Inversion of Crosswell Seismograms via Differential Semblance Optimization

Hua Song* **William W. Symes***

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. The two common inversion methods *traveltime tomography* and waveform inversion via *output least squares* have some unpleasant features from the viewpoint of numerical computations. We present a new waveform inversion method – *differential semblance optimization* (DSO) – in this paper under the non-caustic assumption. The main result 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.

1 Introduction

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

$$\frac{1}{v^2\rho}\frac{\partial^2 p}{\partial t^2} - \nabla(\frac{1}{\rho}\nabla p) = s \tag{1.1}$$

with appropriate side conditions. We shall assume in this thesis that the source is *isotropic* and *punctual*, i.e., $s(x,z,t) = f(t)\delta(x-x_s)\delta(y-y_s)\delta(z-z_s)$, 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

*The Rice Inversion Project, Department of Computational and Applied Mathematics, Rice University, Houston TX 77251-1892

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 *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 *crosswell inverse problem*, which will be studied in this thesis. We also call the later *transmission inverse problem* if only transmission waves are considered, as in this thesis.

Crosswell inverse problems are important in reservoir characterisation and enhanced oil recovery (Justice *et al* ., 1989). The more or less conventional method to resolve these problems is called *traveltime tomography* (TT), or *traveltime inversion*. One extracts the traveltime information from the seismic data, and find a model to fit the traveltime data in the least-squares sense. See Song and Symes (1994) for a theoretical analysis of this method. 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.

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^o *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.

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 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). DSO combines concepts from nonlinear least-squares inversion,migration velocity analysis, and traveltime tomography. 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 cumbersomeness 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.

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 panelty 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.

This paper presents a theoretical analysis of the differential semblance formulation for the crosswell tomography problem of estimating velocity and source wavelet from crosswell waveform data, under the very important simplifying assumption that no caustics are produced by the velocity structure. As shown in, for example, Lines *et al* 1992, caustics and guided waves are extremely common in crosswell surveys. Therefore our current study is not of immediate practical value. Nonetheless it addresses the principal difficulty in waveform inversion.

First, in Section 2, we present a differential semblance formulation for the crosswell inverse problem of estimating velocity and source wavelet from crosswell waveform data. In section 3, we prove smoothness of the DSO objective function. We analyze the gradient of the objective function and give a characterization of all critical points in Section 4, and analyze the hessian in Section 5. In Section 6, we investigate the limit behavior of the DSO objective as the differential semblance parameter approaches zero. 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. A simple numerical example is presented to show the difference between OLS and DSO approaches.

2 Motivations and DSO 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 Fig. 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

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

where δ is the Dirac delta function, with causality intitial condition

$$p(X, t) = 0 \qquad \text{for } t << 0$$

and an appropriate boundary condition. In order to make the idea simple, we will ignore the boundary, hence the refelections 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

$$g(X, X_s, t) = a(X)\delta(t - \tau(X)) + \text{smoother terms}$$

where τ is the traveltime function which satisfies the eiconal equation

$$|\nabla \tau|^2 = \frac{1}{v^2}$$

and a is the amplitude which satisfies the transport equation

$$\nabla a \cdot \nabla \tau + \frac{1}{2} a \Delta \tau = 0$$

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

$$p(X, t) = a(X) f(t - \tau(X)) + \text{smoother terms}$$

If f is highly oscillatory, as we shall assume, to good approximation

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

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 $\{x = x_r, y = y_r, 0 \le z \le Z\}$. 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 τ as functions of z alone. The *data* of the (idealized) crosswell inverse problem are presumed measurements of the pressure field

$$\{S_{data}(z,t) : 0 \le z \le Z, 0 \le t \le T\}$$

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 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

$$\|S[v]f - S_{data}\|^2 \tag{2.1}$$

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

$$(v, f[v] \equiv \frac{a[v^*]}{a[v]} f^*(t - \tau[v^*] + \tau[v]) \tag{2.2}$$

is a global minimizer of the above least squares problem since it gives zero residual, 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 arbitray global minimizer (2.2) of (2.1) to a pair of kinematically velocity and a source wavelet independent of receiver location z. Besides, since

$$\begin{aligned} \frac{\partial}{\partial z} f[v] &= \frac{\partial}{\partial z}\left(\frac{a[v^*]}{a[v]}\right) f^*(t - \tau[v^*] + \tau[v]) \\ &\quad + \frac{a[v^*]}{a[v]} \frac{d}{dt} f^*(t - \tau[v^*] + \tau[v]) \frac{\partial}{\partial z}(\tau[v] - \tau[v^*]), \end{aligned}$$

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 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 $0 < v_{min} < v_{max}$, $v \in V$ implies $v_{min} \leq v(x, y, z) \leq v_{max}$. We also assume that V contains the velocity model v^* which produces the traveltime data τ_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 traveltime 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.

We first make a technical modification to the definition of the forword 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 *a priori* independent model.

We make a technical assumption on the data:

- The noise free data $S_{data} = a[v^*]\frac{\partial}{\partial t}f^*(t - \tau[v^*])$ 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 $z-independent$ 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 $S[v]$ can be viewed as a mapping $H[v] \longrightarrow L^2([0, Z] \times [0, T])$.

