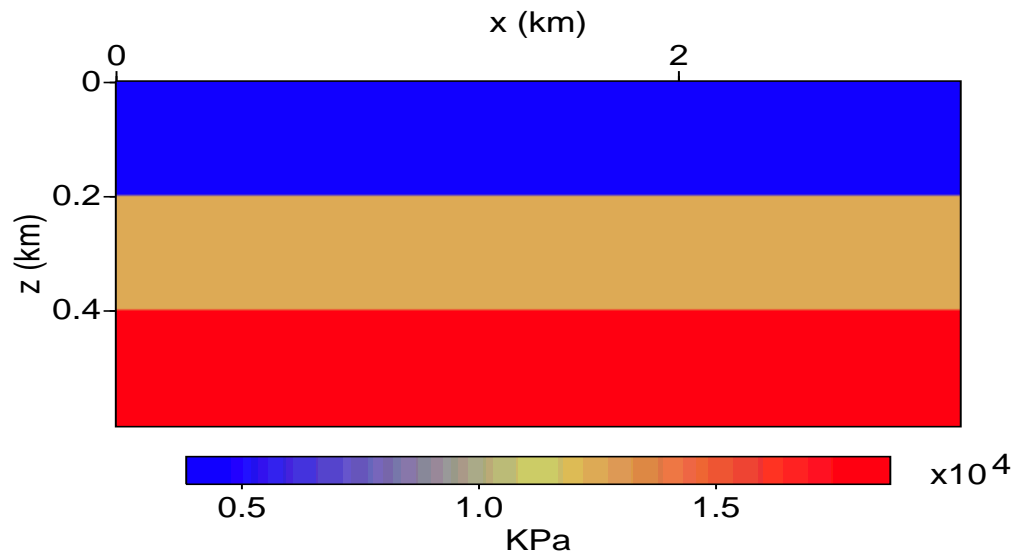


**Figure 3.14:** Top, target data; Middle, data residual for LS inv from  $\bar{m}[m_l^2]$  stack; Bottom, residual for LS inv from  $m_l^0$  (0 slowness panels)



### 3.3 STEP-LIKE THREE LAYERED MODEL WITH FREE SURFACE

This section shows another nonlinear differential semblance based inversion on the three-layer model shown in Figure 3.15 with free surface. The acoustic velocities for the three layers are respectively 1.5, 2.5 and 3.0 km/s. The same source time functions  $f_l$  and  $f_h$  are used (plotted in Figure 3.2).



**Figure 3.15:** Three-layer bulk modulus model (4500, 12500, 18000 KPa) with constant density ( $2 \text{ g/cm}^3$ ) – the acoustic velocities: 1.5, 2.5, and 3.0 km/s

This inversion starts from the homogeneous model (4500 KPa) and updates the initial model via three nonlinear differential semblance iterations, each of which mainly consists of an extended least squares inversion starting from the newly updated initial model, and the computation of updating direction and step-length. Figure 3.16 plots the gradient computed in the 1st extended least squares iteration and the solution to the first extended least squares inversion after 60 LBFGS iterations, which exhibits clearly moveouts due to the wrong kinematic information provided by the current

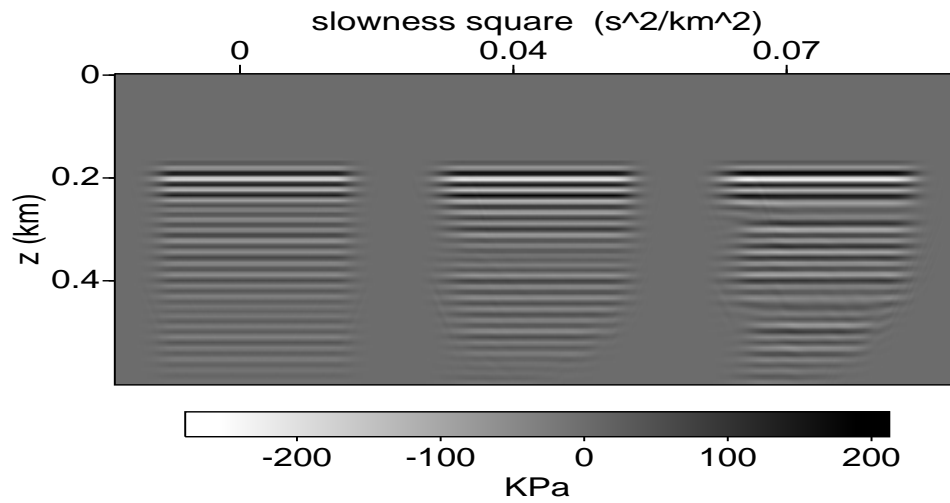
initial model  $m_l^0$ . Figure 3.16 also demonstrates that least squares inversion somehow is capable of suppressing multiple reflections, which make it possible to measure primaries-only moveouts but multiple ghosts.

Figure 3.17 plots the initial model  $m_l^2$  used in the 3rd differential semblance iteration, the solution  $\bar{m}[m_l^2]$  to the extended least squares inversion in the third differential semblance iteration starting from  $m_l^2$ , and a gather of this solution at the middle surface location  $x = 1.5$  km, which is almost flat. Figure 3.18 shows the middle slices at  $x = 1.5$  km of each  $m_l$  computed during the first three nonlinear differential semblance iterations. As one could see, long-scale model updates are achieved. Figure 3.19 demonstrates the data fitting performance of the extended least squares inversion in the 3rd differential semblance iteration, whose relative root mean square error is about 11.4%. Figure 3.20(a) is the stack of  $\bar{m}[m_l^2]$ . As shown, the big chunk of anomaly has been recovered after three nonlinear differential semblance iterations.

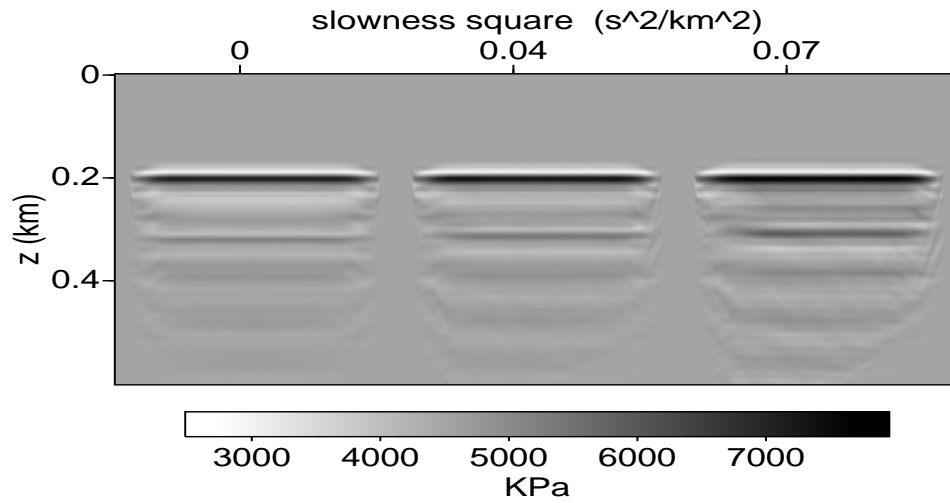
Starting from the stacked model shown in Figure 3.20(a), I run a standard FWI with band-limited source  $f_h$ . Figure 3.20(b) shows the solution to this inversion after 60 LBFGS iterations, whose relative root mean square error is about 14%. As a comparison, I run another FWI starting from the homogeneous model (4500 KPa). Figure 3.20(c) shows its solution after 60 LBFGS iterations. Comparing to the previous one, this inversion only reduced the root mean square residual to about 27% and hung there. Figure 3.21 presents a comparison of their data fitting performance, where only the panels with slowness 0 are plotted for the sake of clarity. Clearly, since the first FWI starts from a more accurate model, it outperforms the second one in all aspects.

In contemporary imaging or inversion practices, multiple reflections are treated as noises and have to be suppressed during pre-processing procedure. While in the previous

