INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700  800/521-0600
RICE UNIVERSITY

On A Transmission Inverse Problem

by

Hua Song

A Thesis Submitted
in Partial Fulfillment of the
Requirements for the Degree

Doctor of Philosophy

Approved, Thesis Committee:

William W. Symes, Chairman
Professor of Computational and Applied Mathematics

Steve Cox
Associate Professor of Computational and Applied Mathematics

John Polking
Professor of Mathematics

Richard Tapia
Noah Harding Professor of Computational and Applied Mathematics

Houston, Texas
August, 1994
On A Transmission Inverse Problem

Hua Song

Abstract

In crosswell seismic experiments, seismic sources are fired in one well, and the wavefields generated are measured in another well. The goal of the crosswell seismology is to find physical parameters, especially the velocities, of the rocks between the wells from these measurements. This amounts to the mathematical problem of solving a coefficient inverse problem of the multidimensional acoustic wave equation. We consider two inversion methods in this thesis: traveltime inversion via traveltime tomography and waveform inversion via differential semblance optimization. The main results are obtained for traveltime tomography and differential semblance optimization under the non-caustic assumption. The main result for traveltime tomography is that the objective function is smooth, and vanishing gradient implies a global minimizer if the data is noise free; The main result for DSO approach is that the objective function is smooth and a critical point is kinematically close to the true velocity model if the noise level is low enough, the source wavelet is oscillatory enough, and the DSO parameter is small enough. We also discuss to some extent the two methods in the presence of caustics.
Acknowledgments

I would like to thank Professors Steve Cox, John Polking, and Richard Tapia for being on my Committee and their interests in my work. I am especially grateful to my advisor, Professor William Symes, for bringing me to Rice, and his guidance during the writing of this thesis.

Most of all, I wish to thank my parents, sisters, and brother for their constant support through the years, and my wife Lanzhi for her love, patience, and courage.

Finally, I am grateful to the financial support from Texas Geophysical Parallel Computation Project during my three year graduate study at Rice University.
Dedicated to my wife Lanzhi and son Kelei
Contents

Abstract ii
Acknowledgments iii
List of Illustrations vii

1 Introduction 1
  1.1 The Problem ........................................ 1
  1.2 Traveltime Tomography .............................. 2
  1.3 Waveform Inversion via Output Least Squares ......... 5
  1.4 Waveform Inversion via Differential Semblance Optimization .... 6
  1.5 The work of this thesis ................................ 7

2 Traveltime Tomography – Single Arrival 9
  2.1 The formulation ....................................... 9
  2.2 Smoothness ........................................... 13
  2.3 Gradient and A Characterization of Critical Points .... 17
  2.4 Hessian, Local Covexity ............................. 21
  2.5 Multiple Sources .................................... 22
3 Traveltime Tomography - Multiple Arrivals 28

3.1 Introduction ........................................ 28
3.2 A New Formulation by Lailly et. al. ......................... 29
3.3 The Beam-Forming Approach ................................ 31

4 Waveform Inversion via DSO - Single Arrival 39

4.1 Motivations and DSO Formulation .......................... 39
4.2 Smoothness of the Objective Function ........................ 44
4.3 The Gradient and Characterization of Critical Points .......... 50
4.4 Hessian and Local Convexity ................................ 59
4.5 The limit problem ($\sigma \rightarrow 0$) ............................. 62
4.6 Numerical Examples ........................................ 64

5 Waveform Inversion via DSO - Multiple Arrivals 73

5.1 Introduction ........................................ 73
5.2 Necessity of Branch-Separation ............................ 76
5.3 The forward mapping as a Fourier integral operator .......... 78
5.4 Microlocal source and microlocal cutoff operators .......... 84
5.5 The Beam-forming Idea ................................... 87

6 Concluding Remarks ........................................ 92
Illustrations

2.1 Experiment Visualization - One Source ........................................ 24
2.2 Ricker source wavelet .............................................................. 24
2.3 Seismogram. ................................................................. 25
2.4 Output least squares objective function. .................................. 25
2.5 Traveltime data. ................................................................. 26
2.6 Traveltime tomography objective function. ............................... 26
2.7 Experiment Visualization - Multiple Sources ............................. 27

3.1 Intersecting rays caused by a low velocity zone. Caustics are the
   envelope on the rays. ............................................................ 35
3.2 Wavefronts for the same velocity model as in Figure 3.1 ............. 35
3.3 First arrivals do not always constrain a low velocity zone ........... 36
3.4 Upper (first) group. ............................................................. 36
3.5 Center (second) group. ......................................................... 37
3.6 Lower (third) group. ............................................................ 37
3.7 The change of variable from $z_r$ to $z$ is multi-valued, and singular on
the caustics if $x$ is to the right of the cusp. ........................... 38

4.1 Ricker source wavelet .................................................. 67
4.2 Receiver-reparametrized source wavelet ............................ 67
4.3 Seismogram ............................................................... 68
4.4 The solid curve is DSO objective function with $\sigma = 1$. The dashed
curve is DSO objective function with $\textit{sigma} = 0$ .................. 68
4.5 OLS objective function .................................................. 69
4.6 Estimation of the source wavelet with a wrong velocity model ..... 69
4.7 Estimation of the source wavelet with the correct velocity model ... 70
4.8 Velocity model with a slow zone ..................................... 70
4.9 Seismogram produced with velocity model in Figure 4.8. The
minimum of the velocity is 1800 m/s. ................................. 71
4.10 DSO objective function with $\sigma = 1$. ............................. 71
4.11 Seismogram produced with velocity model in Figure 4.8. The
minimum of the velocity is 1500 m/s. ................................. 72
4.12 The solid curve is DSO objective function with $\sigma = 1$. The dashed
curve is OLS objective function. ......................................... 72

5.1 Using reflector to produce multiple arrivals ........................... 89
5.2 Seismogram ......................................................... 90
5.3 DSO objective function without branch-separation. .......... 91
5.4 DSO objective function with branch-separation. .............. 91
Chapter 1

Introduction

1.1 The Problem

The constant-density linear acoustic model of small amplitude wave motion in a fluid connects the pressure field \( p(x, y, z, t) \), the sound velocity field \( v(x, y, z) \), the density field \( \rho(x, y, z) \), and a body force divergence ("source") \( s(x, y, z, t) \) through the wave equation

\[
\frac{1}{v^2 \rho} \frac{\partial^2 p}{\partial t^2} - \nabla \left( \frac{1}{\rho} \nabla p \right) = s
\]  

(1.1)

with appropriate side conditions. We shall assume in this thesis that the source is isotropic and punctual, i.e., \( s(x, z, t) = f(t) \delta(x-x_0) \delta(y-y_0) \delta(z-z_0) \), and \( \rho \equiv 1 \). Other more complex classes of sources could be treated by methods similar to those proposed here.

This wave equation is a simple model of seismic wave motion in the earth, regarded as a fluid body (i.e., with shear motion and mode conversion neglected). The relation (1.1) between the sound speed \( v(x, y, z) \), the wavefield \( p(x, y, z, t) \), and the source \( s(x, y, z, t) \) is used in exploration seismology to find part of them by assuming the knowledge of some information about the rest of them. The problem we are concerned with is to determine \( v(x, y, z) \) and and the source wavelet \( f(t) \) from known location,
i.e., \((x_s, y_s, z_s)\), and the trace (measurements, or data) of the wavefield \(p(x, y, z, t)\) on a hypersurface in the earth. When the source and the receiver array are on the same surface, usually on the surface of the earth, and the data contain mostly reflected waves, the problem is the \textit{reflection inverse problem}; when the source and the receiver array are on different surfaces, for example two boreholes, and the velocity of part of the region between the two wells is to be determined, the problem is the \textit{crosswell inverse problem}, which will be studied in this thesis. We also call the later \textit{transmission inverse problem} if only transmission waves are considered, as in this thesis.

1.2 Traveltime Tomography

Crosswell inverse problems are important in reservoir characterization and enhanced oil recovery (Justice \textit{et al.}, 1989). The more or less conventional method to resolve these problems is \textit{traveltime inversion}. One extracts the traveltime information from the seismic data, and looks for a model to fit the traveltime data according to some criterion. When the least squares criterion is used, traveltime inversion is usually called \textit{traveltime tomography}.

Some theoretical analyses on linearized traveltime tomography from reflection data are found in Bube (1992) and Delprat-Jannaud and Lailly (1991). Bube studied the uniqueness of determining reflector depths and non-uniqueness of determining velocity using infinitely many sources and receivers. For that particular problem, he gave
a characterization of the null space of velocity, which has an immediate analogue for traveltime tomography in the crosswell setting. Hence for many reasonable experiment configurations, both the traveltimes as a function of velocity and its derivative as a function of velocity perturbation is not injective, and the least squares problem has many solutions when it has one. This situation occurs when the data is noise-free. Since this non-uniqueness is intrinsic to traveltime tomography, by a “solution” we will mean any velocity model which produces the right traveltimes, and we will say this velocity model is kinematically correct. Since the above mentioned least squares problem is ill-posed, some kind of regularization is need. Delprat-Jannaud and Lailly suggested to regularize the problem by asking more smoothness of the model, and studied approximations of the regularized problem in finite dimensional spaces. For nonlinear traveltime tomography, except the above mentioned non-uniqueness, very little is known theoretically.

As far as numerical experiments are concerned, a lot of work has shown that traveltime tomography behaves well and Newton-like methods converge in a large range of the model. For numerical experiments, see Bishop et al 1985, Chiu and Stewart 1987, Scales 1987, Herman 1992, and Lailly et al 1992. The main drawback of this approach to the inverse problem is that the traveltimes are not primary data of the seismic experiments. Traveltimes must be picked from the waveform data. This picking procedure is often manual, and can be difficult when the structure of the medium is complex. Phase errors are difficult to avoid in the presence of multi-
arrivals. One way to overcome this difficulty is suggested by Herman (1992) who replaces the arrival-time function by an “arrival-time-like” function which can be computed directly from waveform data.

All the above mentioned results assume either there is only one arrival time at a receiver location or only the first arrival is used in the presence of multiple arrivals. The difficulty in dealing with multiple arrivals lies in the fact that the traveltime data are parameterized by receiver location, hence the traveltime function is not single-valued when multiple arrivals are in presence. Lailly et al (1992) propose a new formulation of the traditional traveltime tomography which split the traveltime data into two groups: traveltimes and emerging locations, and reparameterizes these two sets of data with the ray parameters. These two functions are always single valued since for a single ray, there is at most one traveltme and emerging point at the receiver surface. The least squares criterion is suggested to match traveltimes and emerging locations for an estimated velocity model with those calculated from the traveltime data with the same ray parameter. The key point in this new formulation is that for a receiver location and a traveltme, the ray parameter of the corresponding ray can be easily calculated from the data as long as there are a densely placed sources. We have not seen any theoretical analysis on this formulation.
1.3 Waveform Inversion via Output Least Squares

Attempts to resolve the reflection inverse problem using waveform data directly are found in Tarantola 1986, Kolb et al 1986, Gauthier et al 1986, etc.. The idea is to find a model to fit the waveform data in the least-squares sense (output least squares (OLS)). Besides not requiring traveltime-picking, this approach also has some other attractive features, e.g., it can be modified easily to incorporate nonseismic constraints, and it possesses an elegant statistical justification (Tarantola, 1987). Unfortunately, the performance of the method using gradient (Newton-like) optimization is not as good as people would expect. A typical feature of this approach is that the objective function exhibits non-smooth and highly non-convex dependence on velocity trends, i.e., the part of the velocity model which determines the kinematics of the wave propagation. Hence it is impossible to locate the global minimizer using local optimization methods, i.e., quasi-Newton methods. The main reason for this feature is the fact that the long spatial wavelengths that would be sensitive to the background model, are missing from the seismic data. The remaining (short) wavelengths are too sensitive to the background model, since a small (half-wavelength) change in arrivaltime results in a $180^\circ$ phase shift, i.e., a 100% RMS change. Thus the mean square error saturates away from the optimum model. Some global remedies, e.g., genetic algorithms, Monte Carlo search, or simulated annealing, are suggested (Cao et al, 1990, Mosegard and Tarantola 1991, Sen and Stoffa 1991a, 1991b, Scales et al 1991) but are relatively costly.
1.4 Waveform Inversion via Differential Semblance Optimization

Considering the drawbacks of these two traditional inversion methods, a new approach is desirable which should keep the advantages of these methods, yet avoid their weaknesses. A variant of the OLS waveform inversion, *differential semblance optimization* (DSO), was proposed by Symes (1991a, 1991b) and studied both numerically and theoretically to some extent, for various reflection inverse problems (see Symes 1991c, 1992a and Symes and Carazzone 1991 for layered models and plane wave data, see Symes and Kern 1992 for 2D primary-only reflection acoustic models and shot gather (common shot) data, see Symes 1992b for a plane wave detection problem). A detailed account of the philosophy of DSO and its relations to the other methods are found in Symes and Kern (1992). DSO aims to overcome the multi-minima feature of the straightforward OLS waveform inversion, and the cumbersome nature of arrival-time-picking in the traveltime tomography. At the same time, DSO shares the advantages of both methods, i.e., it uses waveform data as in OLS waveform inversion, yet its objective function shares a lot of similarities with that of traveltime tomography. For crosswell seismic data involving caustics, Symes (1994) introduced an interesting idea to reformulate the DSO approach such that the nice features of the DSO objective function with caustic-free data are maintained.

In a reflection inverse problem, the earth model is divided into two components: the slowly-varying ("smooth") part, called "velocity", which is responsible for the
kinematics of the wave propagation; the rapidly-varying ("rough") part, called "reflectivity", which is responsible for the dynamics of the wave propagation. The wavefield depends linearly on the reflectivity but non-linearly on the velocity. The problem of the straightforward OLS approach is that its objective function depends non-smoothly and non-convexly on the velocity, as pointed out in the last subsection. The principle idea of DSO is to expand the space of reflectivity such that its members are shot-dependent, and add a penalty term to the OLS objective function to minimize this dependence since the reflectivity should be the same for all shots. This term is a measure of the differential of the reflectivity with respect to the shot location, i.e., a measure of semblance of the estimates of reflectivity for neighboring shots, hence the name of the method.

1.5 The work of this thesis

This thesis presents a theoretical analysis of the nonlinear traveltime tomography and the differential semblance formulation for the crosswell tomography problem of estimating velocity and source wavelet from crosswell waveform data, and some numerical examples. The main results are obtained under the very important simplifying assumption that no caustics are produced by the velocity structure, and these results are extended to some extent to the cases involving caustics. As shown in, for example, Lines et al (1992), caustics and guided waves are extremely common in crosswell
surveys. Our current study of caustics is not of immediate practical value, nonetheless it addresses the principal difficulty in seismic inversion.

First, we give a theoretical analysis of the nonlinear traveltime tomography in Chapter 2, assuming there is only one arrival. The main result is that under the assumptions that all rays do not turn vertically and do not cross each other, the TT objective function is smooth, and vanishing gradient implies a global minimizer if the data is noise free. This result is extended to a local formulation of traveltime tomography in the presence of caustics in Chapter 3, using the beam-forming idea. Then, in Chapter 4, we present a differential semblance formulation for the crosswell inverse problem of estimating velocity and source wavelet from crosswell waveform data. Our analysis shows that the gradient and hessian of the DSO objective function are closely related to those of the traveltime tomography. Hence local minimization methods, e.g., Newton-like methods, can be used to locate the global minimizer. Thus we can do velocity inversion using waveform data as effectively as using traveltimes, without using those very expensive global methods. Some numerical examples are presented to show the difference between OLS and DSO approaches.