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

$$J[v, f] = \frac{1}{2}\|S[v]f - S_{data}\|^2_{L^2([0,Z]\times[0,T])} + \frac{\sigma^2}{2}\|a[v]\frac{\partial}{\partial z}f\|^2_{L^2(\Omega[v])}. \tag{2.3}$$

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 first two terms zero. The parameters σ 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 $0 \le z_s \le Z_s$, for example, then an obvious modification of the DSO objective function is

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

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 $J[v, f]$ (2.3) 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 $J_{dso}[v] = J[v, f[v]]$, and solve

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

3 Smoothness of the Objective Function

Define an operator

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

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

$$\begin{aligned} T_{-\tau[v]} : L^2(\Omega[v]) &\longrightarrow L^2([0,Z]\times[0,T]) \\ f(z,t) &\mapsto f(z,t-\tau[v](z)) \end{aligned}$$

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

$$\min_{F\in H}\{\frac{1}{2}\|S_1[v]F - S_{data}\|^2 + \frac{\sigma^2}{2}\|a[v](\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})F\|^2$$

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

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

$$\begin{aligned} &< S_1[v]F, S_1[v]\phi > +\sigma^2 < a[v](\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})F, a[v](\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})\phi > \\ &=< S_1[v]\phi, S_{data} >, \quad \forall \phi \in H. \end{aligned}$$

Theorem 3.1 *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*

$$\begin{aligned} \|\frac{\partial}{\partial t}F[v]\| &\leq C\|S_{data}\| \\ \sigma\|(\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})F[v]\| &\leq C\|S_{data}\| \end{aligned}$$

here all $C's$ are constants 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, i.e., which is a first order ordinary differential equation with non-zero initial value, it is nonzero. Denote the mininum of $a[v]$ as a_m.

$$\begin{aligned}
&\int_0^Z \int_0^T a[v]^2 (\frac{\partial}{\partial t} F)^2 + \sigma^2 a[v]^2 \{(\frac{\partial}{\partial z} + \tau[v]' \frac{\partial}{\partial t}) F\}^2 \\
&\geq a_m^2 \min(1, \sigma^2) \int_0^Z \int_0^T (\frac{\partial}{\partial t} F)^2 + \{(\frac{\partial}{\partial z} + \tau[v]' \frac{\partial}{\partial t}) F\}^2 \\
&\geq a_m^2 \min(1, \sigma^2) \int_0^Z \int_{-\tau[v](z)}^{T-\tau[v](z)} (\frac{\partial}{\partial t} f)^2 + (\frac{\partial}{\partial z} f)^2
\end{aligned}$$

Since

$$f(z,t) = 0 \quad \text{for} \quad t = -\tau[v](z) \ \text{ and } \ t = T - \tau[v](z), 0 \leq z \leq Z$$

using the fundamental theorem of calculus we can get

$$\int_0^Z \int_{-\tau[v](z)}^{T-\tau[v](z)} f^2 \leq C \int_0^Z \int_{-\tau[v](z)}^{T-\tau[v](z)} (\frac{\partial}{\partial t} f)^2$$

where $C > 0$. From the above two inequalities we can get

$$\begin{aligned}
&\int_0^Z \int_0^T a[v]^2 (\frac{\partial}{\partial t} F)^2 + \sigma^2 a[v]^2 \{(\frac{\partial}{\partial z} + \tau[v]' \frac{\partial}{\partial t}) F\}^2 \\
&\geq C \int_0^Z \int_{-\tau[v](z)}^{T-\tau[v](z)} (\frac{\partial}{\partial t} f)^2 + (\frac{\partial}{\partial z} f)^2 + f^2 \\
&\geq C \int_0^Z \int_0^T (\frac{\partial}{\partial t} F)^2 + (\frac{\partial}{\partial z} F)^2 + F^2
\end{aligned}$$

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 3.2 *$F[v] \in H$ 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 ODE theory, we know that $\tau[v] \in L^2([0, Z])$ smoothly depends on $v \in V$ (see Song and Symes 1994). Similarly it can be proved that $a[v] \in L^2([0, Z])$ moothly 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 δa be a perturbation of a and λ be a nonzero real number. Let $F(\lambda) = F[a + \lambda\delta a]$. Then $F(\lambda) - F$ satisfies for any $\phi \in H$

$$\begin{aligned}
&\langle a\frac{\partial}{\partial t}(F(\lambda)-F), a\frac{\partial}{\partial t}\phi\rangle + \sigma^2\langle a(\frac{\partial}{\partial z}+\tau'\frac{\partial}{\partial t})(F(\lambda)-F), a(\frac{\partial}{\partial z}+\tau'\frac{\partial}{\partial t})\phi\rangle \\
&= \langle a\frac{\partial}{\partial t}\phi, \lambda\frac{\delta a}{a}S_d - 2\lambda\delta a\frac{\partial}{\partial t}F(\lambda) - \lambda^2\frac{(\delta a)^2}{a}\frac{\partial}{\partial t}F(\lambda)\rangle \\
&+\sigma^2\langle a(\frac{\partial}{\partial z}+\tau'\frac{\partial}{\partial t})\phi, -2\lambda\delta a(\frac{\partial}{\partial z}+\tau'\frac{\partial}{\partial t})F(\lambda) - \lambda^2\frac{(\delta a)^2}{a}(\frac{\partial}{\partial z}+\tau'\frac{\partial}{\partial t})F(\lambda)\rangle
\end{aligned}$$

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

$$\begin{aligned}
&\|a\frac{\partial}{\partial t}(F(\lambda)-F)\|^2 + \sigma^2\|a(\frac{\partial}{\partial z}+\tau'\frac{\partial}{\partial t})(F(\lambda)-F)\|^2 \\
&\le \|a\frac{\partial}{\partial t}(F(\lambda)-F)\|\|\lambda\frac{\delta a}{a}S_d - 2\lambda\delta a\frac{\partial}{\partial t}F(\lambda) - \lambda^2\frac{(\delta a)^2}{a}\frac{\partial}{\partial t}F(\lambda)\| \\
&+\sigma^2\|a(\frac{\partial}{\partial z}+\tau'\frac{\partial}{\partial t})(F(\lambda)-F)\|\|2\lambda\delta a(\frac{\partial}{\partial z}+\tau'\frac{\partial}{\partial t})F(\lambda) + \lambda^2\frac{(\delta a)^2}{a}(\frac{\partial}{\partial z}+\tau'\frac{\partial}{\partial t})F(\lambda)\|
\end{aligned}$$