**Figure 3.16:** Migrated and Inverted bulk-modulus in the 1st DS-iteration



(a) 1st LS gradient in 1st extended LS inversion



(b) Inverted bulk-modulus in 1st DS-iteration



Figure 3.17: Initial and final models in the 3rd DS iteration

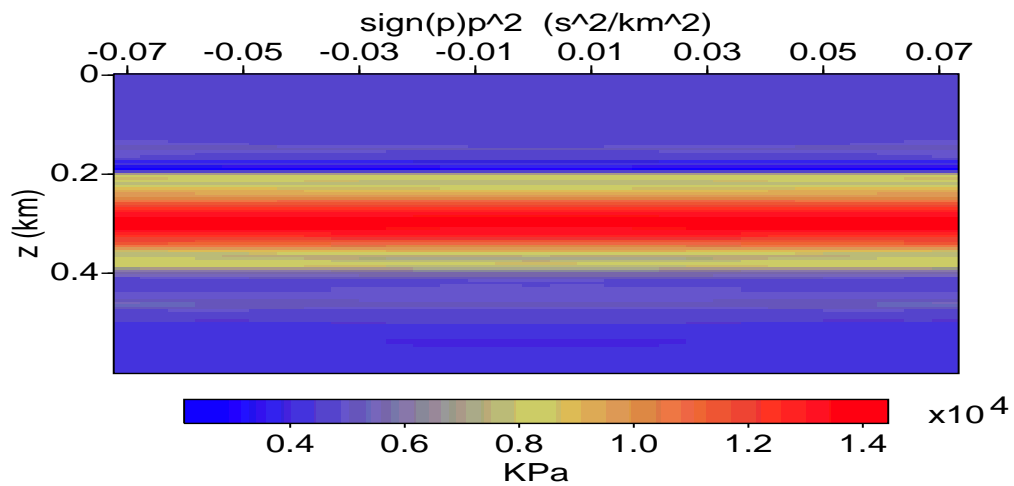
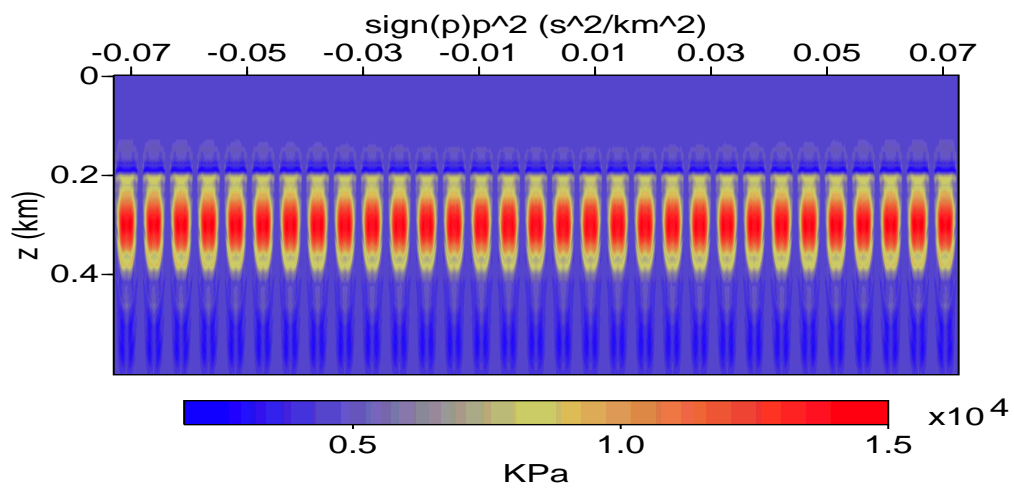
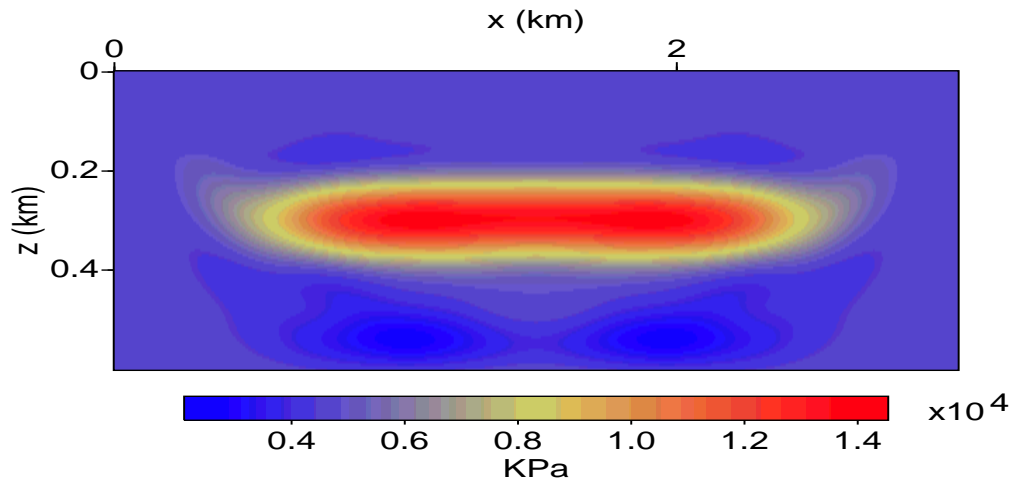
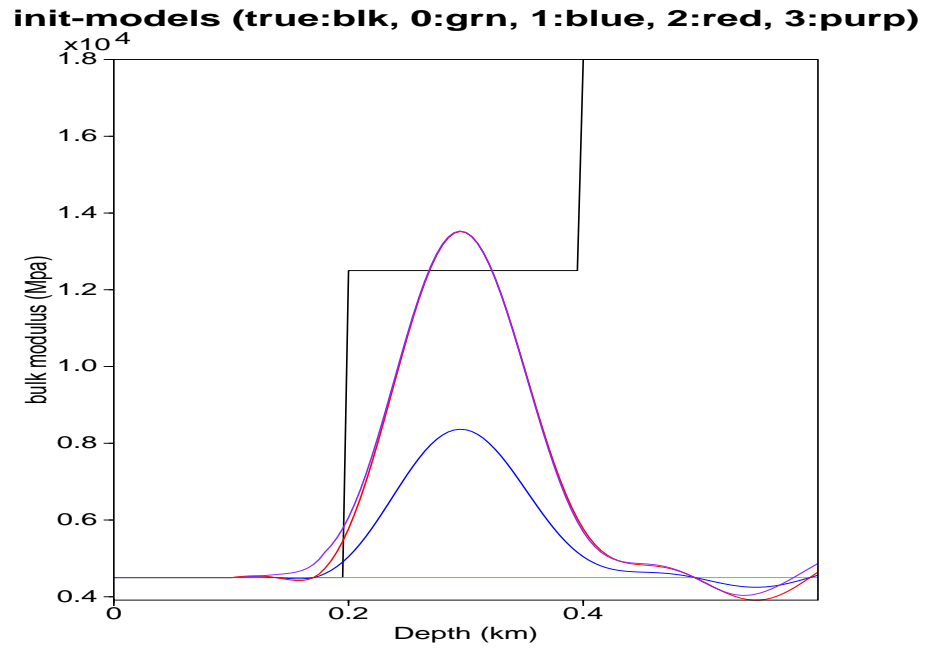
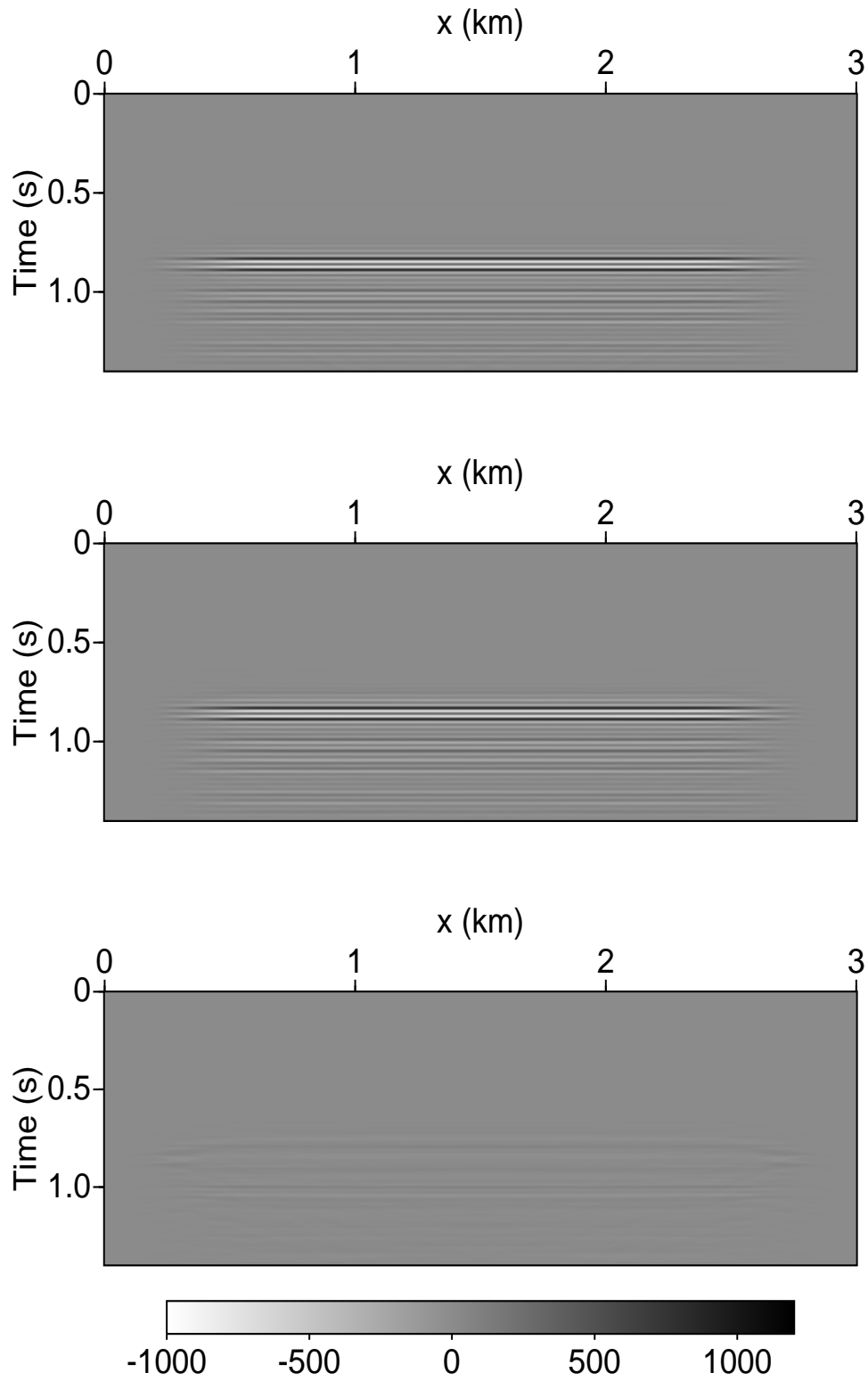


Figure 3.18: Middle slices of  $m_i^k$  ( $k = 0, 1, 2, 3$ )

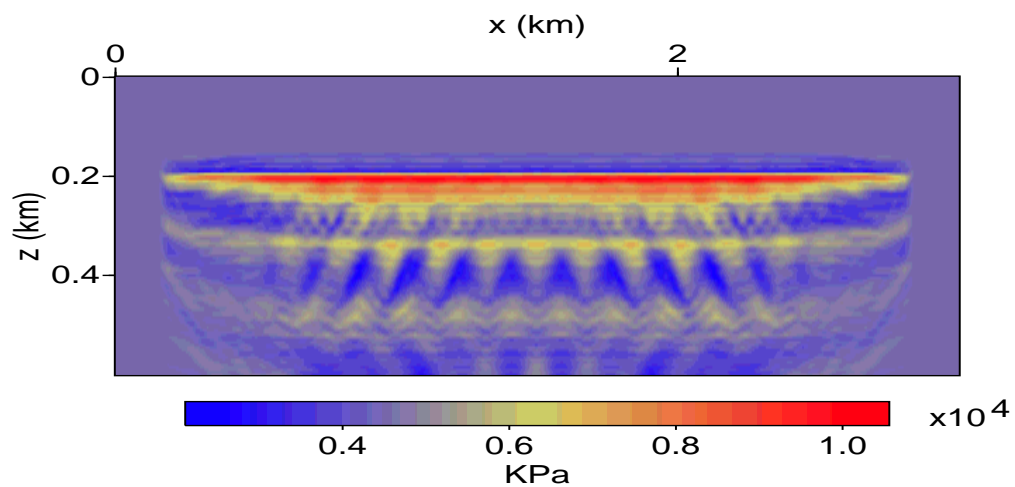
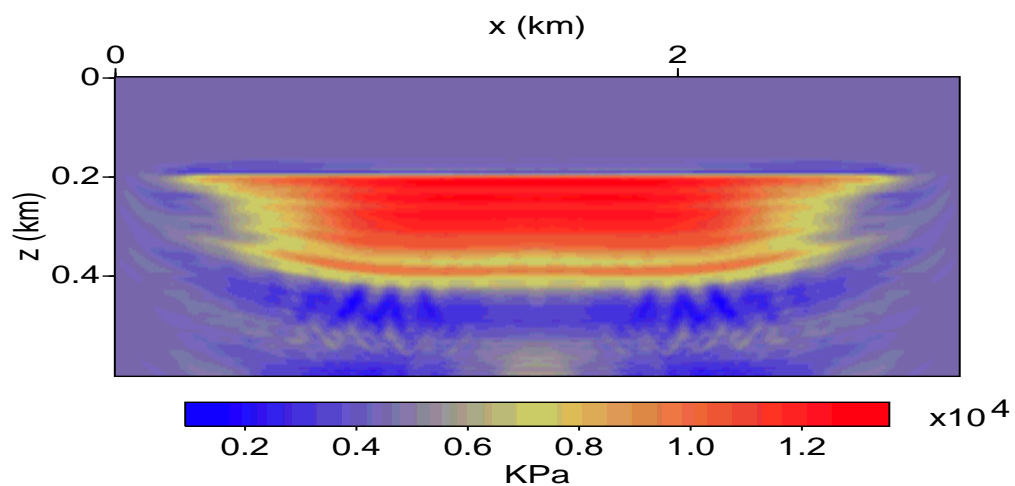
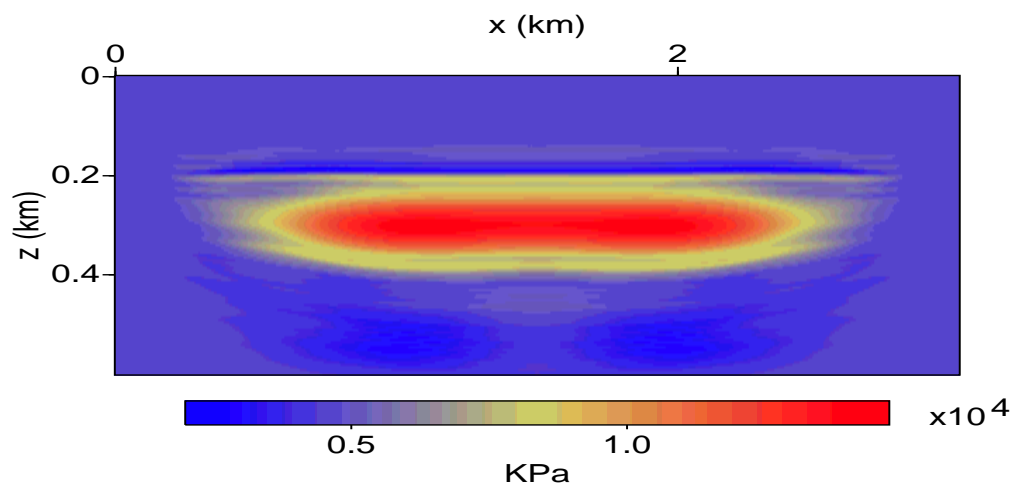


two experiments, the proposed algorithm treats multiple reflections as meaningful information as other data. This owes to the fact that least squares inversion is capable of suppressing multiple reflections. I will demonstrate this fact in the next section.