We gather some miscellaneous results on crosswell wave propagations and waveform inversion of crosswell seismic data involving caustics in Chapter 5, including a local formulation of waveform inversion using the beam-forming idea. Finally, we conclude the thesis with some comments in Chapter 6.
Chapter 2

Traveltime Tomography – Single Arrival

2.1 The formulation

Typical crosswell seismic experiments involve at least two vertical wells in the earth. Sources (e.g., dynamite, air guns, etc.) are fired in some wells, and the wavefields generated are recorded in other wells. Suppose that there is only one source well and one receiver well, and only one source is fired in the source well (see Figure 2.1). If we use the notation \( X = (x, y, z) \) for a point in “flat” earth with \( z \) being the depth, \( v(X) \) for the velocity field, \( \rho(X) \) for the density field, \( t \) for time, \( p(X, t) \) for pressure field, \( f(t) \) for source wavelet, and \( X_s \) for a source location, then the pressure field generated by the source with source wavelet \( f(t) \) at location \( X_s \) is approximately governed by the acoustic wave equation

\[
\frac{1}{v(X)^2 \rho(X)} \frac{\partial^2 p}{\partial t^2} - \nabla \cdot \left( \frac{1}{\rho(X)} \nabla p \right) = f(t) \delta(X - X_s)
\]

where \( \delta \) is the Dirac delta function, with causality initial condition

\[
p(X, t) = 0 \quad \text{for } t << 0
\]

and an appropriate boundary condition. In order to make the idea simple, we will ignore the boundary, hence the reflections from the boundary. Since it is very under-
determined to find the function \( v \) from traveltime data, we assume that the source-receiver plane is perpendicular to the \( y \)-axis, and \( v \) is independent of \( y \). A consequence of this assumption is that all rays which start with directions within the source-receiver plane will always stay in this plane. This fact will be used later in our analysis of the gradient of the objective function.

With our non-caustic assumption about the velocity model, the retarded Green’s function is

\[
g(X, X_s, t) = a(X)\delta(t - \phi(X)) + \text{smoother terms}
\]

where \( \phi \) is the traveltime function which satisfies the eiconal equation

\[
|\nabla \phi|^2 = \frac{1}{v^2}
\]

and \( a \) is the amplitude which satisfies the transport equation

\[
\nabla a \cdot \nabla \phi + \frac{1}{2} a \Delta \phi = 0
\]

By convolving the source wavelet \( f \) and the retarded Green’s function in \( t \), we get an expansion of the wavefield

\[
p(X, t) = a(X)f(t - \phi(X)) + \text{smoother terms}
\]

Usually \( f \) is oscillatory, hence the first term in the above expansion contains most energy in the wavefield, and the data (seismogram) recorded at the receiver well will be approximated as

\[
p(X_r, t) = a(X_r)f(t - \phi(X_r))
\]
where $X_r = (x_r, y_r, z_r)$ is the receiver location, $x_r$ and $y_r$ are fixed, and $0 \leq z_r \leq Z$.

Since usually the source wavelet has a distinct peak, the graph of the data also has distinct peaks, sometimes they can be picked out, which gives the traveltime data. Traveltime tomography tries to find a velocity model which gives the best fit in traveltimes, using some criterion, e.g., least squares. Be aware that traveltimes are not always easy to pick. This is the shortcoming of this method and the main reason that other methods which use the raw (waveform) data directly might be desirable.

Let $\tau_d(z_r), 0 \leq z_r \leq Z$, be the traveltime data extracted from the seismogram:

$$\tau_d(z_r) = \phi_d(X_r),$$

where $\phi_d(X_r)$ is the traveltime field generated by the true velocity model, and $\tau[v](z_r), 0 \leq z_r \leq Z$, be traveltime for a given velocity $v$:

$$\tau[v](z_r) = \phi[v](X_r),$$

where $\phi[v](X_r)$ is the traveltime field generated by the velocity model $v$. A simple least squares formulation is

$$\min_{v \in V} J[v]$$

where $V$ is a bounded set of velocities which will be defined below, and

$$J[v] = \frac{1}{2} \|\tau[v] - \tau_d\|_{L^2([0, Z])}^2$$

We will study the smoothness and the gradient and hessian of the objective function. The cases involving multiple wells, multiple sources, can be studied similarly.
We will use the following finite dimensional set for velocity, which simplifies the analysis without hurting our goal. We take a linearly independent subset \( \{v_i : i = 1, 2, ..., n\} \) of \( C^\infty(\Omega) \), where \( \Omega = [x_s, x_r] \times [0, Z] \), here \( v_i, i = 1, ..., n \), are positive and bounded above and below away from zero, and define a set

\[
V = \{ \sum_{i=1}^{n} c_i v_i : c_i \in I_i \subset R, i = 1, 2, ..., n \}
\]

where \( I_i, i = 1, 2, ..., n \), are bounded intervals such that for some constants \( 0 < v_{\text{min}} < v_{\text{max}} \), \( v \in V \) implies \( v_{\text{min}} \leq v(x, y, z) \leq v_{\text{max}} \). We also assume that \( V \) contains the velocity model \( v^* \) which produces the travelt ime data \( \tau_d \), and that \( v_i \)'s and \( I_i \)'s are chosen in such a way that for each \( v \in V \), all the rays connecting the source location and the receiver locations stay in the region \( (x_s, x_r) \times [0, Z] \) and do not intersect.

The existence of such a set \( V \) will become clear in the proof of smooth dependence of the travelt ime function on velocity. Under these assumptions, \( a[v] \) and \( \tau[v] \), as solutions of the eiconal equation and transport equation respectively, are uniformly bounded above and below by positive constants, and their derivatives are uniformly bounded above. We regard all the functions of the velocity as defined on \( V \), and by smoothness of their dependence on \( v \) we mean they are smooth as functions of \( c = (c_1, c_2, ..., c_n) \in R^n \) through \( v \), although we will not treat them as functions of \( c \) explicitly. Implicitly we identify \( V \) as subset of \( R^n \) and equip it with the euclidean norm.
2.2 Smoothness

All the analysis in this section depends on the theory of propagation of singularities.

An indispensable tool in this theory is the Hamilton system. For the purpose of later use, we introduce one version of the Hamiltonian system governing the rays associated to the acoustic wave propagation and some terminology. Let the Hamiltonian be

\[ H(x, z, t, \xi, \eta, \tau) = \frac{1}{2} \{ v(x, z)^2 (\xi^2 + \eta^2 + \zeta^2) - \tau^2 \}. \]

Then the Hamiltonian system is

\[
\begin{align*}
\dot{x} &= \partial_x H = v(x, z)^2 \xi \\
\dot{y} &= \partial_y H = v(x, z)^2 \eta \\
\dot{z} &= \partial_z H = v(x, z)^2 \zeta \\
\dot{t} &= \partial_t H = -\tau \\
\dot{\xi} &= -\partial_x H = -v(x, z) \partial_x v(x, z)(\xi^2 + \eta^2 + \zeta^2) \\
\dot{\eta} &= -\partial_y H = -v(x, z) \partial_y v(x, z)(\xi^2 + \eta^2 + \zeta^2) \\
\dot{\zeta} &= -\partial_z H = -v(x, z) \partial_z v(x, z)(\xi^2 + \eta^2 + \zeta^2) \\
\dot{\tau} &= -\partial_t H = 0
\end{align*}
\]

where "\( \dot{\ } \)" means \( \frac{d}{ds} \) and \( s \) is a parameter along a ray. The trajectories of this ordinary differential system are called \textit{bicharacteristic strips}. It is easy to check that \( H \) is constant on the bicharacteristic strips. The bicharacteristic strips on which \( H = 0 \) are called \textit{null-bicharacteristic strips}. The projections of the null bicharacteristic
strips into the \((x, y, z)\) space are called \((characteristic)\) rays. Since \(v\) is independent of \(y\) in our case, from the sixth and the second equations we know that the rays which start with directions in the \(x - z\) plane will stay in the plane forever, and for these rays, \(\eta = 0, y = 0\). Hence we can treat the problem as a two-dimensional one, and delete the second and sixth equations. From the last equation in the above system we see that \(\tau\) is a constant, and we only consider the problem forward in time, we can take \(\tau = -1\). Now the Hamiltonian system becomes

\[
\begin{align*}
\frac{dx}{dt} &= v(x, z)^2 \xi \\
\frac{dz}{dt} &= v(x, z)^2 \zeta \\
\frac{d\xi}{dt} &= -v(x, z) \partial_x v(x, z)(\xi^2 + \zeta^2) \\
\frac{d\zeta}{dt} &= -v(x, z) \partial_x v(x, z)(\xi^2 + \zeta^2)
\end{align*}
\]

which in fact is the Hamiltonian system with Hamiltonian \(\frac{1}{2} \{ v(x, z)^2 (\xi^2 + \zeta^2) - 1 \}\).

Under our assumption that no ray turns vertically between the two wells, \(\xi = \frac{dx}{dt}\) is never zero between the two wells. Hence we can reparametrize the rays by \(x\). Also note that for the null-bicharacteristics,

\[\xi^2 + \zeta^2 = \frac{1}{v(x, z)^2}\]

The fact \(\dot{x} \neq 0\) also implies that \(\xi\) does not change sign. Since for all rays connecting the source and receivers, the initial values of \(\xi\) are positive, we can solve \(\xi\) from the identity above. Hence

\[\xi = \sqrt{1/v(x, z)^2 - \zeta^2}\]
The Hamiltonian system above can then be rewritten as

\[
\frac{dt}{dx} = \frac{1}{v(x, z)^2 \sqrt{1/v(x, z)^2 - \zeta^2}}
\]

\[
\frac{dz}{dx} = \frac{\zeta}{\sqrt{1/v(x, z)^2 - \zeta^2}}
\]

\[
\frac{d\zeta}{dx} = -\frac{v_z(x, z)}{v(x, z)^3 \sqrt{1/v(x, z)^2 - \zeta^2}}
\]

Given initial conditions, say,

\[ t(0) = t_0 \]

\[ z(0) = z_0 \]

\[ \zeta(0) = \zeta_0 \]

we can trace the ray with the starting position \((x_0, z_0)\) and starting direction \((\zeta_0, \zeta_0) = (\sqrt{1/v(x_0, z_0)^2 - \zeta_0^2}, \zeta_0)\) through the medium by solving the initial value problem of the Hamilton system.

If the receiver is located at \((x_r, z_r)\), then solving a two-point boundary value problem, if it is solvable, with boundary conditions

\[ t(0) = 0 \]

\[ z(0) = z_s \]

\[ z(x_r) = z_r \]

gives traveltime at the receiver location \((x_r, z_r)\):

\[ \tau[v](z_r) = t(x_r) \]
In order to use the smooth dependence on parameters and initial values results of initial value problems of ODEs, we try to find an initial value for $\zeta$ such that the initial value problem with initial conditions

\[ t(0) = 0 \]
\[ z(0) = z_s \]
\[ \zeta(0) = \zeta_s \]

is equivalent to the above two point boundary value problem. Define map

\[ F(\zeta(0)) = z(x_r) \]

where $z(x_r)$ is part of the solution to the two point boundary value problem. The smooth invertibility of this map corresponds to the absence of caustics.

Apply the smooth dependence result of solutions of ODEs on parameters (Lang 1993, Chapter XIV, Theorem 4.3), we get

**Lemma 2.1** \( \tau[v](z_r) \) is \( C^p, p \geq 1 \) is an integer, as a function of \( v \in V, z_r \in [0, Z] \) in a neighborhood of \( v^* \).

Consequently, we have

**Theorem 2.1** \( J[v] \) is smooth in \( v \in V \).
2.3 Gradient and A Characterization of Critical Points

In this subsection, we investigate the gradient of the objective function $J$, and give a characterization of the critical points. In order to analyze the gradient of $J$, we need the following lemmas.

**Lemma 2.2** The directional derivative of $\tau$ with respect to $\nu$ in direction $\delta \nu$ is

$$D\tau[\nu] \delta \nu(z_r) = \oint_{X_s} \left( -\frac{\delta \nu}{\nu^2} \right) (x(s), z(s)) ds,$$

where the integration path is the raypath from $X_s$ to $X_r$, and $s$ is the raypath-length from the source location $X_s$.

**Proof** From the Eiconal equation

$$|\nabla \phi[v]|^2 = \frac{1}{\nu^2},$$

we have

$$\nabla D \phi[v] \delta \nu \cdot \nabla \phi[v] = -\frac{\delta \nu}{\nu^3},$$

i.e.,

$$\nabla D \phi[v] \delta \nu \cdot (\nu \nabla \phi[v]) = -\frac{\delta \nu}{\nu^2}.$$

Since $\nu \nabla \phi[v]$ is the unit tangential direction along the ray, we have

$$D \phi[v] \delta \nu(X) = \oint_{X_s} \left( -\frac{\delta \nu}{\nu^2} \right) (x(s), z(s)) ds,$$
hence,

$$D\tau[v]\delta v(z_r) = D\phi[v]\delta v(X_r)$$

$$= \int_{X_r}^{X_r} \left( -\frac{\delta v}{v^2} \right) (x(s),z(s)) ds.$$

Q.E.D.

Now we analyze the gradient. As a preparation result, first we prove that, unlike $D\tau[v]$, its $L^2$ adjoint $(D\tau[v])^*$ is injective under our assumptions about $V$. In fact, we have

**Lemma 2.3** The adjoint operator $(D\tau[v])^*$ maps $L^2([0,Z])$ into $L^2(\Omega)$,

and there exists a positive constant $C_0$ such that for any $v \in V$ and any $g \in L^2([0,Z])$,

$$\|(D\tau[v])^*g\| \geq C_0\|g\|.$$  \hspace{1cm} (2.1)

**Proof** First we derive the formula for $(D\tau[v])^*g$. For any $z_r \in [0,Z]$, there is a ray connecting the source location $X_s$ and the receiver location $X_r$. For any point on the ray with x-coordinate $x \in [x_s,x_r]$, we denote its z-coordinate by $z(x;z_r)$. Then we have

$$\int_0^Z (D\tau[v])(z_r)g(z_r)dz_r = \int_0^Z \left\{ \int_{X_s}^{X_r} \left( -\frac{\delta v}{v^2} \right) (x(s),z(s)) ds \right\} g(z_r)dz_r$$

$$= \int_0^Z \left\{ \int_{x_s}^{x_r} \left( -\frac{\delta v}{v^2} \right) (x,z(x;z_r))J_1(x;z_r) dx \right\} g(z_r)dz_r$$

where $J_1(x;z_r)$ is the jacobian of the change of variable from $s$ to $x$ on the ray connecting the source location $X_s$ and the receiver location $X_r$. Now make another
change of variable from \( z_r \in [0, Z] \) to \( z \) for each fixed \( x \), we get

\[
\int_0^Z (D\tau v)(z_r) g(z_r) dz_r = \int_0^Z \int_{x_r}^{x_r'} \left( -\frac{\delta v}{v^2} \right) (x, z) J_1(x; z_r(z; x)) g(z_r, (x, z)) J_2(z; x) dx dz
\]

(2.2)

where \( J_2 \) is the jacobian of the change of variable from \( z_r \) to \( z \) for a fixed \( x \). Note that this change of variables is defined only in the part of the region covered by rays.

Now it is clear that \((D\tau[v])^*\) is defined by

\[
(D\tau[v])^* g(x, z) = \left( -\frac{1}{v^2} \right) (x, z) J_1(x; z_r(z; x)) g(z_r, (x, z)) J_2(z; x) \chi(x, z)
\]

where \( \chi \) is the characteristic function of the region covered by rays. This formula is interpreted as spreading the value of \( g \) at \( z_r \) along the ray connecting \( X_r \) and \( X_s \), or "backward projection", on the part covered by rays. Hence,

\[
\| (D\tau[v])^* g \|^2 = \int_0^Z \int_{x_r}^{x_r'} \left( (D\tau[v])^* g(x, z) \right)^2 dx dz
\]

\[
= \int_0^Z \int_{x_r}^{x_r'} \left\{ \left( -\frac{1}{v^2} \right) (x, z) J_1(x; z_r(z; x)) g(z_r, (x, z)) J_2(z; x) \chi(x, z) \right\}^2 dx dz.
\]

Make an inverse change of variable from \( z \) to \( z_r \) for each fixed \( x \), we get

\[
\| (D\tau[v])^* g \|^2 = \int_0^Z \int_{x_r}^{x_r'} \left\{ \left( -\frac{1}{v^2} \right) (x, z(z_r; x)) J_1(x; z_r) g(z_r) J_2(z(z_r; x); x) \right\}^2 dx dz,
\]

We know that \( v \) is bounded from below and above, and \( J_1(x; z_r) = |\frac{dz_r}{dx}|^{-1} \) is bounded below. The "non-caustic" assumption implies that \( J_2(z(z_r; x); x) \) is bounded below. Hence, there is a constant \( C_0 \), which is independent of \( g \), such that

\[
\| (D\tau[v])^* g \| \geq C_0 \| g \|.
\]
Let $\tau_d(z_r)$ be the noise-free traveltime data extracted from the waveform data $S_{data}(z,t)$ generated by velocity model $v^*$, i.e., $\tau_d = \tau[v^*]$. It can be calculated that the gradient of $J[v]$ is defined by

$$ DJ[v] \delta v = \langle D\tau[v] \delta v, \tau[v] - \tau[v^*] \rangle = \langle \delta v, (D\tau[v])^*(\tau[v] - \tau[v^*]) \rangle $$

Hence the $L^2$ gradient of $J$ is

$$ \text{Grad} J[v] = (D\tau[v])^*(\tau[v] - \tau[v^*]) $$

Thus, according to the previous lemma,

$$ \| \text{Grad} J[v] \| \geq C \| \tau[v] - \tau[v^*] \| $$

where $C$ is independent of $v$, under the non-caustics assumption. This means that all critical points of $J$ produce traveltimes which fit the data exactly, hence they belong to the set of all kinematically correct velocity models. It must be noticed that, with a single, or sometimes even with multiple sources, one can not expect this set to be a singleton, or in other words, the "null space" of the traveltime tomography to be empty (see the first section). We summarize the above discussion as

**Theorem 2.2** All local extrema interior to $V$ of $J[v]$ are global minimizers, and they are kinematically correct velocities.