Now we have got an inequality in the form

$$A^2 + B^2 \le AC + BD$$

Use

$$AC \le \frac{A^2+C^2}{2}, \quad BD \le \frac{B^2+D^2}{2}$$

to the right hand side, we get

$$A^2 + B^2 \le C^2 + D^2$$

Hence we get

$$\begin{aligned}
&\|a\frac{\partial}{\partial t}(F(\lambda)-F)\|^2 + \sigma^2\|a(\frac{\partial}{\partial z}+\tau'\frac{\partial}{\partial t})(F(\lambda)-F)\|^2 \\
&\le \lambda^2\|\frac{\delta a}{a}S_d - 2\delta a\frac{\partial}{\partial t}F(\lambda) - \lambda\frac{(\delta a)^2}{a}\frac{\partial}{\partial t}F(\lambda)\|^2 \\
&+\lambda^2\|2\delta a(\frac{\partial}{\partial z}+\tau'\frac{\partial}{\partial t})F(\lambda) + \lambda\frac{(\delta a)^2}{a}(\frac{\partial}{\partial z}+\tau'\frac{\partial}{\partial t})F(\lambda)\|^2
\end{aligned}$$

Let λ go to zero, we get

$$\|a\frac{\partial}{\partial t}(F(\lambda)-F)\| \longrightarrow 0 \tag{3.1}$$

$$\|a(\frac{\partial}{\partial z}+\tau'\frac{\partial}{\partial t})(F(\lambda)-F)\| \longrightarrow 0 \tag{3.2}$$

Let $DF\delta a$ be the unique solution of the following variational problem

$$\begin{aligned}&\langle a\frac{\partial}{\partial t}DF\delta a, a\frac{\partial}{\partial t}\phi\rangle + \sigma^2\langle a(\frac{\partial}{\partial z}+\tau'\frac{\partial}{\partial t})DF\delta a, a(\frac{\partial}{\partial z}+\tau'\frac{\partial}{\partial t})\phi\rangle\\ &= \langle \delta a\frac{\partial}{\partial t}\phi, \frac{\delta a}{a}S_d - 2\delta a\frac{\partial}{\partial t}F\rangle - \sigma^2\langle a(\frac{\partial}{\partial z}+\tau'\frac{\partial}{\partial t})\phi, 2\delta a(\frac{\partial}{\partial z}+\tau'\frac{\partial}{\partial t})\end{aligned}$$

Then $\frac{F(\lambda)-F}{\lambda} - DF\delta a$ satisfies

$$\begin{aligned}&\langle a\frac{\partial}{\partial t}(\frac{F(\lambda)-F}{\lambda} - DF\delta a), a\frac{\partial}{\partial t}\phi\rangle\\ &\quad+\sigma^2\langle a(\frac{\partial}{\partial z}+\tau'\frac{\partial}{\partial t})(\frac{F(\lambda)-F}{\lambda} - DF\delta), a(\frac{\partial}{\partial z}+\tau'\frac{\partial}{\partial t})\phi\rangle\\ &= \langle a\frac{\partial}{\partial t}\phi, 2\delta a\frac{\partial}{\partial t}(F-F(\lambda)) - \lambda\frac{(\delta a)^2}{a}\frac{\partial}{\partial t}F(\lambda)\rangle\\ &\quad+\sigma^2\langle a(\frac{\partial}{\partial z}+\tau'\frac{\partial}{\partial t})\phi, 2\delta a(\frac{\partial}{\partial z}+\tau'\frac{\partial}{\partial t})(F-F(\lambda)) - \lambda\frac{(\delta a)^2}{a}(\frac{\partial}{\partial z}+\tau'\frac{\partial}{\partial t})F(\lambda)\rangle\end{aligned}$$

Similar to the proof of (3.1)-(3.2) , we can prove that

$$\|a\frac{\partial}{\partial t}(\frac{F(\lambda)-F}{\lambda} - DF\delta a)\|^2 + \sigma^2\|a(\frac{\partial}{\partial z}+\tau'\frac{\partial}{\partial t})(\frac{F(\lambda)-F}{\lambda})\|^2 \longrightarrow 0$$

as λ goes to zero. From the coerciveness of the variational problem, we know that this means F is differentiable and its directional derivative in direction δa is $DF\delta a$. In the same way we can prove that this derivative is continuous and F has contionuous higher derivatives. q.e.d.

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

Theorem 3.3 *$J[v]$ depends smoothly on $v \in V$.*

4 The Gradient and Charaterization of Critical Points

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

$$J_{dso}[v] = \frac{1}{2}\|S_1[v]F[v] - S_{data}\|^2 + \frac{\sigma^2}{2}\|a[v](\frac{\partial}{\partial z}+\tau[v]'\frac{\partial}{\partial t})F[v]\|^2 \tag{4.1}$$

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

Theorem 4.1 *The directional derivative of $J[v]$ is given by*

$$\begin{aligned}DJ_{dso}[v]\delta v \quad = \quad &\sigma^2\langle a[v](\frac{d}{dz}D\tau[v]\delta v)\frac{\partial}{\partial t}F[v], a[v](\frac{\partial}{\partial z}+\tau[v]'\frac{\partial}{\partial t})F[v]\rangle\\ &-\sigma^2\langle a[v](\frac{d}{dz}\frac{Da[v]\delta v}{a[v]})F[v], a[v](\frac{\partial}{\partial z}+\tau[v]'\frac{\partial}{\partial t})F[v]\rangle\end{aligned}$$