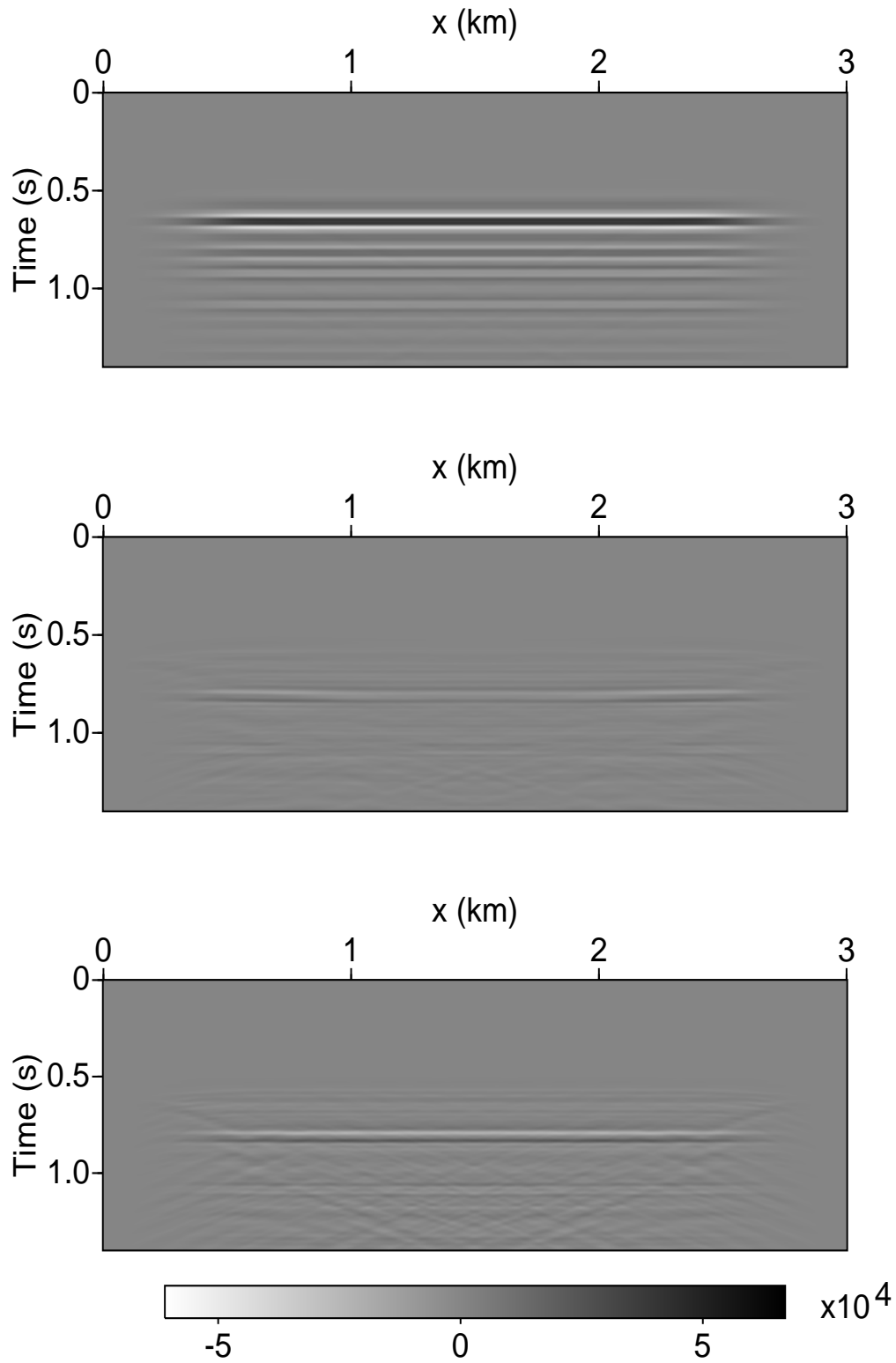
**Figure 3.19:** Top, target data; Middle, predicted data from  $\bar{m}[m_l^2]$ ; Bottom, data residual (0 slowness panels)



**Figure 3.20:** Initial and final models of the LS inversion based on the final model generated by DS inversion



**Figure 3.21:** Top, target data; Middle, data residual for LS inv from  $\bar{m}[m_l^2]$  stack; Bottom, residual for LS inv from  $m_l^0$  (0 slowness panels)



### 3.4 SUPPRESSING MULTIPLE REFLECTIONS WITH LEAST-SQUARES INVERSION

This section shows three inversion experiments with free surface and different models to demonstrate that the extended least squares inversion is capable of suppressing multiple reflections.

The first inversion experiment employs a 2D acoustic model with free surface, constant density  $2 \text{ g/cm}^3$  and four layer bulk modulus plotted in Figure 3.22. The acoustic velocities of the four layers are respectively 1.5, 2.5, 3.0 and 3.5 km/s. This inversion starts from the homogeneous model (4500 KPa) and consists of 30 LBFGS iterations. Clear multiple reflections are observed in the first LS gradient (i.e., the migrated image with respect to homogeneous background) and its middle offset (3000 m) gather shown in Figure 3.23. Those multiple reflections are substantially suppressed during this extended LS inversion, which is shown in the final LS solution and its middle offset gather plotted in Figure 3.24.

The second inversion experiment employs a 2D acoustic model with free surface, constant density  $2 \text{ g/cm}^3$  and three dipped-layer bulk modulus plotted in Figure 3.25. The acoustic velocities of the three zones are respectively 1.5, 2.5 and 2.0 km/s. This inversion starts from the homogeneous model (4500 KPa) and consists of 30 LBFGS iterations. Similar phenomena of multiple suppression is observed. Clear multiple reflections are observed in the first LS gradient (i.e., the migrated image with respect to homogeneous background) and its middle offset (3000 m) gather shown in Figure 3.26. Those multiple reflections are substantially suppressed during this extended LS inversion, which is shown in the final LS solution and its middle offset gather plotted in Figure 3.27.

**Figure 3.22:** Four-layer bulk modulus model (4500, 12500, 18000, and 245000 KPa) with constant density ( $2 \text{ g/cm}^3$ ) – the acoustic velocities: 1.5, 2.5, 3.0 and 3.5 km/s

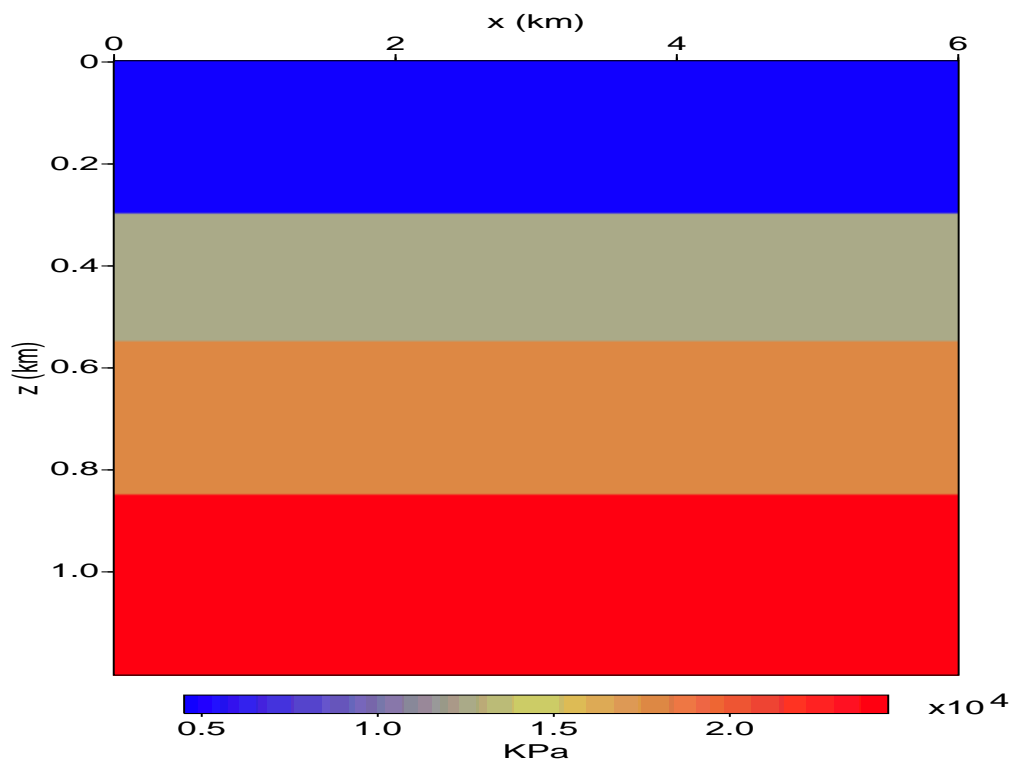
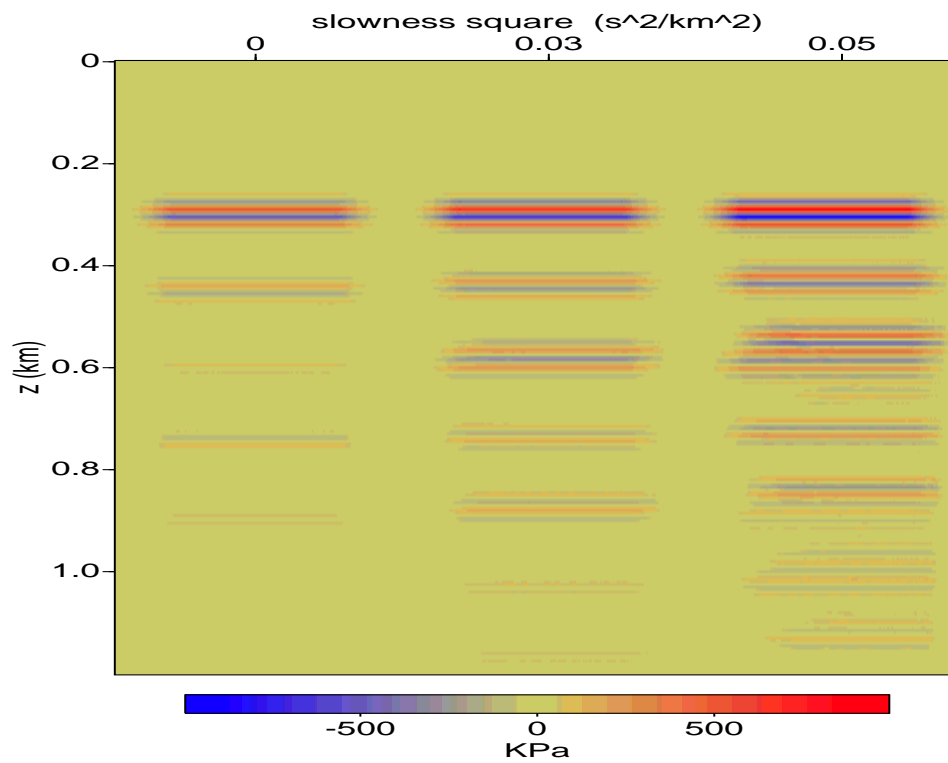
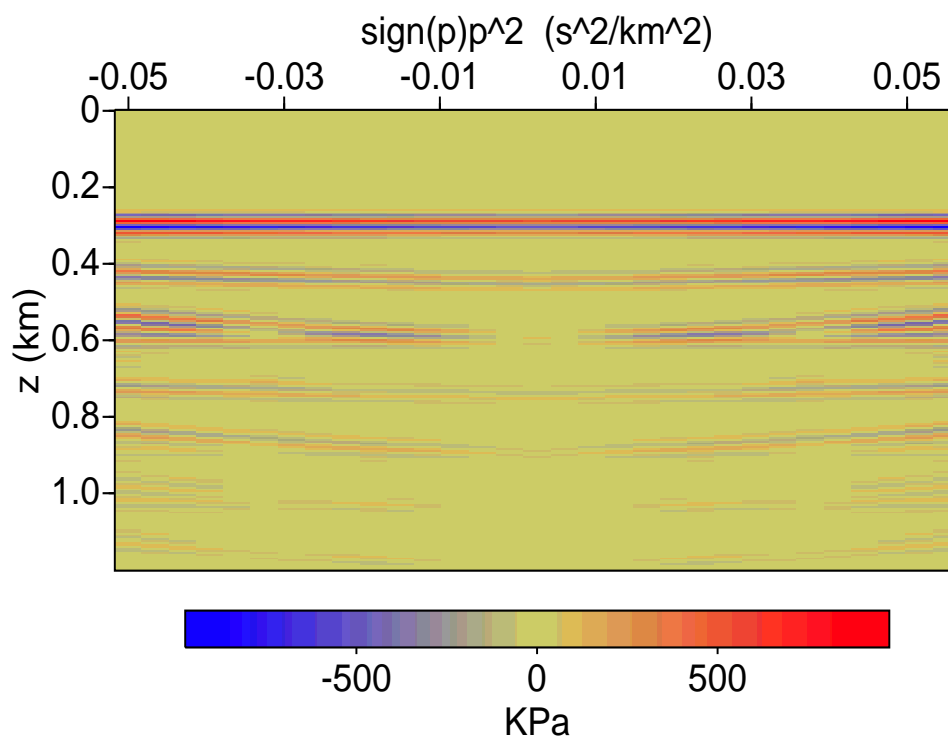