We use a simple numerical example to demonstrate the properties of the objective function and the difference between traveltime tomography and output least squares
waveform inversion. The distance between the two wells is 100 meters. Source is located at the depth 125 meters and 21 receivers are placed evenly in the receiver well, 12.5 meters apart, from the surface of the earth down to depth 250 meters. The sample interval is .25 mini seconds, and there are 501 samples per trace. Fig. 2.2 is a Ricker source wavelet with peak frequency 500 hz. Fig. 2.3 is the seismogram using the above source wavelet and the homogeneous velocity model $v = 2000m/s$. Fig. 2.4 is the output least squares objective function. Fig. 2.5 is traveltime data from the same experiment. Fig. 2.6 is traveltime tomography objective function, which is smooth and convex, hence local minimization methods, i.e., Newton-like methods, can be used to locate the global minimizer.

2.4 Hessian, Local Covexity

It can be calculated that

$$D^2J[v](\delta v, \delta v) = \langle D\tau[v]\delta v, D\tau[v]\delta v \rangle + \langle D^2\tau[v](\delta v, \delta v), \tau[v] - \tau[v^*] \rangle$$

$$= \langle (D\tau[v])^*(D\tau[v])\delta v, \delta v \rangle + \langle D^2\tau[v](\delta v, \delta v), \tau[v] - \tau[v^*] \rangle$$

The non-negative definite operator $(D\tau[v])^*(D\tau[v])$ is not positive definite since $D\tau[v]$ is not injective as mentioned in the first section. It has zero eigenvalues. For a subspace $D_vV_p$ of $D_vV$ (tangent space of $V$ at $v$) in which this operator is positive definite for all $v$ in a neighborhood $B_v*$ of $v^*$, i.e., there exists a positive number $\lambda$ such that

$$\langle (D\tau[v])^*(D\tau[v])\delta v, \delta v \rangle \geq \lambda \|\delta v\|^2, \forall \delta v \in D_vV_p, \forall v \in B_v*$$
then if \( v \) is close enough to \( v^* \), \( \tau[v] - \tau[v^*] \) will be small enough so the second term in \( D^2 J[v](\delta v, \delta v) \) will be dominated by the first term. Hence, \( J[v] \) is convex in certain velocity perturbation directions and flat in others in a neighborhood of \( v^* \).

It is the task of resolution analysis to characterize the directions in which \( J[v] \) is convex, and to find experiment settings which make the subspace \( D_v V_p \) as large as possible.

### 2.5 Multiple Sources

If more than one sources are used in the sources well, we can formulate and analyze the traveltime inversion in a similar fashion. We parametrize a source by an index \( s \in S \), where \( S \) is the finite set of all source indices. We still use the velocity space \( V \) defined in Section 2.1, and assume that the assumptions for the single source case in Section 2.1 holds for each source in the multiple source case. Let \( \tau^s_d \) be the traveltime data generated by the source located at \( X_s \), and let \( \tau^s[v] \) be simulated traveltime data with velocity \( v \) and the source located at \( X_s \). The traveltime inversion is formulated as a minimization problem with objective function

\[
J[v] = \frac{1}{2} \sum_{s \in S} \| \tau^s[v] - \tau^s_d \|^2
\]

Similar calculation as in Section 2.3 leads to

\[
\text{Grad} J[v](x, z) = \sum_{s \in S} (D\tau^s[v])^* (\tau^s[v] - \tau^s_d)
\]

\[
= \sum_{s \in S} \left( \frac{1}{v^2} - \frac{1}{v^2} \right) (x, z) J^s_1(x; z_r(x, z)) J^s_2(z, x) (\tau^s[v] - \tau^s_d)(z_r(x, z)) \chi'(x, z)
\]
where $J_1^*$ and $J_2^*$ are the two jacobians in the proof of Lemma 2.2, respectively, for the source at $X_s$, and $\chi^s$ is the characteristic function of the region covered by rays connecting the source location $X_s$ and all the receiver locations (see Fig 2.7). Since these regions have disjoint subregions near the source well, we can find $x_c \in (x_s, x_r)$ such that

$$\text{Grad}J[v] = 0$$

implies

$$-\frac{1}{v^2} (x, z) J_1^*(x; z, (z; x)) J_2^*(z; x)(\tau^*[v] - \tau_d^*)(z, x) \chi^s(x, z) = 0$$

for all $s \in S$, for $x_s < x < x_c$, and all $z \in [0, Z]$, which in turn implies that

$$(\tau^*[v] - \tau_d^*)(z_r) = 0,$$

for all $s \in S$, and all $z_r \in [0, Z]$. Hence $v$ is a global minimizer.
Figure 2.1  Experiment Visualization – One Source

Figure 2.2  Ricker source wavelet
Figure 2.3 Seismogram.

Figure 2.4 Output least squares objective function.
Figure 2.5 Traveltime data.

Figure 2.6 Traveltime tomography objective function.
Figure 2.7 Experiment Visualization – Multiple Sources
Chapter 3

Traveltime Tomography – Multiple Arrivals

3.1 Introduction

Traditional traveltime tomography either assumes that data do not contain multiple arrivals (e.g., triplications) caused by the presence of caustics (envelopes of rays), as in Chapter 2, or gets rid of the multi-valued feature of the traveltime function by using only the first arrival. Traveltime tomography using only the first arrivals can be analyzed in a similar fashion as in Chapter 2, but it is not satisfactory since the first arrivals are usually very weak and do not carry information of the most interesting features of the medium (Geoltrain and Brac, 1992). A simple example is a medium which is homogeneous except a local low velocity zone. If the receiver line is far enough away from the low velocity zone, a triplication will be seen in the traveltime curve, and most rays which go through the low velocity zone give later arriving traveltimes. Hence traveltime inversion using the first arrivals will not be able to give a reasonable image of the low velocity zone (see Figures 3.1 - 3.3).

The difficulty in dealing with multiple arrivals lies in the fact that the traveltime data are parameterized by receiver location, hence the traveltime function is not
single-valued when multiple arrivals are in presence. In this chapter, we will present two ways to deal with this difficulty.

3.2 A New Formulation by Lailly et. al.

Lailly et. al. (1992) propose a new formulation of the traditional traveltime tomography which split the traveltime data into two groups: traveltime and emerging location, and reparameterizes these two set of data with the ray parameters. These two functions are always single valued since for a single ray, there is at most one traveltime and emerging point at the receiver surface. The least squares criterion is suggested to match traveltimes and emerging locations for an estimated velocity model with those calculated from the traveltime data with the same ray parameter. The key point in this new formulation is that for a receiver location and a traveltime, the ray parameter of the corresponding ray can be easily calculated of the data as long as there are a densely placed sources.

The set-up of the crosswell seismic experiment is still as in Chapter 2 except that the sources are densely located in a portion of the source well. We parameterize the sources by \( s \in S \). For each source location \( X_s = (x_s, z_s) \), traveltime data \( \tau_d(X_s, X_r) \), where \( X_r = (x_r, z_r) \), \( 0 \leq z_r \leq Z \), are collected. For a fixed \( z_s \), \( \tau_d(X_s, X_r) \) as a function of \( z_r \) can be multi-valued.

The first thing we want to do is to split the data into traveltimes and emerging locations of rays, and make them functions of ray parameters. This is possible because
from the eikonal equation

\[ |\nabla_{X_r}\tau(X_s, X_r)|^2 = \frac{1}{v(X_s)^2} \]

we have that, for a given source \( X - s \), a receiver \( X_r \), and a traveltime, the ray parameter associated to that particular ray is

\[ p(X_s, X_r) = \frac{\sin\theta}{v(X_s)} = \frac{\partial\tau(X_s, X_r)}{\partial z_s} \]

where \( \tau \) is the branch of the traveltime curve containing the given traveltime. We denote by \( P^s \) the set of ray parameters associated with source with index \( s \). Since for a single source, there is a one to one correspondence between ray parameters and rays, and a single ray will result in a single traveltime and emerging location at the receiver well, we can separately view the traveltime and emerging location as functions of ray parameters, i.e., for a source \( s \in S \), we represent the data in a new form

\[ \tau^*_d(p), z^*_d(p), p \in P^s \]

Now the new formulation of TT is to find a velocity model \( v \) to minimize

\[ J[v] = \sum_{s \in S} (|\tau^*[v] - \tau^*_d|^2 + |z^*[v] - z^*_d|^2) \]

where \( \tau^*[v](p) \) and \( z^*[v](p) \) are traveltime and emerging location of a ray from source \( s \) and with ray parameter \( p \), for a given velocity model \( v \), and \( \| \cdot \| \) is the standard norm in \( L^2(P^s) \).

This approach worked well for the example presented in Lailly et al (1992). We have not seen any analysis on why this approach should work, and we have not succeeded to analyze it ourselves either.
3.3 The Beam-Forming Approach

The "beam-forming" approach is based on the idea that different branches in the traveltime data correspond to different groups of set-off angles of rays, hence the rays can be grouped in a way that each group generates a single arrival. We are only be able to formulate this approach locally, and will use the triplication as a simple example to illustrate the idea. We assume that the data contains a simple triplication, and the velocity model which produces this data is \( v^* \). According to the smooth dependence result of solutions of ODEs on parameters, any velocity in a neighborhood of \( v^* \) will also produce a triplication in data. The three branches in the data can be separated, and for a given velocity model in the neighborhood of \( v^* \), all the rays can be grouped into three groups by raytracing in the following way: begin ray-tracing with the starting direction turning clockwise, a group is formed and the next group is started when the emerging location of the ray begin to turn back (see Fig. 3.4 – 3.6). Each group of rays formed in this way will produce a single valued branch. We denote the three branches in the data by \( \tau^i_d, i = 1, 2, 3 \), respectively, from up down, and the three branches for a given velocity model \( v \) by \( \tau^i[v], i = 1, 2, 3 \), in the same order. Since \( \tau^i_d \) and \( \tau^i[v] \) may have different supports, we shrink the supports, which are closed intervals, of \( \tau^i_d, i = 1, 2, 3 \), at the ends on caustics properly such that for any \( v \) in the above-mentioned neighborhood of \( v^* \), support of \( \tau^i[v] \) contains the shrunk support of \( \tau^i_d \). We denote these shrunk closed intervals by \( I^i, i = 1, 2, 3 \), then all the traveltime functions defined above are single-valued smooth functions of the
receiver location \( z \) in one of the intervals \( I_i \)'s. The objective function is

\[
J[v] = \sum_{i=1}^{3} \frac{1}{2} ||\tau_i[v] - \tau_d^i ||_{L^2(I_i)}^2
\]

The directional derivative in direction \( \delta v \) is

\[
DJ[v] \delta v = \langle \delta v, \sum_{i=1}^{3} (D\tau_i[v])^*(\tau_i[v] - \tau_d^i) \rangle
\]

where

\[
(D\tau_i[v])^*(\tau_i[v] - \tau_d^i)(x, z) =
\]

\[
\left(-\frac{1}{v^2}\right)(x, z)J_1^i(x; z_r(z;x))J_2^i(z; x)(\tau_i[v] - \tau_d^i)(z_r(x, z))\chi^i(x, z), \quad i = 1, 3
\]

where \( J_1^i \) and \( J_2^i \) are the two Jacobians in the proof of Lemma 2.2, respectively, for \( i \)th group of rays, and \( \chi^i \) is the characteristic function of the region covered by the \( i \)th group of rays. Now we consider \( (D\tau^2[v])^* \). We know that

\[
D\tau^2[v] \delta v(z_r) = \oint_{X_r} \left(-\frac{\delta v}{v^2}\right)(x(s), z(s))ds,
\]

and

\[
\langle D\tau^2[v] \delta v, \tau^2[v] - \tau_d^2 \rangle = \int_0^\pi \left\{ \oint_{X_r} \left(-\frac{\delta v}{v^2}\right)(x(s), z(s))ds \right\} (\tau^2[v] - \tau_d^2)(z_r)dz_r
\]

As in the proof of Lemma 2.2, we can do change of variable from \( s \) to \( x \),

\[
\langle D\tau^2[v] \delta v, \tau^2[v] - \tau_d^2 \rangle = \int_0^\pi \left\{ \int_{X_r} \left(-\frac{\delta v}{v^2}\right)(x, z(x; z_r))dx.J_1^2(x; z_r)dx \right\} (\tau^2[v] - \tau_d^2)(z_r)dz_r
\]

For all fixed \( x \) except the \( x = x^* \) corresponding to the cusped caustic (see Fig. 3.7),

we can do change of variable from \( z_r \) to \( z \). This change of variable is regular if \( x \) is to
the left of \( x^* \) and multi-valued, and singular on caustics if \( x \) is to the right of \( x^* \). The caustics except the cusp are folding singularities, and the multi-value nature of this change of variable is caused by this folding. This multi-valued change of variable can be written down more precisely, but since it is not a key point in our reasoning, we will ignore this multi-valuedness and write it as a single change of variable. We denote the jacobian of this change of variable by \( J^2_i(x; z) \), which is singular on caustics, then

\[
\langle D\tau^2[v] \delta v, \tau^2[v] - \tau^2_d \rangle = \int_0^2 \int_{x_r}^{x_r} \left( -\frac{\delta v}{v^2} \right) (x, z) J^2_1(x; z_r(z; x)) J^2_2(z; x) (\tau^2[v] - \tau^2_d)(z_r(x, z)) \chi^2(x, z) \, dx \, dz
\]

where \( \chi^2 \) is the characteristic function of the region covered by the second group of rays. Hence the \( L^2 \) gradient of \( J[v] \) is

\[
\text{Grad} J[v](x, z) = \sum_{i=1}^3 (D\tau^i[v])^*(\tau^i[v] - \tau^i_d)(x, z)
\]

\[
= \sum_{i=1}^3 \left( -\frac{1}{v^2} \right) (x, z) J^i_1(x; z_r(z; x)) J^i_2(z; x) (\tau^i[v] - \tau^i_d)(z_r(x, z)) \chi^i(x, z)
\]

When \( x_0 \) is close enough to \( x_0 \), the vertical line with \( x \)-coordinate \( x \) intersects the three regions in three disjoint line segments, hence

\[
\text{grad} J[v](x_0, z) = 0
\]

implies

\[
\left( -\frac{1}{v^2} \right) (x, z) J^i_1(x; z_r(z; x)) J^i_2(z; x) (\tau^i[v] - \tau^i_d)(z_r(x, z)) \chi^i(x_0, z) = 0, \quad i = 1, 2, 3
\]

which in turn implies

\[
(\tau^i[v] - \tau^i_d)(z_r) = 0 \quad \text{for} \quad z_r \in J^i, \quad i = 1, 2, 3
\]
Hence the synthetic traveltime data generated by velocity model \( v \) coincide with the data on the intervals \( I^i, i = 1, 2, 3 \). Hence this local beam-forming method has similar properties as the traveltime tomography with single arrivals, i.e., the objective function is smooth and locally every critical point is a velocity model which reproduces the traveltime data, though only mostly.
Figure 3.1 Intersecting rays caused by a low velocity zone. Caustics are the envelope on the rays.