Proof. Take directional derivative of (4.1) in direction δv, we get

$$\begin{aligned} DJ_{dso}[v]\delta v &= \langle Da[v]\delta v\frac{\partial}{\partial t}F[v] + a[v]\frac{\partial}{\partial t}(DF[v]\delta v), a[v]\frac{\partial}{\partial t}F[v] - S_{data}\rangle \\ &+\sigma^2\langle Da[v]\delta v(\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})F[v] + a[v](D\tau[v]\delta v)'\frac{\partial}{\partial t}F[v] \\ &+a[v](\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})(DF[v]\delta v), a[v](\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})F[v]\rangle \end{aligned}$$

Since

$$\begin{aligned} &\langle a[v]\frac{\partial}{\partial t}(\frac{Da[v]\delta v}{a[v]}F[v]), a[v]\frac{\partial}{\partial t}F[v] - S_{data}\rangle \\ &+\sigma^2\langle a[v](\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})(\frac{Da[v]\delta v}{a[v]}F[v]), a[v](\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})F[v]\rangle = 0 \end{aligned}$$

and

$$\begin{aligned} &\langle a[v]\frac{\partial}{\partial t}(DF[v]\delta v), a[v]\frac{\partial}{\partial t}F[v] - S_{data}\rangle \\ &+\sigma^2\langle a[v](\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})(DF[v]\delta v), a[v](\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})F[v]\rangle = 0 \end{aligned}$$

because that $F[v]$ satisfies the variational problem, we have

$$\begin{aligned} DJ_{dso}[v]\delta v &= \sigma^2\langle a[v](\frac{d}{dz}D\tau[v]\delta v)\frac{\partial}{\partial t}F[v], a[v](\frac{\partial}{\partial z} + (\frac{d}{dz}\tau[v])\frac{\partial}{\partial t})F[v]\rangle \\ &-\sigma^2\langle a[v](\frac{d}{dz}\frac{da[v]\delta v}{a[v]})F[v], a[v](\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})F[v]\rangle \end{aligned}$$

q.e.d

Corollary 4.1 *If the data is $S_{data} = a[v^*]\frac{d}{dt}f^*(t - \tau[v^*])$, then $v = v^*$ is a global minimizer of the DSO objective function.*

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 δv. 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

$$\begin{aligned} \max_z |\frac{d}{dz}D\tau[v]\delta v| &\le C\|\delta v\|_2, \\ \max_z |\frac{d}{dz}Da[v]\delta v| &\le C\|\delta v\|_3 \\ \max_z |Da[v]\delta v| &\le C\|\delta v\|_3 \end{aligned}$$

where C is independent of $v \in V$ and δv.

Proof. $D\tau[v]\delta v$ is given by

$$D\tau[v]\delta v(z) = \oint_{X_s}^{(x_r,z)} (-\frac{\delta v}{v^2})(x(x), z(s))ds$$

where s is the raypath length from the source. For any $z \in [0, Z]$, there is a ray linking the source X_s and the point (x_r, z). For any point on the ray with x-axis $x \in [x_s, x_r]$, we denote its z-axis by $y(x, z)$. Now the formula for $D\tau[v]\delta v$ becomes

$$D\tau[v]\delta v = \int_{x_s}^{x_r} (-\frac{\delta v}{v^2})(x, y(x, z))J(x, z)dx \tag{4.2}$$

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

$$\begin{aligned} \frac{\partial}{\partial z} D\tau[v]\delta v &= \int_{x_s}^{x_r} -\frac{\partial}{\partial y}(\frac{\delta v}{v^2})(x, y(x, z))\frac{\partial y(x, z)}{\partial z} J(x, z) \\ &\quad -\frac{\delta v}{v^2}(x, y(x, z))\frac{\partial J(x, z)}{\partial z}dx \end{aligned} \tag{4.3}$$

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 bondary 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 $Da[v]\delta v$:

$$Da[v]\delta v = e^{-\int \frac{1}{2}v\Delta\tau[v]ds} \left\{ \int e^{\int \frac{1}{2}v\Delta\tau[v]} - v\nabla a[v] \cdot \nabla(D\tau[v]\delta v) - v\frac{1}{2}a[v]\Delta(D\tau\delta v) \right\}$$

A conservative estimate gives

$$\begin{aligned} \max_z |\frac{d}{dz} Da[v]\delta v| &\le C\|\delta v\|_3 \\ \max_z |Da[v]\delta v| &\le C\|\delta v\|_3 \end{aligned}$$

where C is independent of v and δv. q.e.d.

Let

$$\begin{aligned} S^*_{data} &= a[v^*]\frac{d}{dt} f^*(t - \tau[v]) \\ S_{data} &= S^*_{data} + e \end{aligned}$$

and

$$r = \frac{\|e\|}{\|S^*_{data}\|}.$$

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

$$F[v] = \frac{a[v^*]}{a[v]} f^*(t - \tau[v^*]) + R(z,t) \tag{4.4}$$

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

$$\begin{aligned} \|\frac{\partial}{\partial t} R\| &\leq C(\sigma^2 + r)\|S^*_{data}\| \\ \|(\frac{\partial}{\partial z} + \tau[v]' \frac{\partial}{\partial t})R\| &\leq C(\sigma + r)\|S^*_{data}\| \\ \|R\| &\leq C(\sigma + r) \end{aligned}$$