Figure 3.23: Gradient in 1st LS iteration and its gather at the middle offset



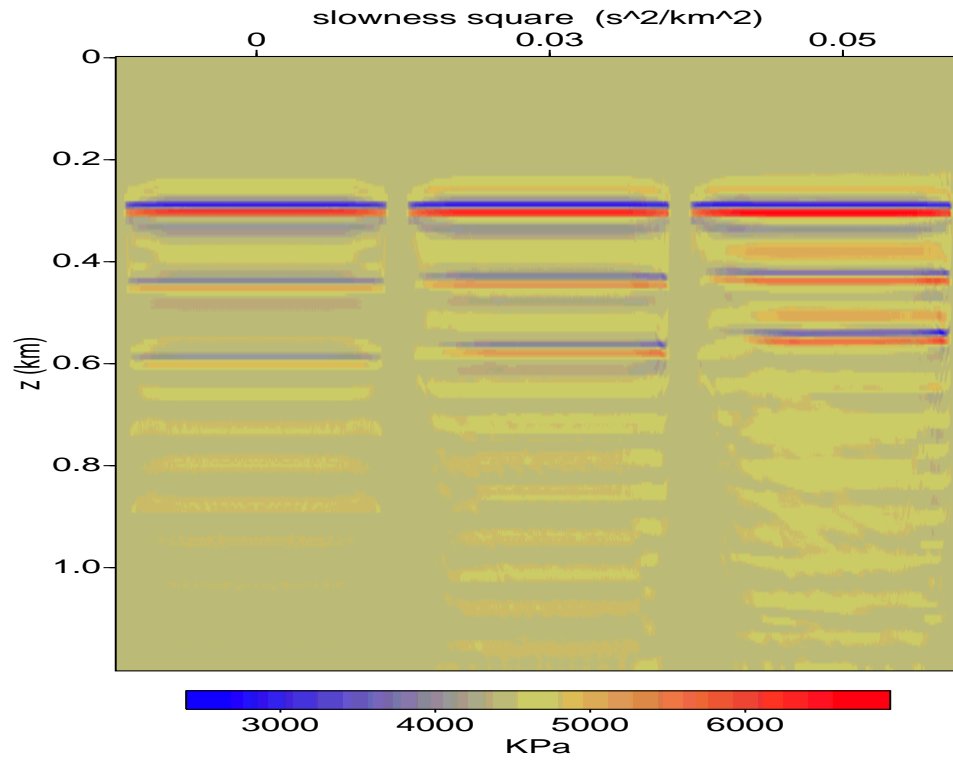
(a) 1st LS gradient



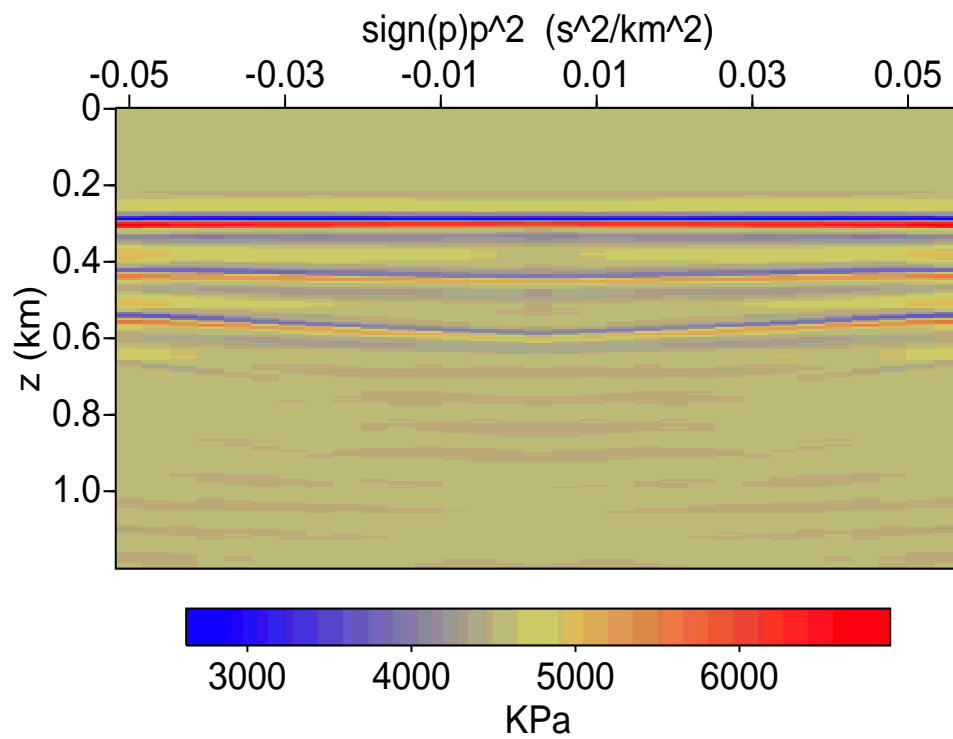
(b) Gather of 1st LS gradient at the middle offset



Figure 3.24: LS solution after 30 LBFGS iterations and its gather at the middle offset

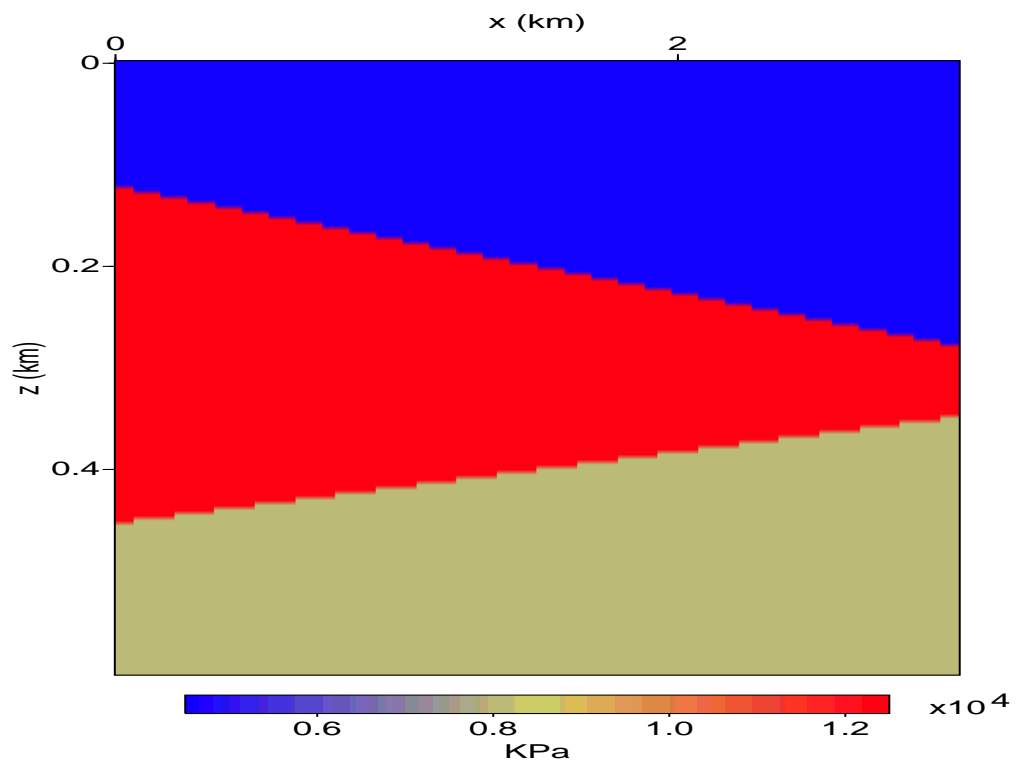


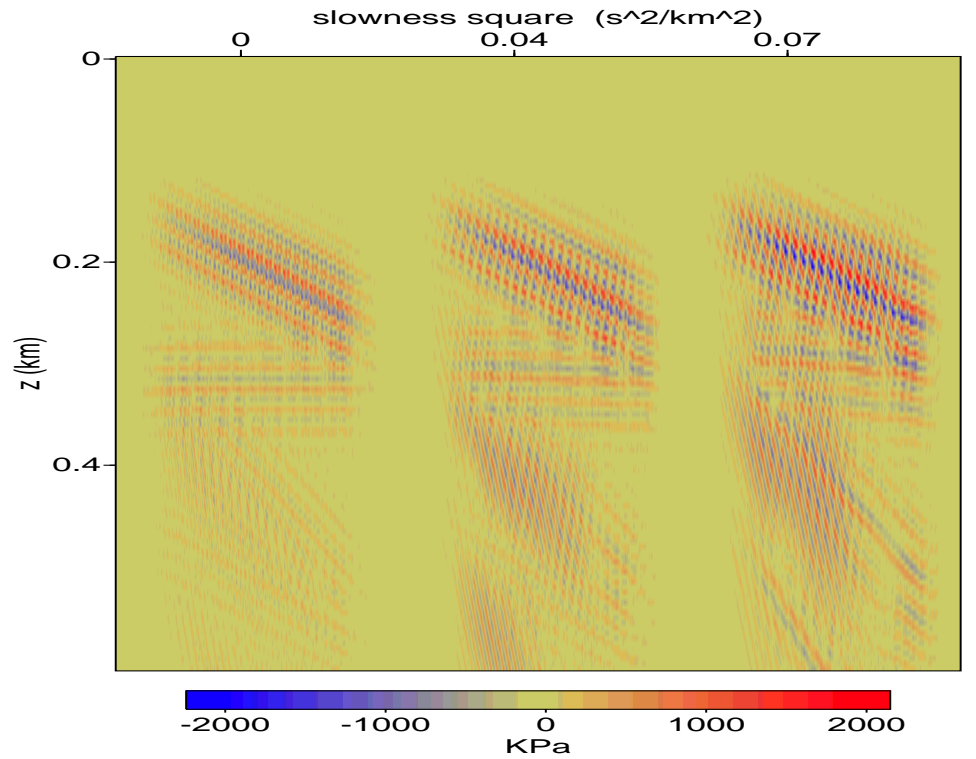
(a) LS solution after 30 iterations



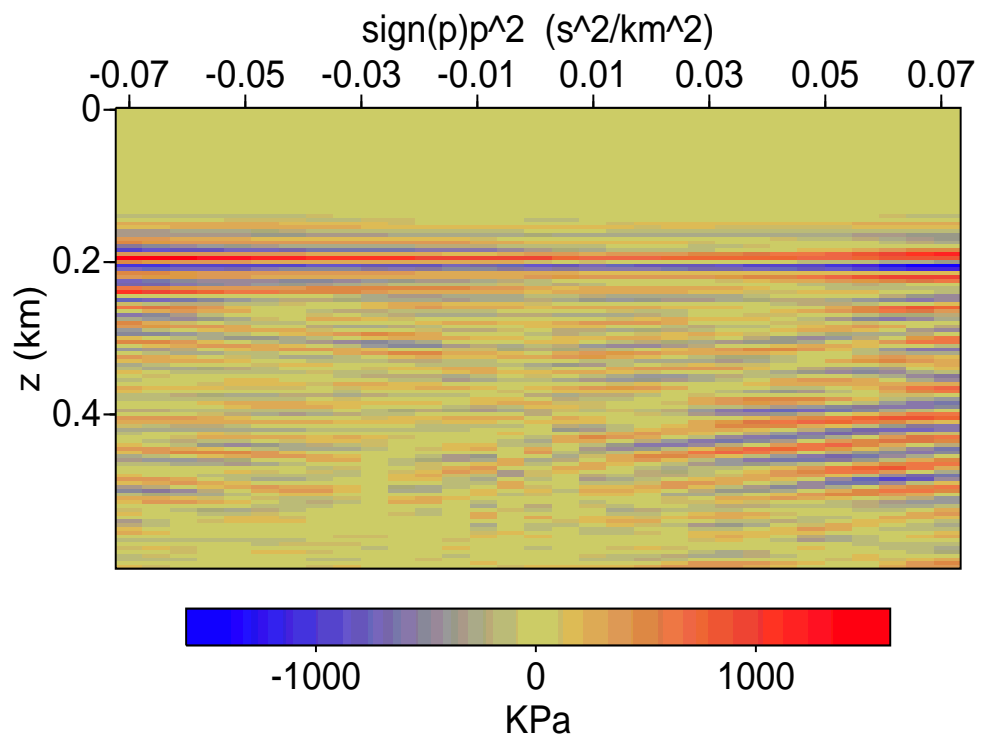
(b) Gather of the final inverted model at the middle offset

**Figure 3.25:** Dipping-layer bulk modulus model (4500, 12500, and 8000 KPa) with constant density ( $2\text{ g/cm}^3$ ) – the acoustic velocities: 1.5, 2.5, and 2.0 km/s



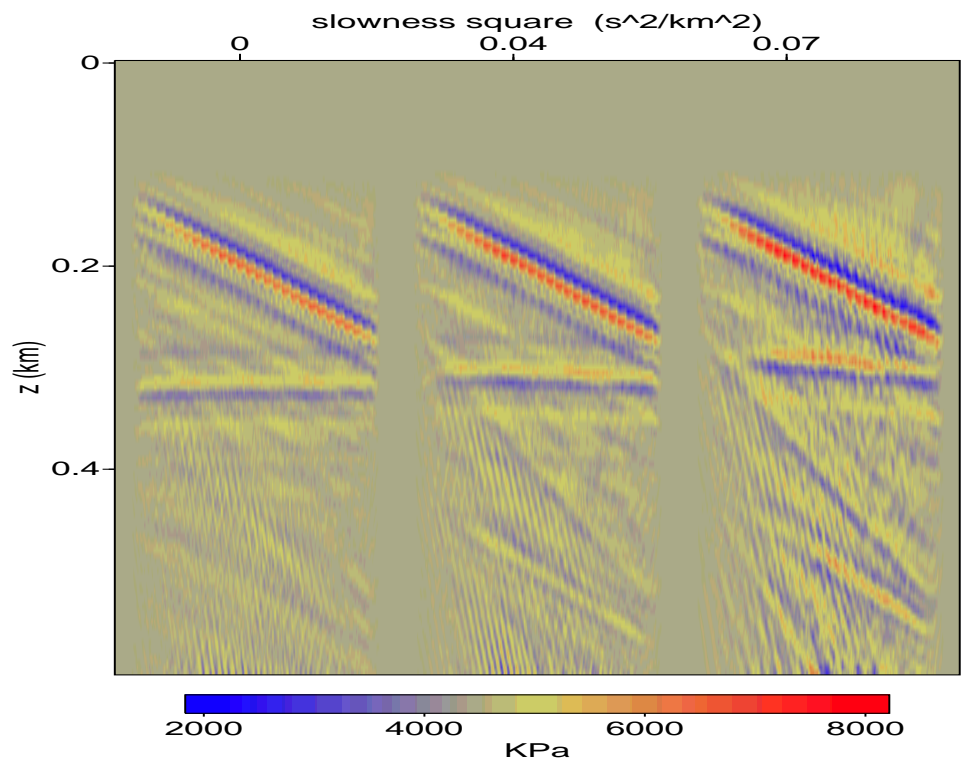
**Figure 3.26:** Gradient in 1st LS iteration and its gather at the middle offset