Figure 3.2 Wavefronts for the same velocity model as in Figure 3.1
Figure 3.3  First arrivals do not always constrain a low velocity zone

Figure 3.4  Upper (first) group.
Figure 3.5 Center (second) group.

Figure 3.6 Lower (third) group.
Figure 3.7  The change of variable from \( z_r \) to \( x \) is multi-valued, and singular on the caustics if \( x \) is to the right of the cusp.
Chapter 4

Waveform Inversion via DSO – Single Arrival

4.1 Motivations and DSO Formulation

According to Chapter 2, if $f$ is highly oscillatory, as we shall assume, to good approximation

$$p(x, y, z, t) \approx a[v](x, y, z)f(t - \tau[v](x, y, z))$$

We shall use the right hand side of this approximate equality as the basic expression for the pressure.

We idealize the measurement of the pressure in the receiver well as the *trace* of the pressure field along the line segment $\{x = x_r, y = y_r, 0 \leq z \leq Z\}$. These measurements depend linearly on the source waveform $f$, and (quite) nonlinearly on the velocity field. Given these two quantities, the model explained so far predicts the measurement

$$S[v]f(z, t) = a[v](x_r, y_r, z)f(t - \tau[v](x_r, y_r, z))$$

$S$ will be called the *forward map*. Since $x_r$ and $y_r$ remain fixed for the remainder of the discussion, we suppress them from the notation, and write $a$ and $\tau$ as functions of $z$ alone. The *data* of the (idealized) crosswell inverse problem are presumed
measurements of the pressure field

\[ \{ S_{data}(z, t) : 0 \leq z \leq Z, 0 \leq t \leq T \} \]

We assume, with most other writers on this topic, that \( S_{data} \in L^2([0, Z] \times [0, T]) \).

This is certainly true if \( S_{data} \) is in the range of \( S[v] \) for some \( v \) and \( f \in L^2_{loc}(R) \).

A natural optimization formulation of the problem is: find \( v \) and \( f \) to minimize the mean square error

\[ \| S[v]f - S_{data} \|^2 \quad (4.1) \]

where the vertical double bars denote the norm in \( L^2([0, Z] \times [0, T]) \). As mentioned in the introduction, this objective function is very poorly behaved as a function of the velocity \( v \).

In order to motivate the DSO approach, let us do some observations first. We take \( S_{data}(z, t) = a[v^*](z)f^*(t - \tau[v^*](z)) \). For any velocity model \( v \), if we allow \( f \) to depend on the receiver location \( z \) also, then the pair

\[ (v, f[v] = \frac{a[v^*]}{a[v]} f^*(t - \tau[v^*] + \tau[v]) \quad (4.2) \]

is a global minimizer of the above least squares problem since it gives zero residue, and \( f[v] \) is independent of \( z \) if and only if \( a[v] = a[v^*] \) and \( \tau[v] = \tau[v^*] \), which implies that \( v \) is a kinematically correct velocity. Since there is only one source wavelet for all the receiver locations, i.e., \( f \) should be independent of \( z \), and we want the velocity model to be at least kinematically correct, we can use “how much” \( f[v] \) depends on
$z$ as a measure of "how correct" the current velocity $v$ is. The measure we use is

$$\| \frac{\partial}{\partial z} f[v] \|.$$

By forcing this term towards zero we expect to drive the almost arbitrary global minimizer (4.2) of (4.1) to a pair of kinematically velocity and a source wavelet independent of receiver location $z$. Besides, since

$$\frac{\partial}{\partial z} f[v] = \frac{\partial}{\partial z} \left( \frac{a[v^*]}{a[v]} \right) f^*(t - \tau[v^*] + \tau[v])$$

$$+ \frac{a[v^*]}{a[v]} \frac{d}{dt} f^*(t - \tau[v^*] + \tau[v]) \frac{\partial}{\partial z} (\tau[v] - \tau[v^*]),$$

when $f^*$ is highly oscillatory, the second term dominates, and it contains a factor

$$\frac{\partial}{\partial z} (\tau[v] - \tau[v^*])$$

which enable us to relate this measure to the traveltime tomography.

Inspired by the above observations, we will now introduce a variant of the mean square error which is much better behaved as a function of $v$, in several steps. We still take $V$ defined in the previous section as the set of velocities. We first make a technical modification to the definition of the forward map $S$. We regard $f$ as the primitive of the actual source waveform, rather than the source waveform itself, thus introducing a derivative into the expression for $S$. Thus from now on

$$S[v]f(z, t) = a[v](z) \frac{\partial f}{\partial t} (t - \tau[v](z))$$

The second change is essential, and in large part responsible for the success of our approach. Each "receiver" recording $\{ p(x_r, y_r, z, t) : 0 \leq t \leq T \}$ for each $z \in [0, Z]$
can be regarded as the result of an independent experiment. In fact, some crosswell surveys have been conducted with a single receiver, the source being setoff repeatedly while the receiver position is varied. For such surveys, the recording at each receiver position is literally the result of a separate experiment.

The essential modification is to alter the source waveform \( f \) to depend on the receiver location as well. That is, each experiment (receiver location) is explained by an \textit{a priori} independent model.

We make a technical assumption on the data:

- The noise free data \( S_{\text{data}} = a[v^*] \frac{\partial}{\partial t} f^*(t - \tau[v^*]) \) satisfies the condition: for any \( v \in V \) and any \( z \in [0, Z] \), the support of \( a[v^*] \frac{\partial}{\partial t} f^*(t - \tau[v^*] + \tau[v]) \) as a function of \( t \) is in \( (0, T) \).

This assumption can be realized by taking \( T \) large enough.

Since the mapping \( S[v] \) is composed of a partial differential operator and a shift in time, and the range is in \( L^2([0, Z] \times [0, T]) \), a reasonable choice for the domain of \( f \) will be

\[
\Omega[v] = \{(z, t) : 0 \leq z \leq Z, -\tau[v](z) \leq t \leq T - \tau[v](z)\}
\]

Physically realizable models, according to the preceding discussion, will have \( z - \text{independent} \) source waveforms. Thus our source models will also be subject to the Neumann boundary condition

\[
\frac{\partial f}{\partial z} = 0, \quad z = 0, Z
\]
although this will appear as a natural boundary condition, imposed implicitly through a variational principle.

Thus the domain of $f$ will be the closed subspace of $H^1(\Omega[v])$:

$$H[v] = \{ f \in H^1(\Omega[v]) : f(z, -\tau[v](z)) \equiv f(z, T - \tau[v](z)) \equiv 0, \forall z \in [0, Z]\}$$

and $S[v]$ is viewed as a mapping $H[v] \rightarrow L^2([0, Z] \times [0, T])$.

With these preliminaries out of the way, we state the objective function of our method:

$$J[v, f] = \frac{1}{2} \| S[v] f - S_{data} \|^2_{L^2([0, Z] \times [0, T])} + \frac{\sigma^2}{2} \| a[v] \frac{\partial}{\partial z} f \|^2_{L^2(\Omega[v])}. \quad (4.3)$$

The first term is the usual mean square error. The second measures the extent to which the sources chosen to represent the various receiver records resemble each other. This semblance is measured differentially, so this term gives the approach its name: Differential Semblance Optimization. Since physically realizable sources have no $z$-dependence, according to the assumptions of our theory, for noise-free data it is possible to make the two terms zero. The parameters $\sigma$ is called the differential semblance weight.

Remarks. In this paper we consider only a single source location $X_s = (x_s, z_s)$. However actual crosswell surveys involve many sources as well as many receivers. Obviously data from many sources more strictly constrains the velocity than does data from a single source. If many sources are used, filling the interval $0 \leq z_s \leq Z_s$,
for example, then an obvious modification of the DSO objective function is

$$J[v, f] = \frac{1}{2} \| S[v] f - S_{\text{data}} \|^2 + \frac{\sigma^2}{2} (\| \frac{\partial f}{\partial z_r} \|^2 + \| \frac{\partial f}{\partial z_s} \|^2)$$

where $\| \cdot \|$'s are norms of the respective function spaces. That is, the data is now compared for all source locations as well, and $f$ is regarded as depending on source location($z_s$) as well, this dependence being penalized in the same way as the dependence on $z_r$.

It is clear that every term in the norms in the objective function $J[v, f]$ (4.3) is linear in $f$, so minimizing $J[v, f]$ with $v$ fixed amounts to solving a linear normal equation. But the dependence of $J$ on $v$ is highly nonlinear. This makes it very difficult to minimize $J$ with respect to $v$ and $f$ simultaneously. In DSO approach, we minimize $J[v, f]$ in two steps:

*Inner State:* fix $v$, solve

$$\min_{f \in H[v]} J[v, f]$$

to get $f[v]$;

*Outer State:* let $J_{\text{des}}[v] = J[v, f[v]]$, and solve

$$\min_{v \in V} J_{\text{des}}[v].$$

4.2 Smoothness of the Objective Function

Define an operator

$$T_{v} : L^2([0, Z] \times [0, T]) \rightarrow L^2(\Omega[v])$$
\[ f(z,t) \mapsto f(z, t + \tau[v](z)) \]

This operator is an isometry, and its inverse and adjoint is

\[ T_{-\tau[v]} : L^2(\Omega[v]) \longrightarrow L^2([0, Z] \times [0, T]) \]

\[ f(z, t) \mapsto f(z, t - \tau[v](z)) \]

Denote

\[ H = \{ f \in H^1([0, Z] \times [0, T]) : f(\cdot, 0) \equiv f(\cdot, T) \equiv 0 \} \]

For fixed \( v \), let

\[ F(z, t) = T_{-\tau[v]}f(z, t) \]

and

\[ S_1[v] = a[v] \frac{\partial}{\partial t} \]

Then we have

\[ S[v]f = S_1[v]F \]

and the inner state is equivalent to the minimization problem

\[ \min_{F \in H} \left\{ \frac{1}{2} \| S_1[v]F - S_{data} \|^2 + \frac{\sigma^2}{2} \| a[v](\frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t})F \| ^2 \right\} \]

where \( \| \cdot \| \) is the standard norm of \( L^2([0, Z] \times [0, T]) \).

The minimization problem in the inner state is equivalent to solving the variational problem

\[ < S_1[v]F, S_1[v] \phi > + \sigma^2 < a[v](\frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t})F, a[v](\frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t})\phi > \]

\[ = < S_1[v] \phi, S_{data} >, \forall \phi \in H. \]
Theorem 4.1  For $S_{data} \in L^2([0,Z] \times [0,T])$, the normal equation has
unique solution $F[v]$ in $H$, and the solution satisfies
\[
\| \frac{\partial}{\partial t} F[v] \| \leq C \| S_{data} \|
\]
\[
\sigma \| (\frac{\partial}{\partial z} + \tau[v]' \frac{\partial}{\partial t}) F[v] \| \leq C \| S_{data} \|
\]
where $C$ is a constant independent of $v$ and $S_{data}$.

Proof  Since for each $z \in [0,Z]$, $a[v](z)$ is the final value of the solution of the
transport equation, which is a first order ordinary differential equation on the ray
with non-zero initial value, it is nonzero. Denote the minimum of $a[v]$ as $a_m$, then
\[
\int_0^Z \int_0^T a[v]^2 (\frac{\partial}{\partial t} F)^2 + \sigma^2 a[v]^2 \{ (\frac{\partial}{\partial z} + \tau[v]' \frac{\partial}{\partial t}) F \}^2
\]
\[
\geq a_m^2 \min (1, \sigma^2) \int_0^Z \int_0^T \frac{\partial}{\partial t} F)^2 + \{ (\frac{\partial}{\partial z} + \tau[v]' \frac{\partial}{\partial t}) F \}^2
\]
\[
\geq a_m^2 \min (1, \sigma^2) \int_0^Z \int_{-\tau[v](z)}^{T-\tau[v](z)} (\frac{\partial}{\partial t} f)^2 + (\frac{\partial}{\partial z} f)^2
\]
Since
\[f(z,t) = 0 \quad \text{for} \quad t = -\tau[v](z) \quad \text{and} \quad t = T - \tau[v](z), 0 \leq z \leq Z\]
using the fundamental theorem of calculus we can get
\[
\int_0^Z \int_{-\tau[v](z)}^{T-\tau[v](z)} f^2 \leq C \int_0^Z \int_{-\tau[v](z)}^{T-\tau[v](z)} \left( \frac{\partial}{\partial t} f \right)^2
\]
where $C > 0$. From the above two inequalities we can get
\[
\int_0^Z \int_0^T a[v]^2 (\frac{\partial}{\partial t} F)^2 + \sigma^2 a[v]^2 \{ (\frac{\partial}{\partial z} + \tau[v]' \frac{\partial}{\partial t}) F \}^2
\]
\[
\geq C \int_0^Z \int_{-\tau[v](z)}^{T-\tau[v](z)} \left( \frac{\partial}{\partial t} f \right)^2 + \left( \frac{\partial}{\partial z} f \right)^2 + f^2 \\
\geq C \int_0^Z \int_0^T \left( \frac{\partial}{\partial t} F \right)^2 + \left( \frac{\partial}{\partial z} F \right)^2 + F^2 
\]

Hence the bilinear form in the variational problem is coercive. The bilinear form and the linear functional on the right hand side of the variational problem is obviously continuous, hence the problem has unique solution \( F[v] \) in \( H \) by Lax-Milgram theorem (see Lions 1967). Substitute \( \phi = F[v] \) in the variational problem, we immediately get the two estimates for the first partial derivatives of the solution. q.e.d.

**Theorem 4.2** \( F[v] \in H \) depends smoothly on \( v \in V \).

The proof of this theorem combines a proof of that \( F[v] \) depends smoothly on \( a[v] \) and \( \tau[v] \) as \( L^2 \) functions and a proof that \( a[v] \) and \( \tau[v] \) as \( L^2 \) functions depend smoothly on \( v \in V \).

**Proof** From Chapter 2 we know that \( \tau[v] \in L^2([0, Z]) \) smoothly depends on \( v \in V \). Similarly it can be proved that \( a[v] \in L^2([0, Z]) \) smoothly depends on \( v \in V \) too. Now we only need to prove that \( F[v] \) depends on \( a[v] \) and \( \tau[v] \) smoothly. The proof of this claim is routine but tedious. In order to make the idea clear, we will study the dependence of \( F[v] \) on \( a[v] \) only. The discussion of the general case is similar.