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 assumtion on data we have $R \in H$ and

$$\begin{aligned} &\langle a\frac{\partial}{\partial t}R, a\frac{\partial}{\partial t}\phi\rangle + \sigma^2 \langle a(\frac{\partial}{\partial z} + \tau'\frac{\partial}{\partial t})R, a(\frac{\partial}{\partial z} + \tau'\frac{\partial}{\partial t})\phi\rangle \\ &= -\sigma^2\langle a\{\frac{d}{dz}(\frac{a^*}{a})\}f^*(t-\tau^*) + a^*(\tau' - \tau^{*'})\frac{d}{dt}f^*(t-\tau^*), a(\frac{\partial}{\partial z} + \tau'\frac{\partial}{\partial t})\phi\rangle \\ &\quad + \langle e, a\frac{\partial}{\partial t}\phi\rangle \end{aligned}$$

Let $\phi = R$, then

$$\begin{aligned} &\|a\frac{\partial}{\partial t}R\|^2 + \sigma^2\|a(\frac{\partial}{\partial z} + \tau'\frac{\partial}{\partial t})R\|^2 \\ &\leq \sigma^2\{\|a\{\frac{d}{dz}(\frac{a^*}{a})\}f^*(t-\tau^*)\| + \|a^*(\tau' - \tau^{*'})\frac{d}{dt}f^*(t-\tau^*)\|\}\|a(\frac{\partial}{\partial z} + \tau'\frac{\partial}{\partial t})R\| \\ &\quad + \|e\|\|a\frac{\partial}{\partial t}R\| \end{aligned}$$

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

$$\begin{aligned} \|a\frac{\partial}{\partial t}R\| &\leq \sigma^2\|a\{\frac{d}{dz}(\frac{a^*}{a})\}f^*(t-\tau^*)\| + \|a^*(\tau' - \tau^{*'})\frac{d}{dt}f^*(t-\tau^*)\| + \|e\| \\ \|a(\frac{\partial}{\partial z} + \tau'\frac{\partial}{\partial t})R\| &\leq \sigma\|a\{\frac{d}{dz}(\frac{a^*}{a})\}f^*(t-\tau^*)\| + \|a^*(\tau' - \tau^{*'})\frac{d}{dt}f^*(t-\tau^*)\| + \|e\| \end{aligned}$$

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 3.1, we have

$$\|R\| \le \|R\|_H \le C\{\|\frac{\partial}{\partial t}R\| + \|(\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})R\|\} \le C(\sigma + r)$$

q.e.d.

Using (4.4), we can get a L^2 estimate of $F[v]$ and improvement of the inequalities in Theorem 3.1.

Corollary 4.2

$$\|F[v]\| \le C(\beta + \sigma + r)\|S^*_{data}\|$$
$$\|a[v]\left(\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t}\right)F[v]\| \le (C_0 + C_1\beta + C_2\sigma + C_3 r)\|S^*_{data}\|$$

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

Proof. The desired inequality follows 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 $DJ_{dso}[v]\delta v$ given in Theorem 4.1 and use the estimates in Theorem 3.1, Lemmas 4.1 and 4.2, and Corollary 4.2, the directional derivative of $J[v]$ can be written as

$$\begin{aligned} DJ_{dso}[v]\delta v &\le \sigma^2\left\{\int\int a[v^*]^2(\frac{d}{dz}D\tau[v]\delta v)\frac{d}{dz}(\tau[v] - \tau[v^*])\left[\frac{d}{dt}f^*(t - \tau[v^*])\right]^2 dz\, dt\right. \\ &\quad \left. + (\sigma + r)C_1(\|\delta v\|_2\| + \|\delta v\|_3)\frac{d}{dt}f^*\|^2 + C_2(\|\delta v\| + \|\delta v\|_3)\|f^*\|\,\|\frac{d}{dt}f^*\|\right\} \\ &= \sigma^2\left\{\int a[v^*]^2(\frac{d}{dz}D\tau[v]\delta v)\frac{d}{dz}(\tau[v] - \tau[v^*])dz\right. \\ &\quad \left. + (\sigma + r)C_1(\|\delta v\|_2 + \|\delta v\|_3) + \beta C_2(\|\delta v\|_2 + \|\delta v\|_3)\right\}\|\frac{d}{dt}f^*\|^2 \end{aligned}$$

where $C_i, i = 1, 2$, are independent of $\delta v, v$, and σ, and $\beta \equiv \|f^*\|/\|\frac{d}{dt}f^*\|$ 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 ray-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 (4.2). 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

$$\begin{aligned} DJ_{dso}[v]\delta v &= \sigma^2 \| \frac{d}{dt} f^* \|^2 \{ - \int \delta v (D\tau[v])^* \frac{d}{dz} \{ a[v^*]^2 \frac{d}{dz} (\tau[v] - \tau[v^*]) \} \\ &\quad + C_1(\sigma + r)(\|\delta v\|_2 + \|\delta v\|_3) + C_2 \beta (\|\delta v\|_2 + \|\delta v\|_3) \}. \end{aligned}$$

Let $GradJ_{dso}[v]$ be the L^2 gradient of $J_{dso}[v]$, i.e.,

$$DJ_{dso}[v]\delta v = < \delta v, GradJ_{dso}[v] >$$

then we have

$$\begin{aligned} \|GradJ_{dso}[v]\| &= \sigma^2 \| \frac{d}{dt} f^* \|^2 \{ \| (D\tau[v])^* \{ \frac{d}{dz} a[v^*]^2 \frac{d}{dz} (\tau[v] - \tau[v^*]) \} \| \\ &\quad + O(\sigma) + O(r) + O(\beta) \} \\ &\geq \sigma^2 C \| \frac{d}{dz} \{ a[v^*]^2 \frac{d}{dz} (\tau[v] - \tau[v^*]) \} \| + O(\sigma) + O(r) + O(\beta) \end{aligned}$$

here we have used the coercivity of $(D\tau[v])^*$ (see Lemma 2.2 of Song and Symes 1994) 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 δv in the rest of the paper.