(a) 1st LS gradient

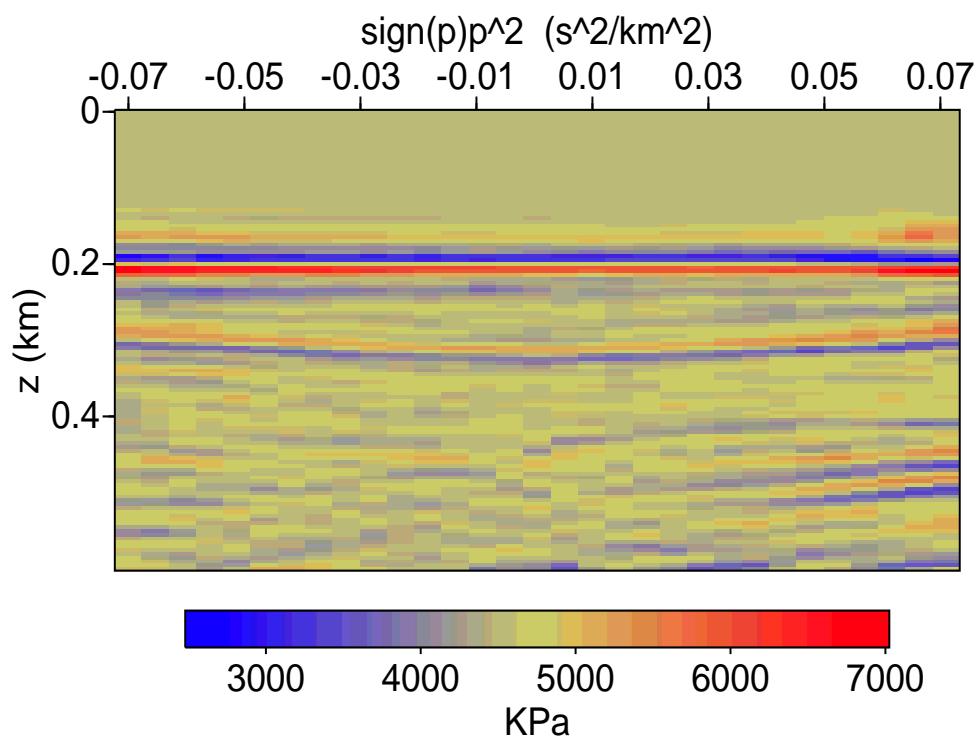


(b) Gather of 1st LS gradient at the middle offset

Figure 3.27: LS solution after 30 LBFGS iterations and its gather at the middle offset



(a) LS solution after 30 iterations

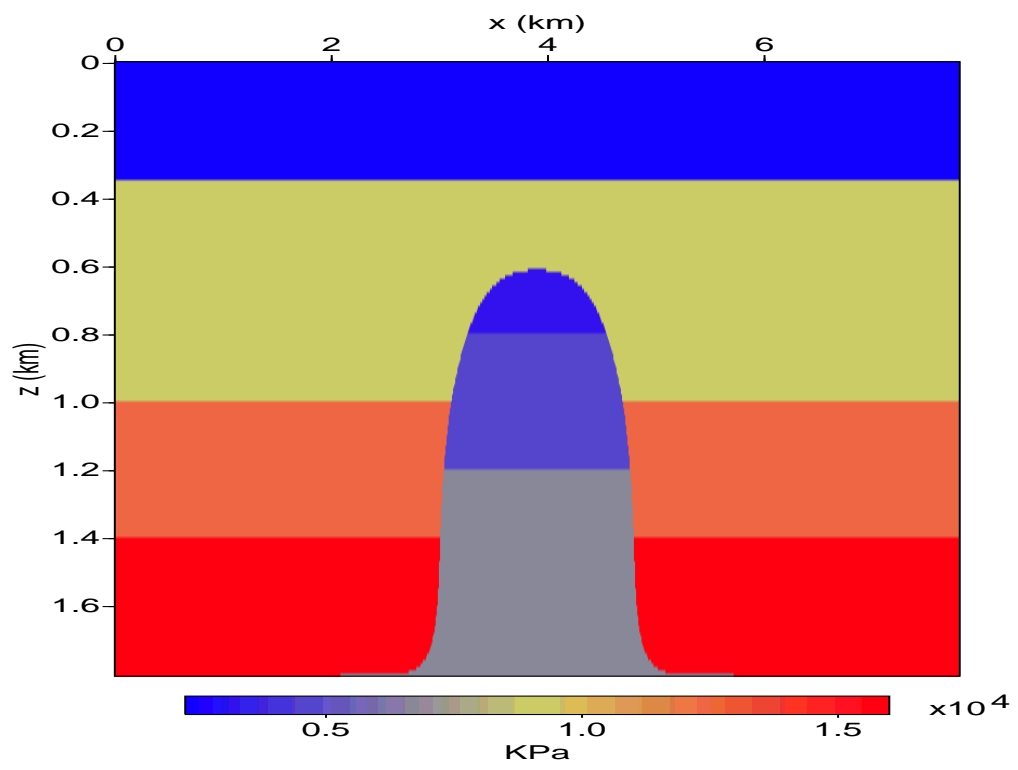


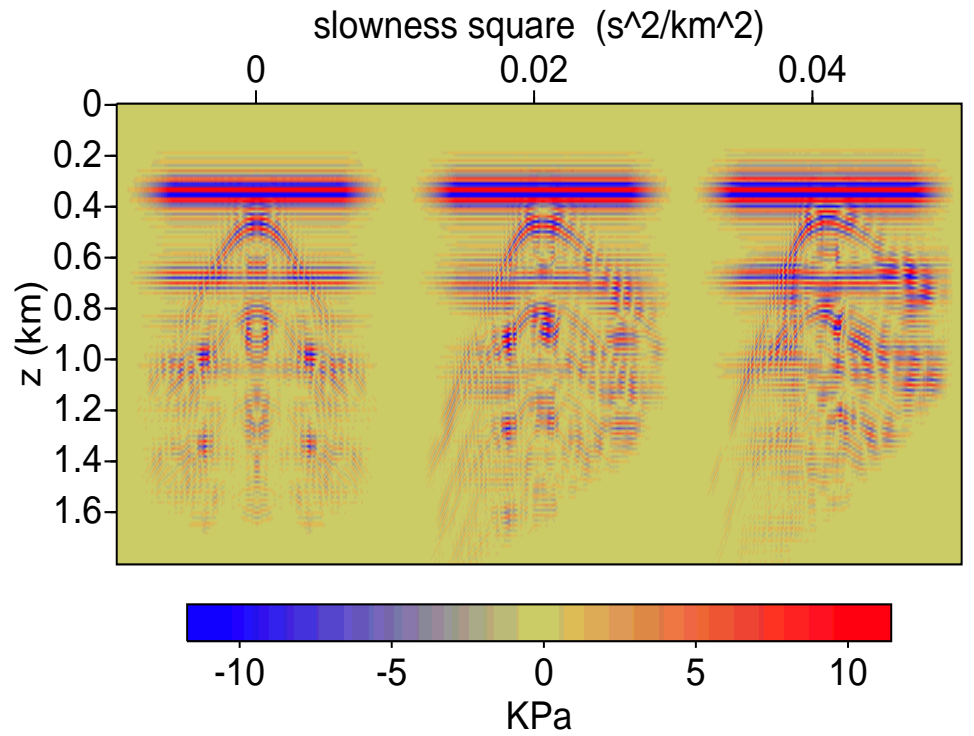
(b) Gather of the final inverted model at the middle offset

The third inversion experiment employs a 2D acoustic model with free surface, constant density  $2 \text{ g/cm}^3$  and dome bulk modulus model plotted in Figure 3.28. Starting from the homogeneous model (4500 KPa), this inversion consists of 30 LBFGS iterations. Clear multiple reflections are observed in the first LS gradient (i.e., the migrated image with respect to homogeneous background) and its middle offset (3000 m) gather shown in Figure 3.29. Those multiple reflections are substantially suppressed during this extended LS inversion, which is shown in the final LS solution and its middle offset gather plotted in Figure 3.30.

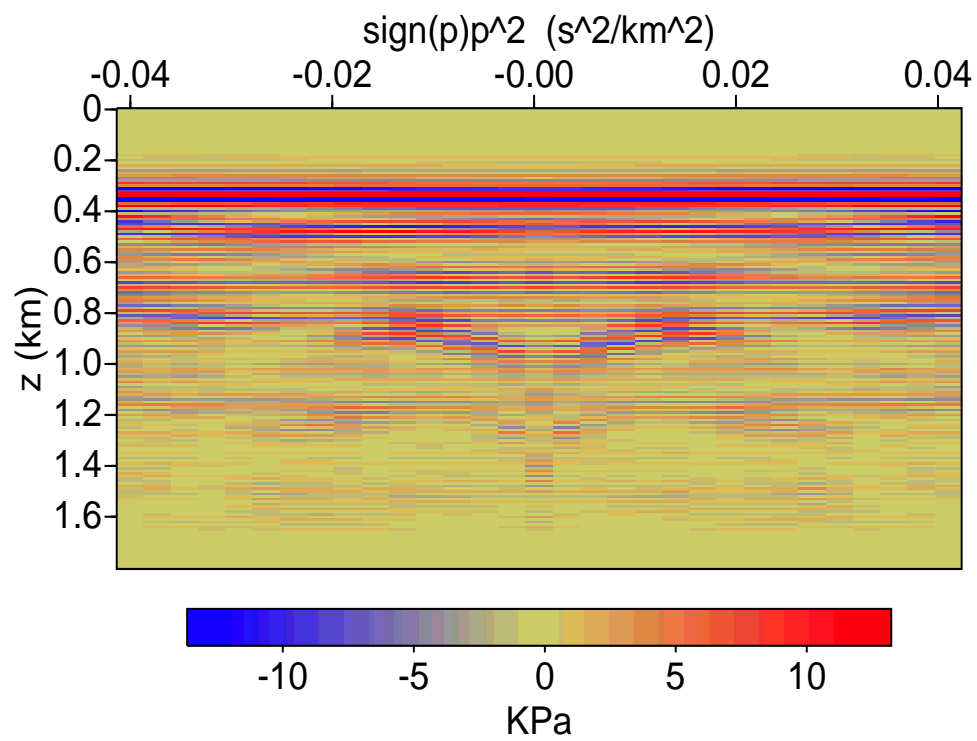
As a summary, the inversion experiments seem to suggest that it is in fact general that extended LS inversion is capable of suppressing multiple reflections. This feature of LS inversion makes it possible to measure primaries-only moveouts of the solution to an extended least-squares inversion but multiple ghosts, which definitely contributes to the working of proposed algorithm.

Figure 3.28: Dome bulk modulus model



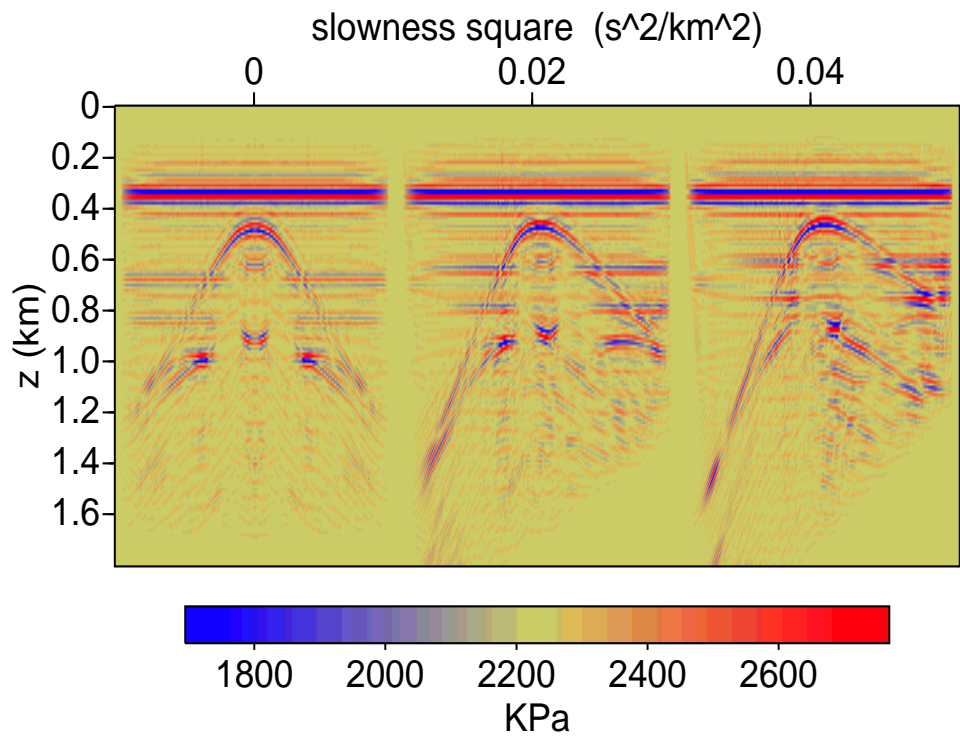
**Figure 3.29:** Gradient in 1st LS iteration and its gather at the middle offset