For the sake of simplicity, we will suppress all the dependences on \( v \). Let \( \delta a \) be a perturbation of \( a \) and \( \lambda \) be a nonzero real number. Let \( F(\lambda) = F[a + \lambda \delta a] \). Then \( F(\lambda) - F \) satisfies for any \( \phi \in H \)

\[
\langle a \frac{\partial}{\partial t} (F(\lambda) - F), a \frac{\partial}{\partial t} \phi \rangle + \sigma^2 \langle a (\frac{\partial}{\partial z} + \tau' \frac{\partial}{\partial t}) (F(\lambda) - F), a (\frac{\partial}{\partial z} + \tau' \frac{\partial}{\partial t}) \phi \rangle
\]
\[
\begin{align*}
  & = \langle a \frac{\partial}{\partial t} \phi, \lambda \frac{\delta a}{a} S_d - 2\lambda \delta a \frac{\partial}{\partial t} F(\lambda) - \lambda^2 \frac{(\delta a)^2}{a} \frac{\partial}{\partial t} F(\lambda) \rangle \\
  & + \sigma^2 \langle a \left( \frac{\partial}{\partial z} + \tau \frac{\partial}{\partial t} \right) \phi, -2\lambda \delta a \left( \frac{\partial}{\partial z} + \tau \frac{\partial}{\partial t} \right) F(\lambda) - \lambda^2 \frac{(\delta a)^2}{a} \left( \frac{\partial}{\partial z} + \tau \frac{\partial}{\partial t} \right) F(\lambda) \rangle
\end{align*}
\]

Substitute \( \phi = F(\lambda) - F \) into the above equality, and apply Cauchy-Schwartz inequality to the right hand side, we get

\[
\begin{align*}
  & \|a \frac{\partial}{\partial t} (F(\lambda) - F)\|^2 + \sigma^2 \|a \left( \frac{\partial}{\partial z} + \tau \frac{\partial}{\partial t} \right) (F(\lambda) - F)\|^2 \\
  & \leq \|a \frac{\partial}{\partial t} (F(\lambda) - F)\| \|\lambda \frac{\delta a}{a} S_d - 2\lambda \delta a \frac{\partial}{\partial t} F(\lambda) - \lambda^2 \frac{(\delta a)^2}{a} \frac{\partial}{\partial t} F(\lambda)\| \\
  & + \sigma^2 \|a \left( \frac{\partial}{\partial z} + \tau \frac{\partial}{\partial t} \right) (F(\lambda) - F)\| : \\
  & \|2\lambda \delta a \left( \frac{\partial}{\partial z} + \tau \frac{\partial}{\partial t} \right) F(\lambda) + \lambda^2 \frac{(\delta a)^2}{a} \left( \frac{\partial}{\partial z} + \tau \frac{\partial}{\partial t} \right) F(\lambda)\|
\end{align*}
\]

Now we have got an inequality in the form

\[
A^2 + B^2 \leq AC + BD
\]

Use

\[
AC \leq \frac{A^2 + C^2}{2}, \quad BD \leq \frac{B^2 + D^2}{2}
\]

to the right hand side, we get

\[
A^2 + B^2 \leq C^2 + D^2
\]

Hence we get

\[
\begin{align*}
  & \|a \frac{\partial}{\partial t} (F(\lambda) - F)\|^2 + \sigma^2 \|a \left( \frac{\partial}{\partial z} + \tau \frac{\partial}{\partial t} \right) (F(\lambda) - F)\|^2 \\
  & \leq \lambda^2 \|\frac{\delta a}{a} S_d - 2\delta a \frac{\partial}{\partial t} F(\lambda) - \lambda \frac{(\delta a)^2}{a} \frac{\partial}{\partial t} F(\lambda)\|^2 \\
  & + \lambda^2 \|2\delta a \left( \frac{\partial}{\partial z} + \tau \frac{\partial}{\partial t} \right) F(\lambda) + \lambda \frac{(\delta a)^2}{a} \left( \frac{\partial}{\partial z} + \tau \frac{\partial}{\partial t} \right) F(\lambda)\|^2
\end{align*}
\]
Let $\lambda$ go to zero, we get
\[
\|a \frac{\partial}{\partial t}(F(\lambda) - F)\| \longrightarrow 0 \quad (4.4)
\]
\[
\|a(\frac{\partial}{\partial z} + \tau' \frac{\partial}{\partial t})(F(\lambda) - F)\| \longrightarrow 0 \quad (4.5)
\]

Let $DF\delta a$ be the unique solution of the following variational problem
\[
\langle a \frac{\partial}{\partial t}DF\delta a, a \frac{\partial}{\partial t} \phi \rangle + \sigma^2 \langle a(\frac{\partial}{\partial z} + \tau' \frac{\partial}{\partial t}) DF\delta a, a(\frac{\partial}{\partial z} + \tau' \frac{\partial}{\partial t}) \phi \rangle
\]
\[
= \langle \delta a \frac{\partial}{\partial t} \phi, \frac{\delta a}{a} \delta a - 2 \delta a \frac{\partial}{\partial t} F \rangle - \sigma^2 \langle a(\frac{\partial}{\partial z} + \tau' \frac{\partial}{\partial t}) \phi, 2 \delta a(\frac{\partial}{\partial z} + \tau' \frac{\partial}{\partial t}) \phi \rangle
\]

Then $\frac{F(\lambda) - F}{\lambda} - DF\delta a$ satisfies
\[
\langle a \frac{\partial}{\partial t} \left( \frac{F(\lambda) - F}{\lambda} - DF\delta a \right), a \frac{\partial}{\partial t} \phi \rangle
\]
\[
+ \sigma^2 \langle a(\frac{\partial}{\partial z} + \tau' \frac{\partial}{\partial t}) \left( \frac{F(\lambda) - F}{\lambda} - DF\delta a \right), a(\frac{\partial}{\partial z} + \tau' \frac{\partial}{\partial t}) \phi \rangle
\]
\[
= \langle a \frac{\partial}{\partial t} \phi, 2 \delta a \frac{\partial}{\partial t} (F - F(\lambda)) \rangle - \lambda \frac{(\delta a)^2}{a} \frac{\partial}{\partial t} F(\lambda)
\]
\[
+ \sigma^2 \langle a(\frac{\partial}{\partial z} + \tau' \frac{\partial}{\partial t}) \phi, 2 \delta a(\frac{\partial}{\partial z} + \tau' \frac{\partial}{\partial t}) (F - F(\lambda)) \rangle - \lambda \frac{(\delta a)^2}{a} \frac{\partial}{\partial t} F(\lambda)
\]

Similar to the proof of (4.4)-(4.5), we can prove that
\[
\|a \frac{\partial}{\partial t} \left( \frac{F(\lambda) - F}{\lambda} - DF\delta a \right)\|^2 + \sigma^2 \|a(\frac{\partial}{\partial z} + \tau' \frac{\partial}{\partial t}) \left( \frac{F(\lambda) - F}{\lambda} \right)\|^2 \longrightarrow 0
\]
as $\lambda$ goes to zero. From the coerciveness of the variational problem, we know that this means $F$ is differentiable and its directional derivative in direction $\delta a$ is $DF\delta a$.

In the same way we can prove that this derivative is continuous and $F$ has continuous higher derivatives.

q.e.d.

The smoothness of $J[v]$ is a consequence of the previous theorem.

**Theorem 4.3** $J[v]$ depends smoothly on $v \in V$. 
4.3 The Gradient and Characterization of Critical Points

In this section, we calculate the gradient of the objective function

\[ J_{dso}[v] = \frac{1}{2} \| S_1[v] F[v] - S_{data} \|^2 + \frac{\sigma^2}{2} \| a[v] (\frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t}) F[v] \|^2 \]  \hspace{1cm} (4.6)\]

where \( F[v] \) is the solution of the inner state optimization problem with given \( v \). From now on we suppress the dependence of \( S \) on \( v \) for the simplicity of the notation.

**Theorem 4.4** The directional derivative of \( J[v] \) is given by

\[
DJ_{dso}[v] \delta v = \sigma^2 (a[v] (\frac{d}{dz} D\tau[v] \delta v) \frac{\partial}{\partial t} F[v], a[v] (\frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t}) F[v]) \]

\[ -\sigma^2 (a[v] (\frac{d}{dz} D a[v] \delta v)^2) F[v], a[v] (\frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t}) F[v]) \]

**Proof** Take directional derivative of (4.6) in direction \( \delta v \), we get

\[
DJ_{dso}[v] \delta v = (Da[v] \delta v \frac{\partial}{\partial t} F[v] + a[v] \frac{\partial}{\partial t} (DF[v] \delta v), a[v] \frac{\partial}{\partial t} F[v] - S_{data})
\]

\[ +\sigma^2 (Da[v] \delta v \frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t}) F[v] + a[v] (D\tau[v] \delta v) \frac{\partial}{\partial t} F[v] \]

\[ +a[v] (\frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t}) (DF[v] \delta v), a[v] (\frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t}) F[v]) \]

Since \( F[v] \) satisfies the variational problem, we have

\[
\langle a[v] \frac{\partial}{\partial t} (\frac{Da[v] \delta v}{a[v]} F[v]), a[v] \frac{\partial}{\partial t} F[v] - S_{data} \rangle
\]

\[ +\sigma^2 (a[v] (\frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t}) (\frac{Da[v] \delta v}{a[v]} F[v]), a[v] (\frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t}) F[v]) = 0 \]

and

\[
\langle a[v] \frac{\partial}{\partial t} (DF[v] \delta v), a[v] \frac{\partial}{\partial t} F[v] - S_{data} \rangle
\]

\[ +\sigma^2 (a[v] (\frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t}) (DF[v] \delta v), a[v] (\frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t}) F[v]) = 0 \]
hence we get

\[
DJ_{dso}[v] \delta v = \sigma^2(a[v])(\frac{d}{dz} D\tau[v] \delta v) \frac{\partial}{\partial t} F[v], a[v](\frac{\partial}{\partial z} + (\frac{d}{dz} \tau[v]) \frac{\partial}{\partial t}) F[v]
\]

\[-\sigma^2(a[v])(\frac{d}{dz} a[v] \delta v) F[v], a[v](\frac{\partial}{\partial z} + \tau[v]\frac{\partial}{\partial t}) F[v]\]

q.e.d

**Corollary 4.1** If the data is noise free: \( S_{data} = a[v^*] \frac{df^*}{dt} (t - \tau[v^*]) \), then

\( v = v^* \) is a global minimizer of the DSO objective function \( J_{dso} \).

**Proof** It is easy to see that at \( v = v^* \), the solution of the normal equation is \( f[v](z, t) = f^*(t) \), hence from the previous theorem, \( DJ[v^*] \delta v = 0 \) for any \( \delta v \). This point is a global minimizer since it gives zero residue to the objective function. q.e.d.

In order to analyze the gradient of \( J[v] \), we need the following lemmas.

**Lemma 4.1**

\[
\max_z |\frac{d}{dz} D\tau[v] \delta v| \leq C \|\delta v\|_2,
\]

\[
\max_z |\frac{d}{dz} Da[v] \delta v| \leq C \|\delta v\|_3
\]

\[
\max_z |Da[v] \delta v| \leq C \|\delta v\|_3
\]

where \( C \) is independent of \( v \in V \) and \( \delta v \).

**Proof** \( D\tau[v] \delta v \) is given by

\[
D\tau[v] \delta v(z) = \int_{X_t}^{(x_t, z)} \left( -\frac{\delta v}{v^2} \right) (x(s), z(s)) ds
\]
where \( s \) is the raypath length from the source. For any \( z \in [0, Z] \), there is a ray connecting the source \( X_s \) and the point \((x_r, z)\). For any point on the ray with \( x \)-coordinate \( x \in [x_s, x_r] \), we denote its \( z \)-coordinate by \( y(x; z) \). Now the formula for \( D\tau[v]\delta v \) becomes

\[
D\tau[v]\delta v(z) = \int_{x_s}^{x_r} (\frac{-\delta v}{v_T})(x, y(x; z)) J(x; z) \, dx
\]

where \( J(x; z) \) is the jacobian of the change of integration variable from arclength \( s \) to \( x \). Hence the first inequality is immediately from

\[
\frac{\partial}{\partial z} D\tau[v]\delta v(z) = \int_{x_s}^{x_r} \left( -\frac{\partial}{\partial y} \left( \frac{\delta v}{v^2} \right)(x, y(x; z)) \right) \frac{\partial y(x; z)}{\partial z} J(x; z)
- \frac{\delta v}{v^2}(x, y(x; z)) \frac{\partial J(x; z)}{\partial z} \, dx
\]

and the trace theorem, and the fact that \( V \) is a bounded set.

Since \( a[v] \) satisfies the transport equation

\[
\nabla a[v] \cdot \nabla \tau[v] + \frac{1}{2} a[v]\Delta \tau = 0
\]

with proper boundary condition, hence \( Da[v]\delta v \) satisfies

\[
\nabla (Da[v]\delta v) \cdot (v \nabla \tau[v]) + \frac{1}{2} (Da[v]\delta v)v\Delta \tau = -v \nabla a[v] \cdot \nabla (D\tau[v]\delta v) - v\frac{1}{2} a[v]\Delta (D\tau\delta v)
\]

with proper boundary condition. As in the proof of the last lemma, we can solve a initial value problem of an ordinary differential equation along rays and get an integral expression for \( Da[v]\delta v \):

\[
Da[v]\delta v = e^{-\int \frac{1}{2} u\Delta \tau[v] \, ds} \left\{ \int e^{\int \frac{1}{2} u\Delta \tau[v]} - v \nabla a[v] \cdot \nabla (D\tau[v]\delta v) - v\frac{1}{2} a[v]\Delta (D\tau\delta v) \right\}
\]
A conservative estimate gives

\[
\max_z \left| \frac{d}{dz} Da[v] \delta v \right| \leq C \| \delta v \|_3
\]

\[
\max_z | Da[v] \delta v | \leq C \| \delta v \|_3
\]

where \( C \) is independent of \( v \) and \( \delta v \). q.e.d.

Let

\[
S^*_{\text{data}} = a[v^*] \frac{d}{dt} f^*(t - \tau[v^*])
\]

\[
S_{\text{data}} = S^*_{\text{data}} + e
\]

and

\[
r = \frac{\| e \|}{\| S^*_{\text{data}} \|}.
\]

**Lemma 4.2** The solution of the inner state can be expressed as follows

\[
F[v] = \frac{a[v^*]}{a[v]} f^*(t - \tau[v^*]) + R(z, t) \quad (4.9)
\]

where \( R \in H^1([0, T] \times [0, X]) \) satisfies

\[
\left\| \frac{\partial}{\partial t} R \right\| \leq C(\sigma^2 + r)\| S^*_{\text{data}} \|
\]

\[
\left\| \frac{\partial}{\partial z} + \tau[v^*] \frac{\partial}{\partial t} \right\| R \right\| \leq C(\sigma + r) \| S^*_{\text{data}} \|
\]

\[
\| R \| \leq C(\sigma + r) \| S^*_{\text{data}} \|
\]

where \( C \) does not depend on \( v \).
Proof Let

\[ R(z, t) = F[v] - \frac{a[v^*]}{a[v]} f^*(t - \tau[v^*]). \]

From the assumption on data we have \( R \in H \) and

\[
\langle a \frac{\partial}{\partial t} R, a \frac{\partial}{\partial t} \phi \rangle + \sigma^2 \langle a \left( \frac{\partial}{\partial z} + \tau' \frac{\partial}{\partial t} \right) R, a \left( \frac{\partial}{\partial z} + \tau' \frac{\partial}{\partial t} \right) \phi \rangle \\
= -\sigma^2 \langle a \left( \frac{d}{dz} \left( \frac{a^*}{a} \right) \right) f^*(t - \tau^*), a \left( \tau' - \tau^* \right) \frac{d}{dt} f^*(t - \tau^*) \rangle, a \left( \frac{\partial}{\partial z} + \tau' \frac{\partial}{\partial t} \right) \phi \rangle \\
+ \langle e, a \frac{\partial}{\partial t} \phi \rangle
\]

Let \( \phi = R \), then

\[
\|a \frac{\partial}{\partial t} R\|^2 + \sigma^2 \|a \left( \frac{\partial}{\partial z} + \tau' \frac{\partial}{\partial t} \right) R\|^2 \\
\leq \sigma^2 \|a \left( \frac{d}{dz} \left( \frac{a^*}{a} \right) \right) f^*(t - \tau^*)\| + \|a \left( \tau' - \tau^* \right) \frac{d}{dt} f^*(t - \tau^*)\| \|a \left( \frac{\partial}{\partial z} + \tau' \frac{\partial}{\partial t} \right) R\| \\
+ \|e\| \|a \frac{\partial}{\partial t} R\|
\]

Again we get an inequality in the form

\[ A^2 + B^2 \leq AC + BD \]

As in the proof of Theorem 3.2, we get

\[ A^2 + B^2 \leq C^2 + D^2 \]

Hence

\[
\|a \frac{\partial}{\partial t} R\| \leq \sigma^2 \|a \left( \frac{d}{dz} \left( \frac{a^*}{a} \right) \right) f^*(t - \tau^*)\| + \|a \left( \tau' - \tau^* \right) \frac{d}{dt} f^*(t - \tau^*)\| + \|e\| \\
\|a \left( \frac{\partial}{\partial z} + \tau' \frac{\partial}{\partial t} \right) R\| \leq \sigma \|a \left( \frac{d}{dz} \left( \frac{a^*}{a} \right) \right) f^*(t - \tau^*)\| + \|a \left( \tau' - \tau^* \right) \frac{d}{dt} f^*(t - \tau^*)\| + \|e\|
\]
The first two desired inequalities are immediate consequences of the above inequality.