If the gradient of $J_{dso}[v]$ is zero at v, then

$$\| \frac{d}{dz} \{ a[v^*]^2 \frac{d}{dz} (\tau[v] - \tau[v^*]) \| \leq C(\sigma + \beta + r)$$

where C is a constant independent of v and δv. We summarize the above discussion as the following theorem.

Theorem 4.2 *For noise free data, all critical points of the DSO objective function satisfies*

$$\| \frac{d}{dz} \{ a[v^*]^2 \frac{d}{dz} (\tau[v] - \tau[v^*]) \} \| \leq C(\sigma + \beta + r)$$

where $\beta = \|f^*\| / \| \frac{d}{dt} f^* \|$, *and* C *is a positive constant independent of* v.

Note that if we assume that v and v^* give same traveltimes near the ends of the receiver array, i.e.,

$$\tau[v](z) = \tau[v^*](z), \text{near } \ z = 0, Z$$

(this asumption is satisfied if v and v^* differ only in the region well-covered by rays), then Poincare's inequality leads to

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

This result says that if the differential semblance σ 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 well-covered by rays, the*

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

This lemma says that if σ and r are small enough and the source wavelet is oscillatory enough, a critical point will be close to the true velocity model kinematically.

For the estimation of the source wavelet, define

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

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

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

Using Corollary 4.3, we get,

Corollary 4.4 *Under the same assumptions as in Corollary 4.3, the source wavelet eastimation $\overline{f}$ satisfies*

$$\|\overline{f}[v] - f^*\|_{L^2([0,T])} \leq C(\sigma + \beta + r)$$

5 Hessian and Local Convexity

Differentiate the directional derivative of J_{dso} , we get

$$\begin{aligned} D^2 J_{dso}(\delta v, \delta v) = \; & \sigma^2 < D\{a[v]^2(\frac{d}{dz}D\tau[v]\delta v)\}\delta v\frac{\partial}{\partial t}F[v], (\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})F[v] > \\ & +\sigma^2 < a[v]^2(\frac{d}{dz}D\tau[v]\delta v)\frac{\partial}{\partial t}(DF[v]\delta v), (\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})F[v] > \\ & +\sigma^2 < a[v]^2(\frac{d}{dz}D\tau[v]\delta v)\frac{\partial}{\partial t}F[v]\delta v, (\frac{d}{dz}D\tau[v]\delta v)\frac{\partial}{\partial t}F[v] > \\ & +\sigma^2 < a[v]^2(\frac{d}{dz}D\tau[v]\delta v)\frac{\partial}{\partial t}F[v]\delta v, (\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})(DF[v]\delta v) > \\ & -\sigma^2 < D\{a[v]^2(\frac{d}{dz}\frac{Da[v]\delta v}{a[v]})\}\delta v F[v], (\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})F[v] > \\ & -\sigma^2 < a[v]^2(\frac{d}{dz}\frac{Da[v]\delta v}{a[v]})(DF[v]\delta v), (\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})F[v] > \\ & -\sigma^2 < a[v]^2(\frac{d}{dz}\frac{Da[v]\delta v}{a[v]})F[v], (\frac{d}{dz}D\tau[v]\delta v)\frac{\partial}{\partial t}F[v] > \\ & -\sigma^2 < a[v]^2(\frac{d}{dz}\frac{Da[v]\delta v}{a[v]})F[v], (\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})(DF[v]\delta v) > \end{aligned}$$

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 $DR[v]\delta v$ and using (4.4).

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

$$\begin{aligned}
&\langle a[v]\frac{\partial}{\partial t}R[v], a[v]\frac{\partial}{\partial t}\phi\rangle + \sigma^2\langle a[v](\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})R[v], a[v](\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})\phi\rangle \\
&= \langle e, a[v]\frac{\partial}{\partial t}\phi\rangle - \sigma^2\langle a[v](\frac{d}{dz} + \tau[v]'\frac{\partial}{\partial t})(\frac{a^*}{a}f^*(t-\tau^*), a[v](\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})\phi\rangle
\end{aligned}$$

Taking derivative with respect to v in direction δv, we get a variational equality satisfied by $DR[v]\delta v$:

$$\begin{aligned}
&\langle a[v]\frac{\partial}{\partial t}DR[v]\delta v, a[v]\frac{\partial}{\partial t}\phi\rangle + \sigma^2\langle a[v](\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})DR[v]\delta v, a[v](\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})\phi\rangle \\
&= \langle e, Da[v]\delta v\frac{\partial}{\partial t}\phi\rangle \\
&-\sigma^2\langle 2Da[v]\delta v(\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})\left[\frac{a[v^*]}{a[v]}f^*(t-\tau[v^*])\right], a[v](\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})\phi\rangle \\
&-\sigma^2\langle a[v](\frac{d}{dz}D\tau[v]\delta v)\frac{\partial}{\partial t}\left[\frac{a[v^*]}{a[v]}f^*(t-\tau[v^*])\right], a[v](\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})\phi\rangle \\
&-\sigma^2\langle a[v](\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})\left[D\frac{a[v^*]}{a[v]}\delta v f^*(t-\tau[v^*])\right], a[v](\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})\phi\rangle \\
&-\sigma^2\langle a[v](\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})\left[\frac{a[v^*]}{a[v]}f^*(t-\tau[v^*])\right], a[v](\frac{d}{dz}D\tau[v]\delta v)\frac{\partial}{\partial t}\phi\rangle \\
&-\langle 2Da[v]\delta v\frac{\partial}{\partial t}R[v], a[v]\frac{\partial}{\partial t}\phi\rangle \\
&-\sigma^2\langle 2Da[v]\delta v(\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})R[v], a[v](\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})\phi\rangle \\
&-\sigma^2\langle a[v](\frac{d}{dz}D\tau[v]\delta v)\frac{\partial}{\partial t}R[v], a[v](\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})\phi\rangle \\
&-\sigma^2\langle a[v](\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})R[v], a[v](\frac{d}{dz}D\tau[v]\delta v)\frac{\partial}{\partial t}\phi\rangle
\end{aligned}$$