(a) 1st LS gradient

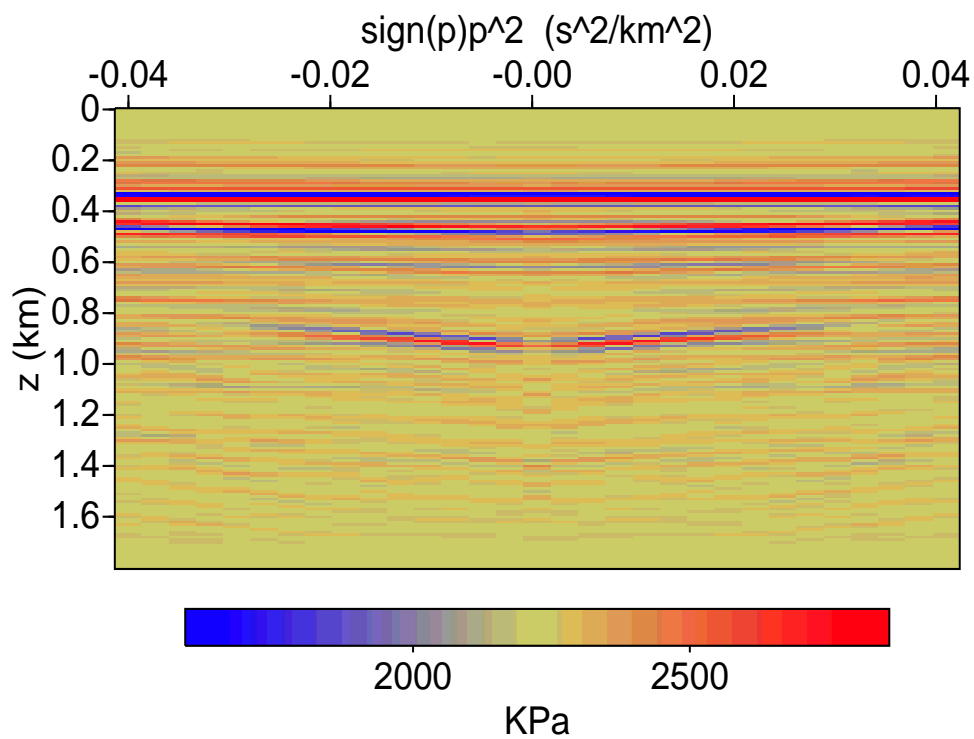


(b) Gather of 1st LS gradient at the middle offset

Figure 3.30: LS solution after 30 LBFGS iterations and its gather at the middle offset



(a) LS solution after 30 LBFGS iterations



(b) Gather of the final inverted model at the middle offset





# Chapter 4

## Discussion and Conclusion

In this thesis, I described and examined a nonlinear differential semblance optimization approach to waveform inversion for reconstructing the subsurface velocity model from reflection recordings. Usually, this type of data does not contain the very low frequency content, nor transmitted energy that is often present in large offset data acquired in wide-angle seismic surveys. Due to the high computational cost of generating the synthetic data, waveform inverse problem dictates the use of local methods based on the calculation of gradient-related descent direction of the misfit functional.

The traditional nonlinear least squares approach can't recover either the long or the medium scale structures, because of the local minima issue due to the interaction of the data spectrum incompleteness and the non-linearity of the least squares misfit functional with respect to model (especially to the low wavenumber content).

On the other hand, Migration velocity analysis, a widely adopted approach in industry, is capable of correcting velocity model at long scales. It uses prestack depth migration to produce an image volume depending on redundant acquisition

parameters and corrects the velocity via analyzing the coherence (semblance) condition of image volumes. With appropriate choice of coherence measurement objective functional, differential semblance strategy turns migration velocity analysis into an automated optimization procedure, for which local methods exhibit little tendency to stagnate at non-global minima. Since prestack depth migration is based on Born or single-scattering approximation (linearization assumption), migration velocity analysis could not account for nonlinear effects, such as multiple reflections.

Based on the extended modeling concept, this work provides a nonlinear generalization of migration velocity analysis that combines the ability of full waveform inversion to account for nonlinear physical effects, such as multiple reflections, with the tendency of differential semblance based migration velocity analysis to avoid local minima. In Chapter 2, I elaborate the reformulation of waveform inversion via the extended modeling concept as a constrained optimization problem aiming to find an extended model from the feasible set of models that satisfy data fitting criterion to minimize the coherence condition. Without reparametrizing the feasible set, the new problem is as difficult as the original one. The main contribution of this work is to re-parametrize the extended model space with the very low-frequency data  $d_l$  and associate the low-frequency data add-in with a controlling model  $m_l \in \mathbf{M}$  via an extra constraint  $d_l = F_l[m_l]$ . Thus, such a reformulation of waveform inversion adopts both the gather-flattening concept from MVA and the data-fitting concept from FWI. In this approach,  $m_l$  plays the same role as the macro model in migration velocity analysis; but no linearization and scale separation assumptions are required by formulation.

In Section 2.2, I examined such a reparametrization with scan tests, which show that the proposed differential semblance objective is convex along the scanning lines

in the low-frequency data space  $\mathbf{D}_l$ . Further, the numerical experiments in Chapter 4 demonstrate that starting from a crude initial guess, the proposed algorithm successfully infers long scale updates and converges to a global minimizer or at least provides an excellent initial estimate for standard FWI. As I have mentioned in Chapter 1, besides addressing the local minima issue, the proposed approach also aims to account for multiple reflections. The numerical experiments in Section 4.2 and 4.3 show that even with the presence of multiple reflections due to free surface, the proposed approach successfully recovers long scale structures and produces excellent initial models for standard FWI. On one hand, the successful experiments with free surface should thank to the fact that the formulation of this approach doesn't make any linearization assumption so that it is capable to account in a natural way for nonlinear effects. On the other hand, the suppression of multiple reflections during LS inversion is an important factor in the working experiments. In Section 4.4, I present three extended least squares inversion exercises on increasingly complex models, which suggest that it is in fact general that the extended least squares inversion is capable of scrubbing multiple reflections given enough iterations. This facilitates the measurement of primaries-only moveouts and contributes the most to the success of this strategy for the experiments with free surface. Hence, this algorithm in fact updates the velocity by flattening primaries-only gathers obtained via nonlinear inversion.

When it comes to the cost of this algorithm, it is expensive. Fortunately, there are strategies to reduce the computational cost. More specifically, this nonlinear differential semblance optimization consists of a few differential semblance iterations (3 to 7 DS-iterations computed in the presented experiments), in which the main computations are one extended least-squares inversion, gradient computation, and step-length computation. Various strategies exist or need to be studied in the future

to improve the efficiency of this approach. For example, as shown in the numerical experiments, the larger the incidence angle of the plane wave shot, the more detectable the moveouts are. Thus, to reduce the cost of extended least-squares inversion, one could only select a small subset of the acquisition parameter to extend the model and do inversion or sample the parameter in a more efficient way. During the gradient and step-length computation, one needs to solve a normal equation or a linear least-squares problem. To reduce the cost, one may refer to various approaches, such as, preconditioning, optimal scaling, etc., all of which are active research topics and worth of further effort. As contrast to reduce the computational cost, some of my other experiments observe that the more complex the model, the more accurate solutions to the extended LS problem and normal equation are needed to get descent directions. This leads to an important question: how accurately the extended LS problem and Normal equation need to be solved to guarantee decency? It is an critical future topic to investigate how the inexact solutions to the extended least-squares inversion and normal equation affect the whole minimization procedure. Until achieving at least a rough estimate of the accuracies needed for solving the extended least-squares inversion and the normal equation, one would have to overly solve those two expensive sub problems, which is another reason besides the limited computational resource that preventing the application of this algorithm to more complex models.

Another limitation of this algorithm is more fundamental. As this algorithm is based on surface oriented extension, it suffers from the common defect shared by all the surface-oriented image volume constructions, in that kinematic artifacts may occur when the velocity model is sufficiently complex such that there are multiple paths in wavefields that connecting source, receiver, and scattering points. The presence of these artifact can easily be as strong as the events corresponding to actual

reflectors, and thus restrict the use of this type of extensions in complex model. Hence, depth-oriented extension may be more appropriate in complex areas and thus in formulating the nonlinear differential semblance optimization for waveform inversion, which is discussed in Symes (2008). But, due to the enormous computational complexity of modeling for depth oriented extension, this work adopts surface oriented extension in order to examine the key low-frequency based reparametrization strategy, which is a general concept that will be used in a nonlinear DSO based on depth-oriented extension in future.

## APPENDIX A

### CALCULATION OF DS GRADIENT VIA ADJOINT STATE METHOD

Recall that the proposed DS optimization problem is

$$\begin{aligned} \min_{\substack{\eta(p,\omega) \\ (p,\omega) \in \tilde{\mathbf{D}}_l}} J_{DS} &:= \frac{1}{2} \|A[\bar{m}]\|^2 \\ \text{such that } &\| \overline{F}_w[\bar{m}](p, t) - d_o(p, t) - d_l[\eta](p, t) \| \simeq 0, \\ &p \in [0, p_{max}] \end{aligned}$$

where  $A[\bar{m}] := \frac{\partial \bar{m}}{\partial p}$ ,  $\tilde{\mathbf{D}}_l := \{(p, \omega) : 0 \leq p \leq p_{max}, |\omega| \leq \omega_l\}$ ,  $p$  is slowness,  $d_l$  are the artificial low-frequency controls that make up the missing very low-frequency data.

Now let's compute the gradient  $\nabla_\eta J_{DS}$ . Assume all the derivatives in the following computation exist. The computation consists of three steps.

Step 1. Compute  $\delta_\eta d_l$

Recall that

$$d_l(p, t) = \int_{|\omega| \leq \omega_l} d\omega e^{2\pi i \omega t} \eta(p, \omega).$$

Applying regular perturbation to the above equation, we have

$$\delta d_l(p, t) \approx \int_{|\omega| \leq \omega_l} d\omega e^{2\pi i \omega t} \delta \eta(p, \omega) = Y \delta \eta(p, \omega) \tag{A-1}$$

Step 2. Compute  $\delta_{d_l} \bar{m}$

Let

$$E[\bar{m}, d_l] := \frac{1}{2} \| \overline{F}_w[\bar{m}](p, t) - d_o(p, t) - d_l[\eta](p, t) \|_{\mathbf{D}}.$$

The first order development of  $E$  gives:

$$\begin{aligned}
\delta E &= \frac{1}{2} \left\{ \|\overline{F}_w[\bar{m} + \delta\bar{m}] - d_o - d_l\|_{\mathbf{D}}^2 - \|\overline{F}_w[\bar{m}] - d_o - d_l\|_{\mathbf{D}}^2 \right\} \\
&\approx \langle \overline{F}_w[\bar{m}] - d_o - d_l, D\overline{F}_w[\bar{m}]\delta\bar{m} \rangle_{\mathbf{D}} \\
&= \langle D\overline{F}_w[\bar{m}]^T (\overline{F}_w[\bar{m}] - d_o - d_l), \delta\bar{m} \rangle_{\mathbf{M}}.
\end{aligned}$$

The first order necessity of the least-squares subproblem gives:

$$D\overline{F}_w[\bar{m}]^T (\overline{F}_w[\bar{m}] - d_o - d_l) = 0.$$

Applying regular perturbation to the above equation, we have

$$D\overline{F}_w[\bar{m}]^T (D\overline{F}_w[\bar{m}]\delta\bar{m} - \delta d_l) \approx 0,$$

i.e.,

$$D\overline{F}_w[\bar{m}]^T D\overline{F}_w[\bar{m}]\delta\bar{m} \approx D\overline{F}_w[\bar{m}]^T \delta d_l.$$

Thus,

$$D_{d_l}\bar{m} = (D\overline{F}_w[\bar{m}]^T D\overline{F}_w[\bar{m}])^\dagger D\overline{F}_w[\bar{m}]^T \tag{A-2}$$

$$\delta_{d_l}\bar{m} = D_{d_l}\bar{m}\delta d_l \tag{A-3}$$

Step 3. Compute  $\delta_{\bar{m}}J_{DS}$



The first order development of  $J_{DS}$  gives:

$$\begin{aligned}
\delta_{\bar{m}} J_{DS} &= \frac{1}{2} \left\{ \left\| \frac{\partial \bar{m} + \delta \bar{m}}{\partial p} \right\|_{\mathbf{M}}^2 - \left\| \frac{\partial \bar{m}}{\partial p} \right\|_{\mathbf{M}}^2 \right\} \\
&\approx \left\langle \frac{\partial \bar{m}}{\partial p}, \frac{\partial \delta \bar{m}}{\partial p} \right\rangle_{\mathbf{M}} \\
&= - \left\langle \frac{\partial^2 \bar{m}}{\partial p^2}, \delta \bar{m} \right\rangle_{\mathbf{M}} + \int dz \left( \frac{\partial \bar{m}}{\partial p} \delta \bar{m} \right) \Big|_0^{p_{max}} \\
&= - \left\langle \frac{\partial^2 \bar{m}}{\partial p^2}, \delta \bar{m} \right\rangle_{\mathbf{M}}.
\end{aligned}$$