Since $R \in H$, similar to the coercivity proof in the proof of Theorem 4.1, we have

$$
\|R\| \leq \|R\|_H \leq C \left\{ \left\| \frac{\partial}{\partial t} R \right\| + \left\| (\frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t} ) R \right\| \right\} \leq C(\sigma + r).
$$

q.e.d.

Using (4.9), we can get an $L^2$ estimate of $F[v]$ and improvement of the inequalities in Theorem 4.1.

\textbf{Corollary 4.2}

$$
\|F[v]\| \leq C(\beta + \sigma + r)\|S_{\text{data}}^*\|
$$

$$
\left\| a[v] \left( \frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t} \right) F[v] \right\| \leq (C_0 + C_1 \beta + C_2 \sigma + C_3 r)\|S_{\text{data}}^*\|
$$

where $\beta = \frac{\|f^*\|}{\|f\|}$, and $C, C_0, C_1, C_2, C_3$ are constants independent of $v$.

\textbf{Proof}  The desired inequalities follow from Lemma 4.2.  q.e.d.

Now we analyze the gradient. Substitute

$$
F[v] = \frac{a[v^*]}{a[v]} f^*(t - \tau[v^*]) + R
$$

in the formula for $DJ_{\text{dso}}[v]\delta v$ given in Theorem 4.4 and use the estimates in Theorem 4.1, Lemmas 4.1 and 4.2, and Corollary 4.2, the directional derivative of $J[v]$ can be written as

$$
1DJ_{\text{dso}}[v]\delta v \leq \sigma^2 \left\{ \int_0^T \int_0^Z a[v^*]^2 \left( \frac{d}{dz} \tau[v] \delta v \right) \frac{d}{dz} (\tau[v] - \tau[v^*]) \left[ \frac{d}{dt} f^*(t - \tau[v^*]) \right]^2 dzdt + \sigma(\sigma + r)C_1(\|\delta v\|_2 + \|\delta v\|_3) \left\| \frac{d}{dt} f^* \right\|^2 + C_2(\|\delta v\| + \|\delta v\|_3) \left\| f^* \right\| \left\| \frac{d}{dt} f^* \right\| \right\}
$$
\[
\sigma^2 \left\{ \int_0^2 a[v^*]^2 \left( \frac{d}{dz} D\tau[v]\delta v \right) \frac{d}{dz} (\tau[v] - \tau[v^*]) \right\} dz \\
+ (\sigma + \tau)C_1(\|\delta v\|_2 + \|\delta v\|_3) + \beta C_2(\|\delta v\|_2 + \|\delta v\|_3) \right\} \left\| \frac{d}{dt} f^* \right\|^2
\]

where \(C_i, i = 1, 2\), are independent of \(\delta v, v, \) and \(\sigma\), and
\[
\beta \equiv \frac{\|f^*\|}{\|\frac{d}{dt} f^*\|}
\]
measures the oscillation of \(f^*\). Obviously the velocity perturbations outside of the region covered by the rays are not constrained by the data, we only consider those inside the covered region. We make a little bit stronger assumption that the velocity perturbations are supported properly inside the covered region, i.e., their supports do not touch the boundary rays for any \(v \in V\). Under this assumption,
\[
D\tau[v]\delta v(0) = D\tau[v]\delta v(Z) = 0
\]
according to the formula of \(D\tau[v]\delta v\) in the proof of Lemma 4.1. Do integration by parts to the integral on the right hand side of the last equality and then rewrite it using the adjoint of \(D\tau[v]\), we get
\[
DJ_{dso}[v]\delta v = \sigma^2 \left\| \frac{d}{dt} f^* \right\|^2 \left( - \int \delta v (D\tau[v])^* \frac{d}{dz} \left\{ a[v^*]^2 \frac{d}{dz} (\tau[v] - \tau[v^*]) \right\} \\
+ C_1(\sigma + \tau)(\|\delta v\|_2 + \|\delta v\|_3) + C_2(\|\delta v\|_2 + \|\delta v\|_3) \right\}
\]

Let \(Grad J_{dso}[v]\) be the \(L^2\) gradient of \(J_{dso}[v]\), i.e.,
\[
DJ_{dso}[v]\delta v = \delta v, Grad J_{dso}[v]
\]
then we have
\[
\|\text{Grad} J_{\text{dso}}[v]\| = \sigma^2 \left\| \frac{d}{dt} f^* \right\|^2 \left\{ \left\| \left( D\tau[v] \right)^* \{ \frac{d}{dz} a[v^*]^2 \frac{d}{dz} (\tau[v] - \tau[v^*]) \} \right\| + O(\sigma) + O(r) + O(\beta) \right\}
\geq \sigma^2 C \left\| \frac{d}{dz} \left\{ a[v^*]^2 \frac{d}{dz} (\tau[v] - \tau[v^*]) \right\} \right\| + O(\sigma) + O(r) + O(\beta)
\]

here we have used the coercivity of \((D\tau[v])^*\) (Lemma 2.2) and the finite-dimensionality of \(DV\), which is a consequence of the finite dimensionality of \(V\). We will also use the finite dimensionality of \(DV\) and not specify the norm of \(\delta v\) in the rest of this chapter.

If the gradient of \(J_{\text{dso}}[v]\) is zero at \(v\), then
\[
\left\| \frac{d}{dz} \left\{ a[v^*]^2 \frac{d}{dz} (\tau[v] - \tau[v^*]) \right\} \right\| \leq C(\sigma + \beta + r)
\]
where \(C\) is a constant independent of \(v\) and \(\delta v\). We summarize the above discussion as the following theorem.

**Theorem 4.5** A critical point \(v\) of the DSO objective function satisfies
\[
\left\| \frac{d}{dz} \left\{ a[v^*]^2 \frac{d}{dz} (\tau[v] - \tau[v^*]) \right\} \right\| \leq C(\sigma + \beta + r)
\]
where \(\beta = \| f^* \| / \| \frac{d}{dz} f^* \|\), and \(C\) is a positive constant independent of \(v\).

Note that if we assume that \(v\) and \(v^*\) give same travel times near the ends of the receiver array, i.e.,
\[
\tau[v](z) = \tau[v^*](z), \text{ near } z = 0, Z
\]
(this assumption is satisfied if \( v \) and \( v^* \) differ only in the region covered by rays connecting the source and the receivers for both velocities), then Poincare’s inequality leads to

\[
\| \tau[v] - \tau[v^*] \| \leq C(\sigma + \beta + r)
\]

This result says that if the differential semblance \( \sigma \) is small enough, and the source wavelet \( f^* \) is oscillatory enough, then any stationary point of \( J[v] \) in \( V \) is kinematically close to the true velocity model \( v^* \).

**Corollary 4.3** If \( v \) is a critical point and \( v \) and \( v^* \) differ only in the region covered by rays connecting the source and the receivers for both velocities, then

\[
\| \tau[v] - \tau[v^*] \| \leq C(\sigma + \beta + r)
\]

For the estimation of the source wavelet, define

\[
\bar{f}(t) = \frac{1}{Z} \int_0^Z dz \frac{a[v](z)}{a[v^*](z)} f[v](z,t)dz
\]

where \( v \) is a minimizer, then using Lemma 4.2, we can get immediately

\[
\| \bar{f}[v] - f^* \|_{L^2([0,T])} \leq \| \frac{1}{Z} \int_0^Z dz \frac{a[v](z)}{a[v^*](z)} \frac{\partial}{\partial t} \left[ f[v] - \frac{a^*[v](z)}{a[v](z)} f^* \right] \|
\leq C(\| \tau[v] - \tau[v^*] \| + \sigma + r)
\]

Using Corollary 4.3, we get,
Corollary 4.4 Under the same assumptions as in Corollary 4.3, the source wavelet estimation $\tilde{f}$ satisfies

$$\|\tilde{f}[v] - f^*\|_{L^1([0,T])} \leq C(\sigma + \beta + r)$$

4.4 Hessian and Local Convexity

Differentiate the directional derivative of $J_{deo}$, we get

$$D^2 J_{deo}[v](\delta v, \delta v) = \sigma^2 < D\{a[v]^2 \frac{d}{dz} D\tau[v]\delta v\} \delta v \partial F[v], (\frac{\partial}{\partial z} + \tau[v]' \frac{\partial}{\partial t}) F[v] >$$

$$+ \sigma^2 < a[v]^2 (\frac{d}{dz} D\tau[v]\delta v) \partial F[v], (\frac{\partial}{\partial z} + \tau[v]' \frac{\partial}{\partial t}) F[v] >$$

$$+ \sigma^2 < a[v]^2 (\frac{d}{dz} D\tau[v]\delta v) F[v]\delta v, (\frac{\partial}{\partial z} + \tau[v]' \frac{\partial}{\partial t}) F[v] >$$

$$- \sigma^2 < a[v]^2 (\frac{d}{dz} D\tau[v]\delta v) F[v]\delta v, (\frac{\partial}{\partial z} + \tau[v]' \frac{\partial}{\partial t}) F[v] >$$

In order to estimate the hessian, we need to estimate the $L^2$ norm of $DF[v]\delta v$ and its derivatives first. We will do so by estimating $D\tau[v]\delta v$ and using (4.9).

Substitute

$$R[v](z,t) = F[v] - \frac{a[v]}{a[v]} f^*(t - \tau[v^*])$$
to the variational equality satisfied by \( F[v] \), we get
\[
\langle a[v] \frac{\partial}{\partial t} R[v], a[v] \frac{\partial}{\partial t} \phi \rangle + \sigma^2 \langle a[v] \left( \frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t} \right) R[v], a[v] \left( \frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t} \right) \phi \rangle
\]
\[
= \langle e, a[v] \frac{\partial}{\partial t} \phi \rangle - \sigma^2 \langle a[v] \left( \frac{d}{dz} + \tau[v] \frac{\partial}{\partial t} \frac{a^*}{a} f^*(t - \tau^*), a[v] \left( \frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t} \right) \phi \rangle
\]

Taking derivative with respect to \( v \) in direction \( \delta v \), we get a variational equality satisfied by \( DR[v] \delta v \):
\[
\langle a[v] \frac{\partial}{\partial t} DR[v] \delta v, a[v] \frac{\partial}{\partial t} \phi \rangle + \sigma^2 \langle a[v] \left( \frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t} \right) DR[v] \delta v, a[v] \left( \frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t} \right) \phi \rangle
\]
\[
= \langle e, Da[v] \delta v \frac{\partial}{\partial t} \phi \rangle
\]
\[
- \sigma^2 \langle 2Da[v] \delta v \left( \frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t} \right) \left[ \frac{a[v^*]}{a[v]} f^*(t - \tau[v^*]) \right], a[v] \left( \frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t} \right) \phi \rangle
\]
\[
- \sigma^2 \langle Da[v] \left( \frac{d}{dz} DR[v] \delta v \right) \frac{\partial}{\partial t} \left[ \frac{a[v^*]}{a[v]} f^*(t - \tau[v^*]) \right], a[v] \left( \frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t} \right) \phi \rangle
\]
\[
- \sigma^2 \langle a[v] \left( \frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t} \right) \left[ D \left[ \frac{a[v^*]}{a[v]} f^*(t - \tau[v^*]) \right], a[v] \left( \frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t} \right) \phi \rangle
\]
\[
- \sigma^2 \langle a[v] \left( \frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t} \right) \left[ \frac{a[v^*]}{a[v]} f^*(t - \tau[v^*]) \right], a[v] \left( \frac{d}{dz} DR[v] \delta v \right) \frac{\partial}{\partial t} \phi \rangle
\]
\[
- \langle 2Da[v] \delta v \frac{\partial}{\partial t} R[v], a[v] \frac{\partial}{\partial t} \phi \rangle
\]
\[
- \sigma^2 \langle 2Da[v] \delta v \left( \frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t} \right) R[v], a[v] \left( \frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t} \right) \phi \rangle
\]
\[
- \sigma^2 \langle a[v] \left( \frac{d}{dz} DR[v] \delta v \right) \frac{\partial}{\partial t} \left[ \frac{a[v^*]}{a[v]} f^*(t - \tau[v^*]) \right], a[v] \left( \frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t} \right) \phi \rangle
\]
\[
- \sigma^2 \langle a[v] \left( \frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t} \right) \left[ D \left[ \frac{a[v^*]}{a[v]} f^*(t - \tau[v^*]) \right], a[v] \left( \frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t} \right) \phi \rangle
\]
\[
- \sigma^2 \langle a[v] \left( \frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t} \right) \left[ \frac{a[v^*]}{a[v]} f^*(t - \tau[v^*]) \right], a[v] \left( \frac{d}{dz} DR[v] \delta v \right) \frac{\partial}{\partial t} \phi \rangle
\]

Let \( \phi = DR[v] \delta v \), use the same technique we have used in the proof of Lemma 4.2, and assume that \( \sigma < 1 \), we can get
\[
\| \frac{\partial}{\partial t} DR[v] \delta v \| \leq C(\sigma + r) \| S_{data}^* \| \| \delta v \|
\]
\[
\| (\frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t}) DR[v] \delta v \| \leq C(\sigma + r) \| S_{data}^* \| \| \delta v \|
\]
which in turn implies

\[ \| DR[v] \delta v \| \leq C(\sigma + r)\| S_{data}^* \| \| \delta v \| \]

Now from

\[ DF[v] \delta v = D \left( \frac{a[v^*]}{a[v]} \right) \delta v f^*(t - \tau[v^*]) + DR[v] \delta v \]

we get

\[ \left\| \frac{\partial}{\partial t} DF[v] \delta v \right\| \leq C \| S_{data}^* \| \| \delta v \| \]

\[ \left\| \left( \frac{\partial}{\partial z} + \tau[v] \frac{\partial}{\partial t} \right) DF[v] \delta v \right\| \leq C(\beta + \max_z \left| \frac{d}{dz} (\tau[v] - \tau[v^*]) \right| + \sigma + r) \cdot \| S_{data}^* \| \| \delta v \| \]

\[ \| DF[v] \delta v \| \leq C \| S_{data}^* \| \| \delta v \| \]

where \( C \) is a constant independent of \( v \) and \( \delta v \). Hence

\[
D^2 J_{dso}[v](\delta v, \delta v) \geq \sigma^2 a[v] \left( \frac{d}{dz} D\tau[v] \delta v \right) \frac{\partial}{\partial t} F[v], a[v] \left( \frac{d}{dz} D\tau[v] \delta v \right) \frac{\partial}{\partial t} F[v]\)

\[
-\sigma^2 O(\max_z \left| \frac{d}{dz} (\tau[v] - \tau[v^*]) \right| + \beta + \sigma + r) \| S_{data}^* \| \| \delta v \| ^2
\]

\[
= \sigma^2 (a[v^*] \left( \frac{d}{dz} D\tau[v] \delta v \right) \frac{d}{dt} f^*(t - \tau[v^*]), a[v^*] \left( \frac{d}{dz} D\tau[v] \delta v \right) \frac{d}{dt} f^*(t - \tau[v^*])
\]

\[
-\sigma^2 O(\max_z \left| \frac{d}{dz} (\tau[v] - \tau[v^*]) \right| + \beta + \sigma + r) \| S_{data}^* \| \| \delta v \| ^2
\]

\[
\geq \sigma^2 \min_z \left| a[v^*] \right|^2 \frac{d}{dt} f^* \| ^2 \| \frac{d}{dz} D\tau[v] \delta v \| ^2
\]

\[
-\sigma^2 O(\max_z \left| \frac{d}{dz} (\tau[v] - \tau[v^*]) \right| + \beta + \sigma + r) \| S_{data}^* \| \| \delta v \| ^2
\]

We will consider the problem in the conic set

\[ D_v V_p \equiv \{ \delta v : \frac{d}{dz} D\tau[v] \delta v \| \geq \gamma \| \delta v \| \} \]
for some $\gamma > 0$.

**Theorem 4.6** If the noise level $\tau$ is small enough, $f^*$ is oscillatory enough, and $\sigma$ is small enough, $J[v]$ is convex in $D_0 V_\rho$ for $v$ in a neighborhood of $\nu^*$ in $V$.

### 4.5 The limit problem ($\sigma \to 0$)

The results in the previous sections suggests that the DSO objective behaves nicely when the differential semblance parameter $\sigma$ is small, hence it seems very interesting to see what the limit problem as $\sigma \to 0$ is. In this section we will investigate this issue. Temporarily we assume that the data is noise free.