Let $\phi = DR[v]\delta v$, use the same technique we have uesd in the proof of Lemma 4.2, and assume that $\sigma < 1$, we can get

$$\begin{aligned}
&\|\frac{\partial}{\partial t}DR[v]\delta v\| \le C(\sigma + r)\|S^*_{data}\|\|\delta v\| \\
&\|(\frac{\partial}{\partial z} + \tau[v]'\frac{\partial}{\partial t})DR[v]\delta v\| \le C(\sigma + r)\|S^*_{data}\|\|\delta v\|
\end{aligned}$$

which in turn implies

$$\|DR[v]\delta v\| \le C(\sigma + r)\|S^*_{data}\|\|\delta v\|$$

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

$$\begin{aligned}
&\|\frac{\partial}{\partial t}DF[v]\delta v\| \le C\|S^*_{data}\|\|\delta v\| \\
&\|(\frac{\partial}{\partial z}+\tau[v]'\frac{\partial}{\partial t})DF[v]\delta v\| \le C(\beta+\max_z|\frac{d}{dz}(\tau[v]-\tau[v^*])|+\sigma+r)\|S^*_{data}\|\|\delta v\| \\
&\|DF[v]\delta v\| \le C\|S^*_{data}\|\|\delta v\|
\end{aligned}$$

where C is a constant independent of v and δv. Hence

$$\begin{aligned}
|D^2J_{dso}(\delta v,\delta v)| &\le \sigma^2\langle a[v](\frac{d}{dz}D\tau[v]\delta v)\frac{\partial}{\partial t}F[v], a[v](\frac{d}{dz}D\tau[v]\delta v)\frac{\partial}{\partial t}F[v]\rangle \\
&\quad -\sigma^2 O(\max_z|\frac{d}{dz}(\tau[v]-\tau[v^*])|+\beta+\sigma+r)\|S^*_{data}\|\|dv\|^2 \\
&= \sigma^2\langle a[v^*](\frac{d}{dz}D\tau[v]\delta v)\frac{d}{dt}f^*(t-\tau[v^*]), a[v^*](\frac{d}{dz}D\tau[v]\delta v)\frac{d}{dt}f^*(t-\tau[v^*])\rangle \\
&\quad -\sigma^2 O(\max_z|\frac{d}{dz}(\tau[v]-\tau[v^*])|+\beta+\sigma+r)\|S^*_{data}\|\|dv\|^2 \\
&\le \sigma^2\min_z a[v^*]^2\|\frac{d}{dt}f^*\|^2\|\frac{d}{dz}D\tau[v]\delta v\|^2 \\
&\quad -\sigma^2 O(\max_z|\frac{d}{dz}(\tau[v]-\tau[v^*])|+\beta+\sigma+r)\|S^*_{data}\|\|dv\|^2
\end{aligned}$$

We will consider the problem in the conic set

$$DV = \{\delta v : \|\frac{d}{dz}D\tau[v]\delta v\| \ge \gamma\|\delta v\|_2, \forall v \in V\}$$

for some $\gamma > 0$.

Theorem 5.1 *If the noise level r is small enough, f^* is oscillatory enough, and σ is small enough, $J[v]$ is convex in DV for v in a neighborhood of v^* in V.*

6 Limilt Behavior ($\sigma \longrightarrow 0$)

The results in the previous sections suggests that the DSO objective behave nicely when the differential semblance parameter σ is small, hence it seems very interesting to see the limit behavior of the objective function as $\sigma \longrightarrow 0$. In this section we will investigate this limit. For the sake of clearness of the idea, we assmue that the the data is noise free.

Let us take a look at the DSO directional derivative first:

$$\begin{aligned}
DJ_{dso}[v]\delta v &= \sigma^2\langle a[v](\frac{d}{dz}D\tau[v]\delta v)\frac{\partial}{\partial t}F[v], a[v](\frac{\partial}{\partial z}+\tau[v]'\frac{\partial}{\partial t})F[v]\rangle \\
&\quad -\sigma^2\langle a[v](\frac{d}{dz}\frac{Da[v]\delta v}{a[v]})F[v], a[v](\frac{\partial}{\partial z}+\tau[v]'\frac{\partial}{\partial t})F[v]\rangle
\end{aligned}$$

According to Lemma 4.2, when σ is small and the noise level is low (i.e., r is small), then

$$\begin{aligned} DJ_{dso}[v]\delta v &\approx \sigma^2\langle a[v](\frac{d}{dz}D\tau[v]\delta v)\frac{\partial}{\partial t}\left(\frac{a[v^*]}{a[v]}f^*(t-\tau[v^*])\right), \\ & a[v](\frac{\partial}{\partial z}+\tau[v]'\frac{\partial}{\partial t})\left(\frac{a[v^*]}{a[v]}f^*(t-\tau[v^*])\right)\rangle \end{aligned}$$

which is very much like the derivative of the second term of the DSO objective alone, with $F[v]$ being replaced by

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

which is the solution of the least squares problem in the expanded space. This observation suggests that we consider

$$J_0[v]=\frac{1}{2}\|a[v](\frac{\partial}{\partial z}+\tau[v]'\frac{\partial}{\partial t})F_0[v]\|^2$$

where

$$F_0[v]=\frac{a[v^*]}{a[v]}f^*(t-\tau[v^*])$$

as a candidate of the limit case.