Thus,

$$\begin{aligned}
\delta_{\eta} J_{DS} &= - \left\langle \frac{\partial^2 \bar{m}}{\partial p^2}, \delta \bar{m} \right\rangle_{\mathbf{M}} \\
&= - \left\langle \frac{\partial^2 \bar{m}}{\partial p^2}, D_{a_i} \bar{m} \delta d_i \right\rangle_{\mathbf{M}} \\
&= - \left\langle (D_{a_i} \bar{m})^T \frac{\partial^2 \bar{m}}{\partial p^2}, \delta d_i \right\rangle_{\mathbf{D}} \\
&= - \left\langle (D_{a_i} \bar{m})^T \frac{\partial^2 \bar{m}}{\partial p^2}, Y \delta \eta \right\rangle_{\mathbf{D}} \\
&= - \int_{|\omega| \leq \omega_l} d\omega Y^T (D_{a_i} \bar{m})^T \frac{\partial^2 \bar{m}}{\partial p^2} \delta \eta
\end{aligned}$$

Since

$$\begin{aligned}
\delta_{\eta} J_{DS} &= \int_{|\omega| \leq \omega_l} d\omega \nabla_{\eta} J_{DS} \delta \eta, \\
\nabla_{\eta} J_{DS} &= Y^T (D_{a_i} \bar{m})^T \frac{\partial^2 \bar{m}}{\partial p^2}.
\end{aligned} \tag{A-4}$$

## APPENDIX B

### ONE-DIMENSIONAL LEAST-SQUARES INVERSION

In this appendix, I demonstrates the continuum low-frequency inversion procedure used in Section 2.2 to solve one-dimensional least-squares problems.

Recall that the 1-D wave equation

$$\frac{1}{v(z)^2} \frac{\partial^2 U}{\partial t^2} - \frac{\partial^2 U}{\partial z^2} = \omega(t) \delta(z),$$

together with appropriate initial and boundary conditions defines a forward map  $F_\omega[v] := \frac{\partial U}{\partial t}$ .

The least-squares problem is

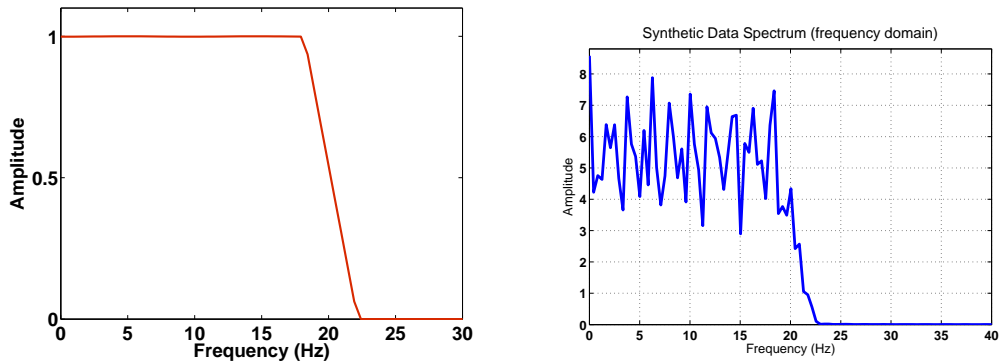
$$\min_v \frac{1}{2} \|F_\omega[v] - d\|^2,$$

where  $\omega(t)$  and  $d$  are known, and both of them have the very low-frequency information down to 0 Hz. Hence, I adopt the following procedure to solve the above least-squares problem.

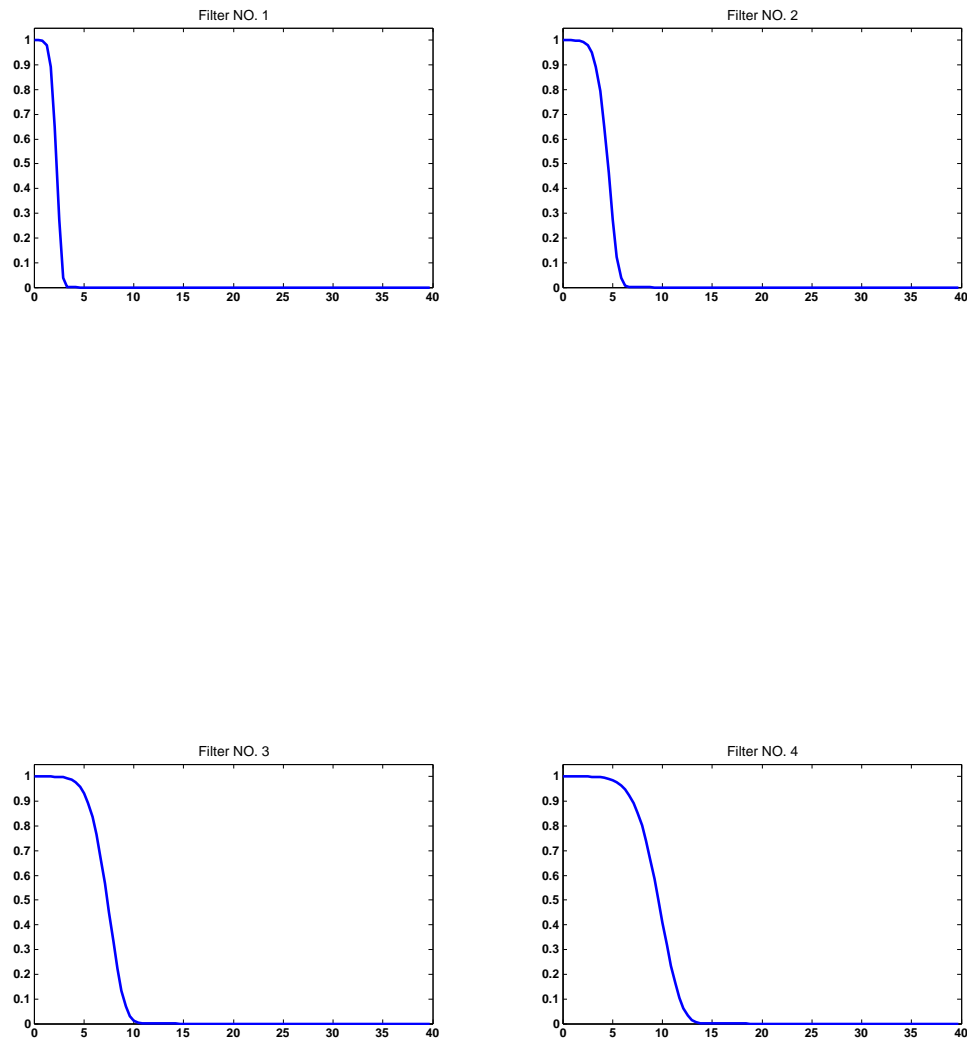
#### Continuum Low-Frequency Inversion:

- Split the whole inversion into multiple runs for source and data with increasing frequency bands, which is achieved by applying appropriate low-pass filters to both source and data;
- Regard the final estimate for the current run as the initial guess for the next run.

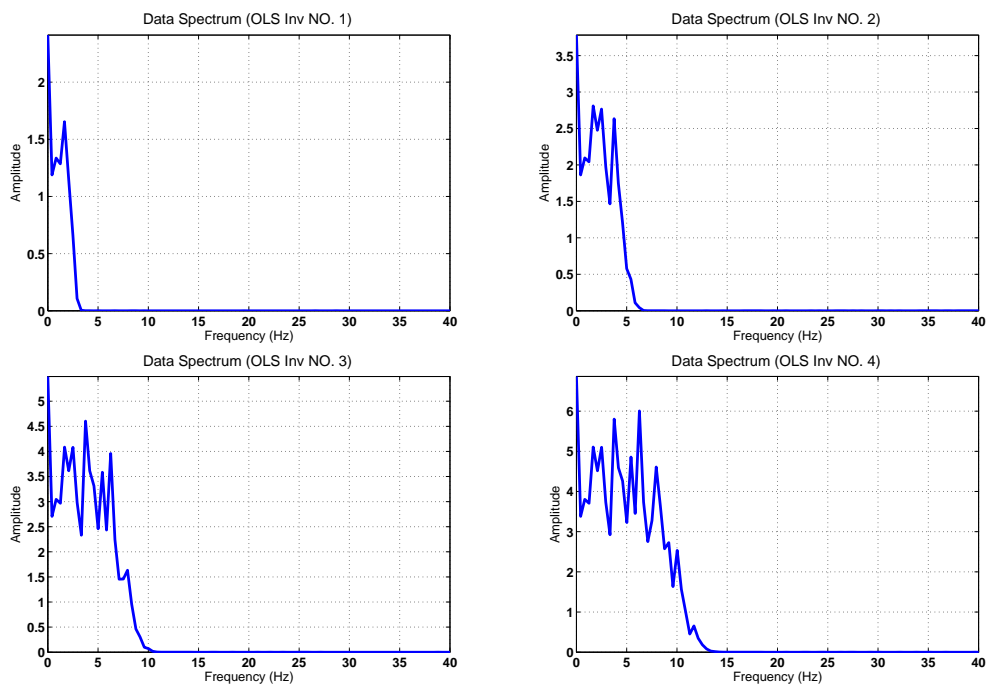
The following inversion for a four-layer model demonstrates the above procedure. The figure Figure B-1 shows the spectra of the given source and data. I applied four low-pass filters to the source and data, and generate four pairs of source and data with increasing frequency bands. The spectra of those four filters are shown in Figure B-2. The spectra of the four data with increasing frequency bands are shown in Figure B-3. Figure B-4 presents the models got from the four continuum low-frequency inversions. Figure B-5 is the model got from the final inversion, which used the whole band of the given source and data. Using only one inversion for the given data and source, I got the model shown in Figure B-6, which took much more computing time than the continuum low-frequency procedure, but led to much worse result. The data fitting performance for the two different inversion procedures is provided in Figure B-7.



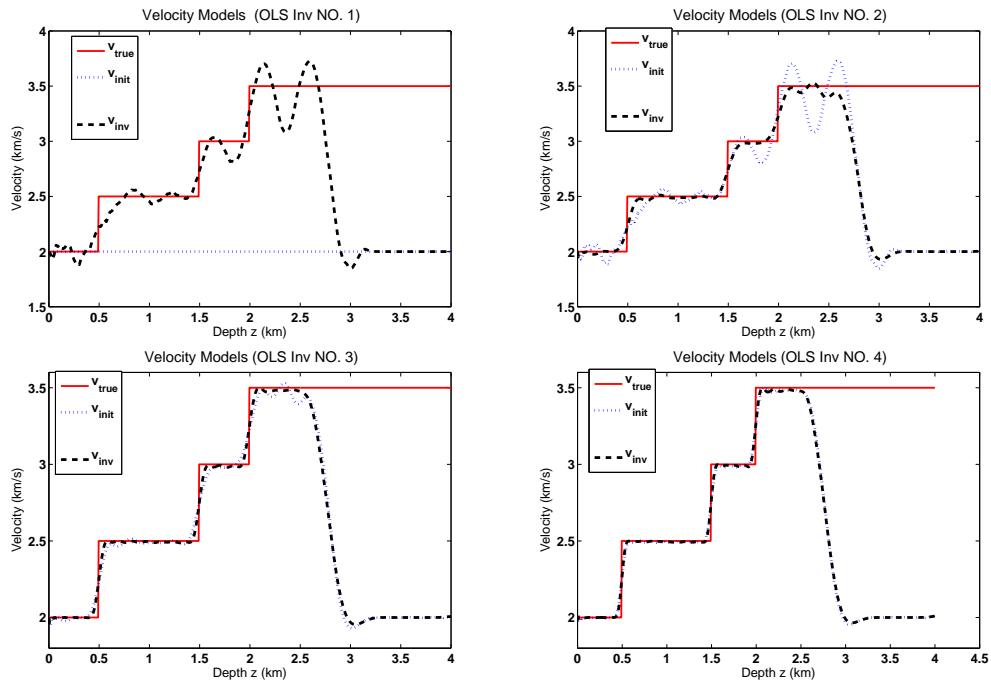
**Figure B-1:** Source Spectra (left one) and Data Spectra (right one)



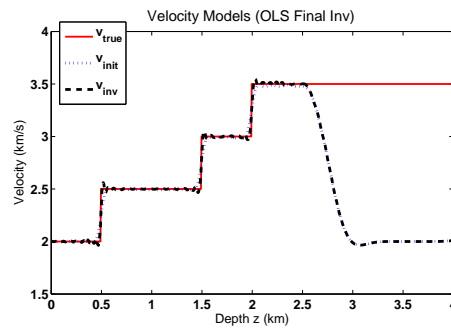
**Figure B-2:** Four low-pass filters used to filter the given source and data to provide the source and data with increasing frequency bands, which were used in the four continuum low-frequency inversions).