Consider the following two-step approach to the original inverse problem, which served as motivation at the beginning of this chapter:

1. for fixed $v$, do
   
   $$\min_f \frac{1}{2} \|S[v]f - S_{data}\|_{L^2([-\tau(v)(z), \tau(v)(z)])}^2$$

   treating $z$ as a parameter, to get $f_0[v]$;

2. do
   
   $$\min_v \frac{1}{2} \|a[v] \frac{\partial}{\partial z} f_0[v]\|_{L^2(\Omega[v])}^2$$

It is easy to see that

$$f_0[v] = \frac{a[v^*(z)]}{a[v](z)} f^*(t - \tau[v^*(z)] + \tau[v](z)),$$
hence it is differentiable with respective to z, and the above two-step procedure is equivalent to minimizing

\[ J_0[v] = \frac{1}{2}\|a[v](\frac{\partial}{\partial z} + \tau[v]\frac{\partial}{\partial t})\frac{a[v^*]}{a[v]} f^*(t - \tau[v^*](z))\|_{L^2([0,T] \times [\omega,\tau])}^2 \]

In order to make the notation more meaningful we denote \( J_{dso}[v] \) by \( J_\sigma[v] \) and \( F[v] \) by \( F_\sigma[v] \). Then using (4.9) we get

\[ J_\sigma[v] = \frac{1}{2}\|a[v]\frac{\partial}{\partial t} R\|^2 + \frac{\sigma^2}{2}\|a[v](\frac{\partial}{\partial z} + \tau[v]\frac{\partial}{\partial t})\left(\frac{a[v^*]}{a[v]} f^*(t - \tau[v^*])\right) + a[v](\frac{\partial}{\partial z} + \tau[v]\frac{\partial}{\partial t}) R\|^2 \]

Lemma 4.2 implies

\[ \left|\frac{1}{\sigma^2} J_\sigma[v] - J_0[v]\right| \leq O(\sigma) \]

Straightforward calculation shows that the directional derivative of \( J_0 \) is

\[ DJ_0[v]\delta v = \left\langle a[v] \left( \frac{d}{dz} D\tau[v]\delta v \right) \frac{\partial}{\partial t} \left( \frac{a[v^*]}{a[v]} f^*(t - \tau[v^*]) \right), \right. \]

\[ \left. a[v] \left( \frac{\partial}{\partial z} + \tau[v]\frac{\partial}{\partial t} \right) \left( \frac{a[v^*]}{a[v]} f^*(t - \tau[v^*]) \right) \right\rangle - \left\langle a[v] \left( \frac{d}{dz} a[v]\delta v \right) \left( \frac{a[v^*]}{a[v]} f^*(t - \tau[v^*]) \right), \right. \]

\[ \left. a[v] \left( \frac{\partial}{\partial z} + \tau[v]\frac{\partial}{\partial t} \right) \left( \frac{a[v^*]}{a[v]} f^*(t - \tau[v^*]) \right) \right\rangle \]

Hence, from Lemma 4.2, we have

\[ \left|\frac{1}{\sigma^2} DJ_\sigma[v]\delta v - DJ_0[v]\delta v\right| \leq O(\sigma)\|\delta v\| \]

Similarly, we can calculate the hessian of \( J_0[v] \) and show that

\[ \left|\frac{1}{\sigma^2} D^2 J_\sigma[v](\delta v, \delta v) - D^2 J_0[v](\delta v, \delta v)\right| \leq O(\sigma)\|\delta v\|^2 \]
From the proof of Lemma 4.2, we know that all the bounds are independent of \( v \in V \), hence we get that \( \frac{1}{\sigma^2} J_\sigma[v] \) uniformly converges to \( J_0[v] \) up to at least the second derivative.

**Theorem 4.7** If the data is noise free, \( \frac{1}{\sigma^2} J_\sigma[v] \) uniformly converges to \( J_0[v] \) up to at least the second derivative.

If the data is not noise free:

\[
S_{data}(z, t) = a[v^*]f^*(t - \tau[v^*](z)) + e(z, t)
\]

where \( e \in L^2([0, Z] \times [0, T]) \), then

\[
f_0[v](z, t) = \frac{a[v^*](z)}{a[v](z)}f^*(t - \tau[v^*](z) + \tau[v](z)) + \frac{1}{a[v](z)}e(z, t + \tau[v](z))
\]

which is not differentiable with respect to \( z \). Hence the above two-step procedure is ill-posed. The DSO approach can be regarded as a regularization of the two-step procedure.

### 4.6 Numerical Examples

We have done a simple numerical experiment to illustrate our theoretical results and the different behaviors of the DSO approach and the usual output least-squares approach. The distance between the two wells is 100 meters. Source is located at the depth 125 meters and 21 receivers are placed evenly in the receiver well, 12.5 meters apart, from the surface of the earth down to depth 250 meters. The sample
interval is .25 mini seconds, and there are 501 samples per trace. Fig. 4.1 is a Ricker source wavelet with peak frequency 500 hz. Fig. 4.2 is the receiver-reparametrized source wavelet. Fig. 4.3 is the seismogram using the above source wavelet and the homogeneous velocity model \( v = 2000 \text{m/s} \). The solid curve in Fig. 4.4 is the DSO objective function with \( \sigma = 1 \), and the dashed curve in Fig. 4.4 is the DSO objective function with \( \sigma = 0 \) over 11 evenly distributed constant velocity models from 1600 m/s to 2400 m/s. This comparison indicates that the convexity of the DSO objective function is a contribution of the differential semblance term. Fig. 4.5 is the OLS objective function over the same range of velocity models. The difference is clear. The DSO objective function is nicely convex while the OLS objective function is flat away from the global minimizer. Fig. 4.6 is the estimate of the source wavelet using a wrong velocity \( v = 2400 \text{m/s} \). The receiver-dependent nature of the estimated source wavelet indicates that the velocity is not correct. Fig. 4.7 is the estimate of the source wavelet, which is flat, \( i.e., \) independent of the receiver location, at the correct velocity \( v = 2000 \text{m/s} \). This difference is an illustration of our idea that the flatness of the estimated source wavelet is a measure of the correctness of the current velocity.

We have also done a more realistic numerical example to demonstrate that the DSO approach is very much like the OLS approach in the presence of caustics. Fig. 4.8 is a velocity model with a slow velocity zone. Outside the slow velocity zone, the velocity is 2000 m/s. The model is plotted with a 90° clockwise rotation. The source is placed at the middle of the "100" line, and 21 receivers are evenly placed on the
“200” line from 50 meters to 250 meters. Fig. 4.9 is the seismogram produced by this velocity model with minimum 1800 m/s. It can be seen that caustics are about to be formed. Fig. 4.10 is the DSO objective. Fig. 4.11 is the seismogram produced by this velocity model with minimum 1500 m/s, which contains multiple arrivals and caustics. Fig. 4.12 is a comparison of the DSO and OSL objective functions, which shows that our non-caustic assumption is necessary for our formulation of the DSO approach.
Figure 4.1  Ricker source wavelet

Figure 4.2  Receiver-reparametrized source wavelet
Figure 4.3  Seismogram

Figure 4.4  The solid curve is DSO objective function with $\sigma = 1$. The dashed curve is DSO objective function with $\sigma = 0$
Figure 4.5 OLS objective function

Figure 4.6 Estimation of the source wavelet with a wrong velocity model
Figure 4.7  Estimation of the source wavelet with the correct velocity model

Figure 4.8  Velocity model with a slow zone
Figure 4.9  Seismogram produced with velocity model in Figure 4.8. The minimum of the velocity is 1800 m/s.

Figure 4.10  DSO objective function with $\sigma = 1$. 
Figure 4.11  Seismogram produced with velocity model in Figure 4.8. The minimum of the velocity is 1500 m/s.

Figure 4.12  The solid curve is DSO objective function with $\sigma = 1$. The dashed curve is OLS objective function.
Chapter 5

Waveform Inversion via DSO – Multiple Arrivals

5.1 Introduction

The analysis and numerical examples in the previous chapter fail in the presence of multi-arrivals, which usually appear along with caustics, in the seismograms. A ready resolution of the problem is that we separate the multi-arrivals, as we did for traveltime data, but it is much more complicated with waveform data. In this chapter we present some analytical results about crosswell wave propagations and an analogue of the local beam-forming formulation in Chapter 3 for waveform inversion of seismic data with multiple arrivals.

All the analysis in this chapter depends on the theory of propagation of singularities. A indispensable tool in this theory is the Hamilton system. For the purpose of later use, we review one version of the Hamilton system governing the rays associated to the acoustic wave propagation and some terminology again. Let the Hamiltonian be

\[ H(x, z, t, \xi, \eta, \tau) = \frac{1}{2} \{-\tau^2 + v(x, z)^2(\xi^2 + \eta^2)\}. \]

Then the Hamilton system is

\[ \dot{x} = \partial_\xi H = v(x, z)^2 \xi \]
\[
\begin{align*}
\dot{z} &= \partial_{\eta}H = v(x, z)^2 \eta \\
\dot{t} &= -\tau \\
\dot{\xi} &= -\partial_x H = -v(x, z)\partial_x v(x, z)(\xi^2 + \eta^2) \\
\dot{\eta} &= -\partial_z H = -v(x, z)\partial_z v(x, z)(\xi^2 + \eta^2) \\
\dot{\tau} &= 0
\end{align*}
\]

where \( \cdot \) means \( \frac{d}{ds} \) and \( s \) is a parameter along a ray. The trajectories of this ordinary differential system are called \textit{bicharacteristic strips}. It is easy to check that \( H \) is constant on the bicharacteristic strips. The bicharacteristic strips on which \( H = 0 \) are called \textit{null bicharacteristic strips}. The projections of the null bicharacteristic strips into the \((x, z, t)\) space are called \textit{(characteristic) rays}. From the last equation in the above system we see that \( \tau \) is a constant, and we only consider the problem forward in time, we can assume that \( \tau = 1 \). Now the Hamilton system becomes

\[
\begin{align*}
\dot{x} &= v(x, z)^2 \xi \\
\dot{z} &= v(x, z)^2 \eta \\
\dot{\xi} &= -v(x, z)\partial_x v(x, z)(\xi^2 + \eta^2) \\
\dot{\eta} &= -v(x, z)\partial_z v(x, z)(\xi^2 + \eta^2)
\end{align*}
\]

which in fact is the Hamilton system with Hamiltonian \( \frac{1}{2}v(x, z)^2(\xi^2 + \eta^2) \). Given initial conditions, say,

\[ x(0) = x_0 \]
\[ z(0) = z_0 \]
\[ \xi(0) = \xi_0 \]
\[ \eta(0) = \eta_0 \]

we can trace the ray with the starting position \((x_0, z_0)\) and starting direction \((\xi_0, \eta_0)\) through the medium by solving the initial value problem of the Hamilton system.

An outline of this chapter is as follows. In Section 2, a numerical example which shows that when multiple arrivals are present in the data, straightforward DSO objective function is very much like that of output least squares, and an initial analysis why this happens are presented. In section 3, we treat the forward mapping of the crosswell wave propagation problem as a Fourier integral operator and study its wavefront relation and the wavefront set of the solution. In order to single out an arrival branch, we need to microlocalize the source. In Section 4, we study the effects of the microlocalization of the source on the solution, and construct microlocal cutoff operators whose action on a synthetic full data (i.e., generated with a full source and known velocity) is the “same” as the data produced by a microlocal source in the sense that they have same wavefront sets. In Section 5, we present a local formulation of waveform inversion via DSO for seismic data with multiple arrivals, using the beam-forming idea.
5.2 Necessity of Branch-Separation

We use a simple example to demonstrate the necessity of separating arrival branches. We can use the limit behavior of the DSO objective function to see formally why multiple arrivals cause trouble.

Suppose we have two arrivals, linked by a folding caustic. Let \( \tau_1 \) denote the direct arrival time, which corresponds to the rays not tangent to the caustic, and \( \tau_2 \) denote the caustic arrival time, which corresponds to the rays tangent to the caustic. According to Stickler et al. (1981), away from caustics, the impulsive response near the direct arrival time is asymptotically

\[
A_1 \delta(t - \tau_1)
\]

and the impulsive response near the caustic arrival time is asymptotically

\[
A_2 (t - \tau_2)^{-1}
\]

where \( A_1 \) and \( A_2 \) are quantities related to the solutions of the generalized transport equations. Given source time function \( f \), we can get the wavefield asymptotically by convolving the above functions with \( f \). The contribution from the direct rays to the wavefield is

\[
A_1 f(t - \tau_1)
\]

as expected; the contribution from the caustic rays to the wavefield is

\[
A_2 (f * \frac{1}{t})(t - \tau_2).
\]
For the sake of simplicity we drop the amplitude. The forward map is

$$S[v]f = \sum_{i=1}^{2} \frac{\partial}{\partial t} f(t - \tau_i[v])$$

Let the data be

$$S_{\text{data}} = \sum_{i=1}^{2} \frac{\partial}{\partial t} f^*(t - \tau_i[v^*])$$

The objective function of the limit case is

$$J_0[v] = \frac{1}{2} \| \frac{\partial}{\partial z} f_0[v] \|^2$$

where

$$f_0[v] = (T_{-\tau_1[v]} + T_{-\tau_2[v]})^{-1} \bigg( \sum_{i=1}^{2} \frac{\partial}{\partial t} f^*(t - \tau_i[v^*]) \bigg)$$

$$= (I + T_{\tau_1[v]-\tau_2[v]})^{-1} \bigg( \frac{\partial}{\partial t} f^*(t) + \frac{\partial}{\partial t} f^*(t - \tau_2[v^*] + \tau_1[v]) \bigg)$$

Even when $v$ is close to $v^*$, $\tau_2[v^*] - \tau_1[v]$ still might be large, hence the cross term

$$f^*(t - \tau_2[v^*] + \tau_1[v])$$

might depend on $z$ quite non-trivially, in the sense that the norm of its $z$ derivative is large. Hence $J_0$ might be quite non-convex. From Theorem 4.7, we expect that this is also true for the DSO objective function.

In order to use a simple constant velocity to illustrate this observation, we use a reflector on the top of the two wells to produce a later arrival. Figure 5.1 is an illustration of the rays. Figure 5.2 is the seismogram. Figure 5.3 is the objective function of the conventional DSO formulation, which is not convex.
5.3 The forward mapping as a Fourier integral operator

In order to apply the "beam-forming" idea to waveform inversion, we need theory and techniques from microlocal analysis for the wave equation. First, we prove that the forward map is a Fourier integral operator. The initial value problem of the wave equation in 2 dimensions:

\[
\frac{1}{v^2(x,z)} \frac{\partial^2}{\partial t^2} u(x,z,t) - \Delta u(x,z,t) = f(t) \delta(x - x_s, z - z_s)
\]

\[
u(x,z,t < 0) = 0
\]

\[
\frac{\partial}{\partial t} u(x,z,t < 0) = 0
\]

where \(v\) is the sound speed of the medium, \(f(t)\) is the source wavelet, and \((x_s, z_s)\) is the source location.

For the sake of simplicity, we consider the problem in the whole space, that is \((x,z) \in \mathbb{R}^2\), and \(t \in \mathbb{R}^+\). The forward crosswell problem is that, given \(f\) and \(v\), looking for \(u(x = x_r, z, t)\) where \(x_r\) is a fixed number. We will consider the mapping from \(f\) to \(u(x = x_r, z, t)\).