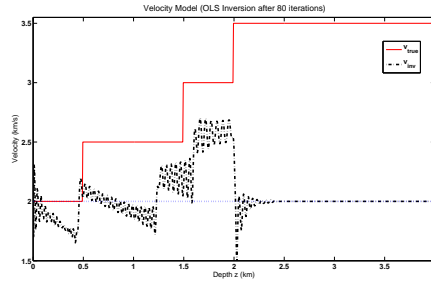
**Figure B-3:** Spectra of the data used in the four continuum low-frequency inversions (Left-top: first run, Right-top: run 2, Left-bottom: run3, Right-bottom: run4)



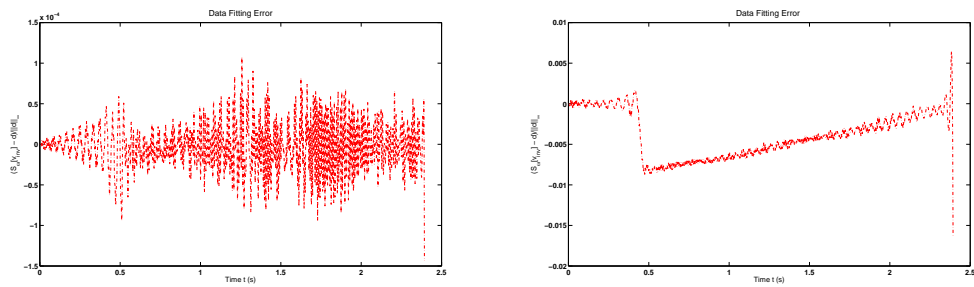
**Figure B-4:** Velocity models resulting from the four continuum low-frequency inversions (Left-top: first run, Right-top: run 2, Left-bottom: run3, Right-bottom: run4)



**Figure B-5:** Velocity model resulting from the final inversion, which used the source and data with the whole band



**Figure B-6:** Velocity Models resulting from the one-inversion procedure, which consists of only one inversion that used the source and data with the whole band.



**Figure B-7:** Relative Data Fitting Error from two inversion procedures: the left figure is for the continuum low-frequency inversion procedure, the right one is for the one-inversion procedure.

## REFERENCES

- Albertin, U., P. Sava, J. Etgen, and M. Maharramov, 2006, Adjoint wave equation velocity analysis: 76th Annual International Meeting, Expanded Abstracts, TOM2.1, Society of Exploration Geophysicists.
- Ali, B. H., O. H., S., J. Virieux, and F. Sourbier, 2009a, Efficient 3d frequency-domain full waveform inversion with phase encoding: Presented at the 71st Conference & Technical Exhibition, EAGE. Expanded Abstract.
- , 2009b, Three-dimensional frequency-domain full-waveform inversion with phase encoding: 79th Annual International Meeting, Expanded Abstracts, 2288–2292, Society of Exploration Geophysicists.
- Bamberger, A., G. Chavent, and P. Lailly, 1977, Etude mathématique et numérique d'un problème inverse pour l'Équation des ondes à une dimension: Rapport Interne 14, Centre de Mathématiques Appliquées, École Polytechnique, Paris.
- , 1979, About the stability of the inverse problem in 1-d wave equation — application to the interpretation of seismic profiles: Applied Mathematics and Optimization, **5**, 1–47.
- Brenders, A. and G. Pratt, 2007a, Efficient waveform tomography for lithospheric imaging: implications for realistic, 2-d acquisition geometries and low frequency data: Geophysical Journal International, **168**, 152–170.
- , 2007b, Full waveform tomography for lithospheric imaging: results from a blind test in a realistic crustal model: Geophysical Journal International, **168**, 133–151.
- Brenders, A. and R. Pratt, 2007c, Efficient waveform tomography for lithospheric imaging: implications for realistic, two-dimensional acquisition geometries and low-frequency data: Geophysical Journal International, **168**, 152–170.
- , 2007d, Full waveform tomography for lithospheric imaging: Results from a blind test in a realistic crustal model: Geophysical Journal International, **168**, 133–151.
- , 2007e, Waveform tomography of marine seismic data: What can limited offset offer?: Presented at the 2007 SEG Annual Meeting.
- Bube, K. and R. Burridge, 1983, The one dimensional inverse problem of reflection seismology: SIAM Review, **25**, 497–559.
- Bunks, C., F. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.
- Cao, D., S. Singh, and A. Tarantola, 1990, Simultaneous inversion for background velocity and impedance maps: Geophysics, **55**, 458–469.
- 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.
- De Hoop, M. V., S.-K. Foss, and B. Ursin, 2005, Depth-consistent reflection tomography using PP and PS seismic data: Geophysics, **70**, U51–U65.
- Ellefsen, K., 2009, A comparison of phase inversion and travelttime tomography for



- processing near-surface refraction traveltimes: *Geophysics*, **74**, WCB11–WCB24.
- Gauthier, O., A. Tarantola, and J. Virieux, 1986, Two-dimensional nonlinear inversion of seismic waveforms: *Geophysics*, **51**, 1387–1403.
- Hole, J., 1992, Nonlinear high-resolution three-dimensional seismic travel time tomography: *Journal of Geophysical Research*, **97**, 6553–6562.
- Kolb, P., F. Collino, and P. Lailly, 1986, Prestack inversion of a 1D medium: *Proceedings of the IEEE*, **74**, 498–506.
- Krebs, J. R., J. E. Anderson, D. Hinkley, R. Neelamani, S. Lee, A. Baumstein, and M.-D. Lacasse, 2009, Fast full-waveform seismic inversion using encoded sources: *Geophysics*, **74**, WCC177–WCC188.
- Lailly, P., 1983, The seismic inverse problem as a sequence of before-stack migrations, *in* Bednar, J., ed., *Conference on Inverse Scattering: Theory and Applications*, 206–220. Society for Industrial and Applied Mathematics.
- Li, J. and W. Symes, 2007, Interval velocity estimation via nmo-based differential semblance: *Geophysics*.
- Lines, L. and S. Treitel, 1984, A review of least-squares inversion and its application to geophysical problems: *Geophysical Prospecting*, **32**, 159–186.
- Min, D. and C. Shin, 2006, Refraction tomography using a waveform-inversion back-propagation technique: *Geophysics*, **71**, R21–R30.
- Minkoff, S. E. and W. W. Symes, 1997, Full waveform inversion of marine reflection data in the plane-wave domain: *Geophysics*, **62**, 540–553.
- Mulder, W. A. and A. P. E. ten Kroode, 2002, Automatic velocity analysis by differential semblance optimization: *Geophysics*, **67**, 1184–1191.
- Nocedal, J. and S. Wright, 1999, *Numerical Optimization*: Springer Verlag.
- Nolet, G., 1987, *Seismic tomography*: D. Reidel & Co.
- 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.
- Sacks, P. and F. Santosa, 1987, A simple computational scheme for determining the sound speed of an acoustic medium from its surface impulse response: *SIAM J. SCI. STAT. COMPUT.*, **8**, 501–520.
- Santosa, F. and W. Symes, 1989, An analysis of least-squares velocity inversion. Number 4 *in* *Geophysical monograph series*: SEG.
- Sen, M. and P. Stoffa, 1991a, Nonlinear multiparameter optimization using genetic algorithms: Inversion of plane wave seismograms: *Geophysics*, **56**, 1794–1810.
- , 1991b, Nonlinear one-dimensional seismic waveform inversion using simulated annealing: *Geophysics*, **56**, 1624–1636.
- Shah, N., M. Warner, T. Nangoo, A. Umpleby, I. Stekl, J. Morgan, and L. Guasch, 2012a, Quality assured full-waveform inversion: Ensuring starting model adequacy: Presented at the 82th Annual International Meeting, Society of Exploration Geophysicists.
- Shah, N., M. Warner, J. Washbourne, L. Guasch, and A. Umpleby, 2012b, A phase-

- unwrapped solution for overcoming a poor starting model in full-wavefield inversion: 74th EAGE Conference & Exhibition, Expanded Abstracts, P014, European Association of Geoscientists & Engineers.
- Shen, P. and W. W. Symes, 2008, Automatic velocity analysis via shot profile migration: *Geophysics*, **73**, VE49–60.
- Shen, P., W. W. Symes, S. Morton, and H. Calandra, 2005, Differential semblance velocity analysis via shot profile migration: 75th Annual International Meeting, Expanded Abstracts, SPVA1.4, Society of Exploration Geophysicists.
- Shen, P., W. W. Symes, and C. C. Stolk, 2003, Differential semblance velocity analysis by wave-equation migration: 73rd Annual International Meeting, Expanded Abstracts, 2135–2139, Society of Exploration Geophysicists.
- Shin, C. and Y. Cha, 2008, Waveform inversion in the laplace domain: *Geophysical Journal International*, **173**, 922–931.
- Shin, C. and D.-J. Min, 2006, Waveform inversion using a logarithmic wavefield: *Geophysics*, **71**, R31–R42.
- Sirgue, L., 2003, Inversion de la forme donde dans le domaine fréquentiel de données sismiques grands offsets: PhD thesis, Queens University, Canada.
- Sirgue, L., O. I. Barkved, J. P. V. Gestel, O. J. Askim, and J. H. Kommedal, 2009, 2d waveform inversion on valhall wide-azimuth obs: Presented at the 71st Conference & Technical Exhibition, EAGE.
- Sun, D., 2008, A nonlinear differential semblance algorithm for plane waves in layered media: Technical Report 09-04, Department of Computational and Applied Mathematics, Rice University.
- Sun, D. and W. W. Symes, 2010a, Iwave implementation of adjoint state method: Technical Report TR10-06, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, USA.
- , 2010b, Iwave implementation of born simulation: Technical Report TR10-05, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, USA.
- Symes, W., 1993, A differential semblance criterion for inversion of multioffset seismic reflection data: *J. Geoph. Res.*, **98**, 2061–2073.
- Symes, W. and J. Carazzone, 1992, Velocity inversion by coherency optimization: *Geophysical Inversion*, 59–89.
- Symes, W. and R. Versteeg, 1993, Velocity model determination using differential semblance optimization: 63rd annual international meeting: Expanded abstracts, 696-699, Society of Exploration Geophysicists.
- Symes, W. W., 1981, The inverse reflection problem for a smoothly stratified elastic medium: *SIAM Journal on Mathematical Analysis*, **12**, 421–453.
- , 1986, On the relation between coefficient and boundary values for solutions of webster’s horn equation: *SIAM Journal on Mathematical Analysis*, **17**, 1400–1420.
- , 1991, Layered velocity inversion: a model problem from reflection seismology: *SIAM Journal on Mathematical Analysis*, **22**, 680–716.

- , 1999, All stationary points of differential semblance are asymptotic global minimizers: layered acoustics: Technical Report 99-09, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, USA.
- , 2007, Migration velocity analysis and waveform inversion: Technical Report 07-05, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, USA.
- , 2008, Migration velocity analysis and waveform inversion: Geophysical Prospecting, **56**, 765–790.
- Symes, W. W. and J. J. Carazzone, 1991, Velocity inversion by differential semblance optimization: Geophysics, **56**, 654–663.
- Symes, W. W., D. Sun, and M. Enriquez, 2011, From modeling to inversion: Designing a well-adapted simulator: Geophysics Prospecting, **59**, 814833.
- Tarantola, A., 1984a, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266.
- , 1984b, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266.
- , 1987, Inverse Problem Theory: Elsevier.
- Tarantola, A., E. Crase, M. Jervis, Z. Konen, J. Lindgren, K. Mosegard, and M. Noble, 1990, Nonlinear inversion of seismograms: State of the art: 60th Annual International Meeting, Expanded Abstracts, S13.7, Society of Exploration Geophysicists.
- Tarantola, A. and B. Valette, 1982, Generalized nonlinear inverse problems solved using the least squares criterion: Reviews of Geophysics and Space Physics, **20**, 219–232.
- van Leeuwen, T. and F. J. Herrmann, 2012, Fast waveform inversion without source-encoding: Geophysical Prospecting, 1365–2478.
- Verm, R. and W. Symes, 2006, Practice and pitfalls in nmo-based differential semblance velocity analysis: 75th Annual International Meeting, Expanded Abstracts, SEG.
- Vigh, D. and E. W. Starr, 2008, 3d plane-wave full-waveform inversion: Geophysics, **73**, VE135–VE144.
- Vigh, D., W. Starr, J. Kapoor, and H. Li, 2010, 3d full waveform inversion on a Gulf of Mexico WAZ data set: 81st Annual International Meeting, Expanded Abstracts, 957–961, Society of Exploration Geophysicists.
- Virieux, J. and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, WCC1–WCC26.
- Zelt, C. and P. Barton, 1998, Three-dimensional seismic refraction tomography: A comparison of two methods applied to data from the faeroe basin: Journal of Geophysical Research, **103**, 7187–7210.