The solution operator of the above problem, denoted as

\[
S : D' (\mathbb{R}^+) \rightarrow D' (\mathbb{R} \times \mathbb{R}^+),
\]

can be written as a composition of several simpler operators when each composition makes sense:

\[
S = T \circ A \circ M
\]
where $M$ is defined as

$$M : D'(R^+) \longrightarrow D'(R^2 \times R^+),$$

$$f(t) \longmapsto f(t)\delta(x - z, z - z),$$

which is well-defined since it is nothing but a tensor product of distributions, and it is easy to show that its wavefront relation is

$$WF'(M) \subset \{(x, z; \xi, \eta, \tau), (t; \tau) :\}

\{(\xi, \eta) \text{ and } \tau \text{ are not zero at the same time};\}

$A$ is the solution operator of the initial value problem of the wave equation:

$$A : D'(R^2 \times R^+) \longrightarrow D'(R^2 \times R^+)$$

which sends $f\delta$ to $u$, and its wavefront relation is (see Duistermaat, Theorem 5.1.6)

$$WF'(A) \subset \{(x, z; \xi, \eta, \tau), (x', z'; \xi', \eta', \tau') :\}

\{\text{they are on the same null bicharacteristic};\}

Finally $T$ is the restriction operator

$$T : D'(R^2 \times R^+) \longrightarrow D'(R \times R^+),$$

$$u(x, z, t) \longmapsto u(x = L, z, t),$$

and its wavefront relation is (see Duistermaat p.56)

$$WF'(T) \subset \{(z; \eta, \tau), (x, z, t; \xi, \eta, \tau) \text{ for arbitrary } \xi \text{ and } \eta\}$$
Applying the composition theorem of wavefront relations (Theorem 1.3.7 of Duistermaat), we get

**Theorem 5.1** The operators $A \circ M$ and $S$ are well defined and their wavefront relations are

\[ WF'(A \circ M) \subset \{(x, z, t; \xi, \eta, \tau), (t'; \tau') : t \geq t', \]

there exist $(\xi', \eta')$ such that $(\xi', \eta', \tau')$ nonzero and

that $(x, z, t; \xi, \eta, \tau)$ and $(x_s, z_s, t'; \xi', \eta', \tau')$ are on the

same nullbi – characteristic\}

and

\[ WF'(S) \subset \{(z, t; \eta, \tau), (t'; \tau') : \text{there exist } \xi, \xi' \text{ and } \eta' \]

such that $(x_r, z, t; \xi, \eta, \tau)$ and $(x_s, z_s, t'; \xi', \eta', \tau')$

on same null bicharacteristic\}

**Proof** The calculations on the wavefront relations are routine. We only need to show that the trace operator makes sense. This trace operator applies to all distributions whose wavefront set does not contain phase variables at any point on the hypersurface $\{x = x_r\}$ in $\mathbb{R}^3$ which have the tangent space of the hypersurface at that point in its null space, or in other words for this simple instance, the wavefront set restricted to $\{x = x_r\}$ does not perpendicular to $\{x = x_r\}$. This means that the trace operator applies to a solution of the wave equation if the wavefront set of the solution does
not contain elements like \((x_r, z, t; \xi, 0, 0)\). But this is true since from the Hamiltonian function, if \(\tau\) is zero, \(\xi\) and \(\eta\) must also be zero for any point on a null bi-characteristic strip, hence the temporal-spectral variables in every element of \(WF'(A)\) is nonzero, and the claim follows from the way in which the wavefront set of the solution is calculated from \(WF(A)\) and the wavefront set of the source term. q.e.d.

As a consequence of the last theorem, we have

**Corollary 5.1** The wave front set of \(Sf\) is

\[
WF(Sf) \subset \{(z, t; \eta, \tau) : \text{there exist } \xi^0, \eta^0 \text{ in } R, (t^0, \tau^0) \text{ in } WF(f) \text{ and } \xi \text{ in } R, \text{such that } (x_s, z_s, t^0; \xi^0, \eta^0, \tau^0) \text{ and } (x_r, z, t; \xi, \eta, \tau) \text{ are on the same null bicharacteristic strips}\}
\]

**Theorem 5.2** \(S\) is a Fourier integral operator of order \(-3/4\) with canonical relation

\[
\Gamma \subset \{(x, z, t, \xi, \eta, \tau; t', \tau') : \text{there exist } (\xi', \eta') \text{ such that } (\xi', \eta', \tau') \text{ nonzero and that } (x, z, t; \xi, \eta, \tau) \text{ and } (x_s, x, t'; \xi', \eta', \tau') \text{ are on the same null bi-characteristics}\}
\]

if no ray is tangent to the receiver well.

**Proof** Let \(\chi(x, z, t)\) be a smooth function with support away from the source and be one on the hypersurface \(\{(x, z, t) : x_r, z \in R, t \in R^+\}\). Then we can write the
forward mapping as

\[ Sf = T \chi A(f \delta). \]

From Theorem 5.1.6 of Duistermaat, we know that \( A \) can be treated as a Fourier integral operator \( A_1 \) away from the diagonal of \( T^\ast(R^3_{x,t}) \) of order \(-3/2\), associated to the canonical relation

\[ \Gamma_1 \in \{(x, z, t, \xi, \eta, \tau; x', z', t', \xi', \eta', \tau') : t \geq t', \tau > 0, (x, z, t, \xi, \eta, \tau) \text{ and } (x', z', t', \xi', \eta', \tau') \text{ are on the same null bicharacteristic strip}\} \]

Since the singular supports of \( \chi \) and \( f \delta \) are disjoint, we only need \( S \) away from the diagonal. Therefore upto a smoothing operator we have

\[ Sf = T \circ A_1 (f \delta). \]

or

\[ S = T \circ A_1 \circ M. \]

First we prove that \( A_1 \circ M \) is an FIO. Assume that the kernel of \( A_1 \) has local representation

\[ \int a(x, z, t, x', z', t', \theta) e^{i\phi(x, z, t, x', z', t', \theta)} d\theta \]

with \( \theta \in R^N \), \( \phi \) a non-degenerate phase function associated to \( \Gamma_1 \), and \( a \in S^{N}_{1,0}(R^3 \times R^3 \times R^N) \). Then

\[ (A_1 \circ M)f = \int e^{i\phi(x, z, t, x, z, t', \theta)} a(x, z, t, x, z, t', \theta) f(t') dt' d\theta \]
According to Theorem 4.1.7 of Hormander(1971), in order to prove that $a(x, z, t, x_s, z_s, t', \theta)$ is a nondegenerate phase function in $R^3 \times R \times R^N$, we need to prove that if $\lambda \in \Gamma_1$, then $\lambda$ is not in $N(R^3 \times R)$, the normal bundle of $R^3 \times R$, and that $\Gamma_1$ intersects $T^*(R^3 \times R^3)|_{R^3 \times R}$ transversally. The proof of the first part is similar to the proof of Theorem 1. For the second part, it suffices to prove that the map

$$\Gamma_1 \longrightarrow R^4_{x,z,t,t'}$$

$$\begin{array}{ccc}
(x, z, t, \xi, \eta, \tau; x', z', t', \xi', \eta', \tau') & \longrightarrow & (x, z, t, t')
\end{array}$$

has full rank. This is true because the map is surjective, i.e., given any $(x, z, t, t') \in R^4_{x,z,t,t'}$, we can choose arbitrary $\xi, \eta$ and $\tau$, and solve the Hamilton system for $x', z', \xi', \eta'$ and $\tau'$. Therefore from Theorem 4.1.7 of Hormander(1971), $\Gamma$ is a lagrangean manifold, $\phi(x, z, t, x_s, z_s, t', \theta)$ is non-degenerate and is associated to the lagrangean manifold $\Gamma$. Further $a(x, z, t, x_s, z_s, t', \theta) \in S_{1,0}(R^3 \times R^N)$ and is non-zero at each point $\{(x, z, t, t') : \phi(x, z, t, t', \theta)\}$. The FIO has order -1, associated to the canonical relation $\Gamma$.

It is well known that the trace operator $T$ is an FIO of order $\frac{1}{4}$ associated to the canonical relation $WF(T)$ (Duistermaat, chapter 5). Now apply the composition theorem of FIO's, we get that $S$ is an FIO of order $-3/4$ associated to the canonical relation $WF(S)$, if no ray is tangent to the receiver well.

q.e.d.

In DSO, we very often need to treat the single source, single receiver mapping as the forward map. Note that the trace operator is an FIO of order $1/4$, hence the forward map defined by a single source, single receiver is an FIO of order $-1/2$, as long
as the ray connecting the source location and the receiver location does not tangent to the receiver well.

5.4 Microlocal source and microlocal cutoff operators

Wavefield generated by a point impulsive source expands into the medium in all directions. Sometimes it is desired to concentrate on just part of the wave field propagating in certain limited directions. One way wave equations used in constructing absorbing boundary conditions and seismological migration are examples of the idea, where the wavefield is decomposed into two parts, called upgoing and downgoing, or ingoing and outgoing, etc, respectively with a chosen reference direction. This can be understood clearly if we use the language of rays and wavefronts. The rays are divided into two groups and a wavefront is divided two semi-spheres. In the case of varying velocities, rays might cross each other and a wavefront might get self-folded, consequently multi-arrivals are generated (see Figures 3.1 - 3.2). We are interested in tracing a certain group of rays which do not cross each other, or a portion of the wavefront which does not get self-folded, in a certain period of time. This is the motivation of the utilization of the microlocal sources.

We will concentrate our study of microlocal sources on the model problem, the acoustic wave equation with constant density in two dimensions:

\[
\frac{1}{v(x,z)^2} \frac{\partial^2}{\partial t^2} - \Delta)u(x,z,t) = f(t)\delta(x - x_s, z - z_s).
\] (5.1)
Denote by $R^2$ the two dimensional Euclidean space as the configrational space, and $R_2$ the two dimensional Euclidean space as the spectral space. Then the cotangent bundle of $R^2$ is $R^2 \times R_2$. We use $(x, z)$ denote an element in $R^2$ and $(\xi, \eta)$ denote an element in $R_2$. Let $\Gamma$ be a conic open subset of $R_2$ and $\Gamma_0$ be a conic open subset of $\Gamma$, and their boundaries do not tangent to each other. Let $p(\xi, \eta)$ be a non-negative $C^\infty(R_2 \setminus 0)$ function which is 1 in $\Gamma_0$ and 0 outside of $\Gamma$ constructed by defining it on the unit circle first then extending it homogeneously radially. This function can be changed slightly near the origin to make a symbol function. We assume the this has been taken care of. Let $P$ denote the pseudodifferential operator of order one with symbol $p(\xi, \eta)$. We consider the following initial value problem

$$
\left( \frac{1}{u(x, z)^2} \frac{\partial^2}{\partial t^2} - \Delta \right) u(x, z, t) = f(t)(P\delta)(x - x_s, z - z_s)
$$

$$
u(x, z, t) = \frac{\partial}{\partial t} u(x, z, t) = 0 \text{ for } t \ll 0
$$

We are interested in the singular support of the "seismogram" $s(z, t) \equiv u(x_r, z, t)$, i.e., the trace of the wavefield at the receiver well.

**Theorem 5.3** SingSupp $s(z, t) \subset \{(z, t) : z \in [z_1, z_2], t \in \{ \text{SingSupp } f \text{ shifted by the travel time at } (x_r, z) \} \}$, where the interval $[z_1, z_2]$ is the subset of the receiver line covered by the rays from the source point $(x_s, z_s)$ with directions lying in the cone $\Gamma$.

**Proof** Note that

$$WF(P\delta) \subset ess.supp P \cap WF(\delta) = \{(x_s, z_s; \xi, \eta) : (\xi, \eta) \in \Gamma\}$$
\[ WF(fP\delta) \subset \{(x_s, z_s, t; \xi, \eta, \tau) : (\xi, \eta) \in \Gamma, (t, \tau) \in WF(f)\} \]

and use Theorem 5.1.

q.e.d.

Now we present a way to construct cutoff operators for the source and corresponding cutoff operators for synthetic seismic data, giving the velocity model. In order to explain the idea, we assume that the model has a slow velocity region and produces a simple triplization in the seismogram. By shooting rays clockwise and marking the places where the emerging position has a critical point, i.e., it begins to return, we can divide the shooting angles into three groups. Denote the three conic subsets of \( R^{2}_{\xi, \eta} \) by \( \Gamma_1, \Gamma_2, \) and \( \Gamma_3 \), the three corresponding cutoff functions, as constructed earlier this section, by \( p_1, p_2, \) and \( p_3 \), and the corresponding pseudodifferential operators \( P_1, P_2, \) and \( P_3 \). Let \( u(x, z, t) \) be the solution of the acoustic wave equation. Then

\[
\left( \frac{1}{v(x, z)^2} \frac{\partial^2}{\partial t^2} - \Delta \right)(P_i u)(x, z, t) = f(t)(P_i \delta)(x - x_s, z - z_s) + R(f\delta)
\]

where \( R \) is a smoothing operator. Hence the wavefield produced by the microlocal source \( fP_i \delta \) is \( P_i u \), neglecting a smooth error. Since the seismogram \( s(z, t) = u(x, z, t) \) is only a trace of the wavefield \( u \), we need to construct cutoff operators \( Q_i \)'s from \( P_i \)'s, such that \( Q_i s \) and the trace of \( P_i u \) on the hypersurface \( \{x = x_r\} \), have the same wavefront sets. Since

\[
P_i u(x, z, t) = \int e^{i(x - x')\xi + (z - z')\eta + (t - t')\tau} p_i(\xi, \eta) u(x', z', t') dxdz'dtd\xi d\eta d\tau,
\]

note that

\[ WF(P_i u) \subset \{(x, z, t; \xi, \eta, \tau) : (\xi, \eta) \in \Gamma_i, \text{ and } (x, z, t; \xi, \eta, \tau) \text{ is on a} \]
null bicharacteristic from the source

and

\[ \text{WF}(P_i u|_{x=x_r}) \subset \{ (z, t; \eta, \tau) : \text{there exists } \xi \text{ such that } (\xi, \eta) \in \Gamma_i, \text{ and} \]
\[ (x_r, z, t; \xi, \eta, \tau) \text{ is on a null bicharacteristic from the source} \} \]

Define \( Q_i \) by

\[ Q_i s(z, t) = \int e^{i[(z-z') n + (t-t') \tau]} p_i(\sqrt{n^2(x_r, z)^2 + \eta^2, \eta}) s(z', t') dz'd\eta'd\tau, \]

It is easy to see that \( (P_i u)|_{x=x_r} \) and \( Q_i s \) have the same wavefront sets under the assumption that there is no returning rays.

5.5 The Beam-forming Idea

As in the case of traveltime tomography in the presence of multiple arrivals (Chapter 3, Section 2), we are only able to formulate the beam-forming approach for waveform inversion locally. We still use the data with a single triplication as example. If a velocity model is close enough to the true velocity model, the synthetic seismogram produced from this velocity will also have a triplication. As in Section 2 of Chapter 3, the rays are grouped into three groups, from each group we can construct a microlocal cut-off operator whose action on the synthetic data will single out a single branch. If the velocity model is close enough to the true model, and the three groups of rays are
shrunked from the bounding rays properly, then the actions of the three microlocal cut-off operators on the data will also produce single arrival branches. Denote the three microlocal cutoff operators constructed in the previous section by \( Q_i, i = 1, 2, 3 \), respectively, then the objective function is

\[
J[v, f] = \frac{1}{2} \sum_{i=1}^{3} (\| Q_i S[v] f - Q_i S_{data} \| + \sigma^2 \| \frac{\partial}{\partial z} f \|^2)
\]

Then, similar to Chapter 3 and Chapter 4, we can prove that the DSO objective function is smooth and \textit{locally} every critical point is close to a kinematically correct model if \( \sigma \) is small enough, the source wavelet is oscillatory enough, and the noise level is low enough.

For the constant velocity example presented in Section 2 of this chapter, it is very easy to group the rays and separate the branches. The DSO objective function with the branches separated is Figure 5.4. The convexity is restored.
Figure 5.1 Using reflector to produce multiple arrivals
Figure 5.3  DSO objective function without branch-separation.

Figure 5.4  DSO objective function with branch-separation.
Chapter 6

Concluding Remarks

In this thesis, we considered traveltime tomography and waveform inversion of cross-well seismograms via DSO, and some related issues. For the cases which do not involve caustics, we formulated the problem as optimization problems, and studied the smoothness and critical points of the objective functions. We also extended the results to local formulations for cases involving caustics.

There are two possible directions for further studies. First, we need to globalize the beam-forming approaches for both traveltime inversion and waveform inversion, and do some more numerical experiments with more realistic velocity models; Second, it would be interesting to carry out similar analyses for reflection inverse problems. Since the forward mapping of the reflection problem is a much more complicated FIO, we expect that the generalization of DSO approach to reflection inversion is even more involved. A lot could been done in these two directions.
Bibliography