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 (4.4) implies that

$$\begin{aligned} J_\sigma[v] &= \frac{1}{2}\|a[v]\frac{\partial}{\partial t}R\|^2 \\ &+\frac{\sigma^2}{2}\|a[v](\frac{\partial}{\partial z}+\tau[v]'\frac{\partial}{\partial t})\left(\frac{a[v^*]}{a[v]}f^*(t-\tau[v^*])\right)+a[v](\frac{\partial}{\partial z}+\tau[v]'\frac{\partial}{\partial t})R\|^2 \end{aligned}$$

hence

$$|\frac{1}{\sigma^2}J_\sigma[v]-J_0[v]|\le O(\sigma)$$

Straightforward calculation shows that the directional derivative of J_0 is

$$\begin{aligned} DJ_0[v]\delta v &= \langle a[v](\frac{d}{dz}D\tau[v]\delta v)\frac{\partial}{\partial t}\left(\frac{a[v^*]}{a[v]}f^*(t-\tau[v^*])\right), \\ & a[v](\frac{\partial}{\partial z}+\tau[v]'\frac{\partial}{\partial t})\left(\frac{a[v^*]}{a[v]}f^*(t-\tau[v^*])\right)\rangle \\ & -\langle a[v](\frac{d}{dz}\frac{Da[v]\delta v}{a[v]})\left(\frac{a[v^*]}{a[v]}f^*(t-\tau[v^*])\right), \\ & a[v](\frac{\partial}{\partial z}+\tau[v]'\frac{\partial}{\partial t})\left(\frac{a[v^*]}{a[v]}f^*(t-\tau[v^*])\right)\rangle \end{aligned}$$

Hence, from Lemma 4.2, we have

$$|\frac{1}{\sigma^2}DJ_\sigma[v]\delta v-DJ_0[v]\delta v|\le O(\sigma)\|\delta v\|$$

Similarly, we can calculate the hessian of $J_0[v]$ and show that

$$|\frac{1}{\sigma^2}D^2J_\sigma[v](\delta v,\delta v) - D^2J_0[v](\delta v,\delta v)| \leq O(\sigma)\|\delta v\|^2$$

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 6.1 *$\frac{1}{\sigma^2}J_\sigma[v]$ uniformly converges to $J_0[v]$ up to at least the second derivative.*

7 Numerical Example

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. 2 is a Ricker source wavelet with peak frequency 200 hz. Fig. 3 is the receiver-reparametrized source wavelet.Fig. 4 is the seismogram using the above source wavelet and the homogeneous velocity model $v = 2000m/s$. Fig. 5 and fig. 6 are DSO and OLS objective functions respectively. For the DSO approach, we choose $\sigma = 1$. The difference is clear. The DSO objectice function is nicely convex while the OLS objective function is not. Fig. 7 is the estimate of the source wavelet using a wrong velocity $v = 2400m/s$. The receiver-dependent nature of the estimated source wavelet indicates that the velocity is not correct. Fig. 8 is the estimate of the source wavelet, which is flat, *i.e.*, independent of the receiver location, at the correct velocity $v = 2000m/s$.

Acknowlegement This work was partially supported by the National Science Foundation, the Office of Naval Research, the Texas Geophysical Parallel Computation Project, the Schlumberger Foundation, and The Rice Inversion Project. TRIP Sponsors for 1994 are: Advance Geophysical, Amoco Production Co., Conoco Inc., Cray Research Inc., Exxon Production Research Co., Interactive Network Technologies, Mobil Research and Development Corp., and Texaco Inc.

Bibliography

1. T.N.BISHOP, K.P.BUBE,R.T.CUTLER,R.T.LAINGAN,P.L.LOVE, J.R.RESNICK, R.T.SHUEY,D.A.SPINDLER,and H.W.WYLD. Tomographic determination of velocity and depth in laterally varying media. *Geophysics*, 50:903-923,1987.

2. K.P.BUBE. Uniqueness of reflector depths and characterization of the slowness null space in linearized seismic reflection tomography. Manuscript, 1992.

3. D. CAO, S. SINGH, and A. TARANTOLA. Simultaneos inversion for background velocity and impedance maps. *Geophysics*,55:458-469,1990.

4. S.CHIU and R.STEWART. Tomographic determination of three-dimensional seismic velocity structure using well logs, vertical seismic profiles, and surface seismic data. *Geophysics*, 52:1085-1098,1987.

5. F.DELPRAT-JANNAUD and PATRICK LAILLY. What information is contained in reflection tomography data? Part I. Manuscript, 1991.

6. O.GAUTHIER, A.TARANTOLA, and J. VIRIEUX. Two-dimensional nonlinear inversion of seismic waveforms. *Geophysics*,51:1387-1403,1986.

7. G.C.HERMAN. Generalization of traveltime inversion. *Geophysics*, 57:9-14,1992.

8. J.H.JUSTICE,A.A.VASSILIOU,S.SINGH,J.D.LOGEL,P.A.HANSEN, B.R.HALL, P.R.HUTT,J.J.SOLANKI. Acoustic tomography for monitoring enhanced oil recovery. *Geophysics: The Leading Edge of Exploration*,1989.

9. P. KOLB, F.COLLINO, and P. LAILLY. Prestack inversion of a 1d medium. *Proceedings of IEEE* 74, papes 498-506, 1986.

10. P. LAILLY, F. DELPRAT-JANNAUD, E. BECACHE, and E. CHOVET. Reflection tomography: how to cope with multiple arrivals? Extended Abstracts, SEG 62th annual international meeting and exposition, New Orleans, 1992. 741-744.

11. S.LANG. Real and Functional Analysis. Third Edition. Springer-Verlag,1993.

12. L. R. LINES, K. R. KELLY and J. QUEEN. Channel waves in cross-borehole data. *Geophysics*, 57:334-342,1992.

13. J. L. LIONS. Problemes Aux Limites Dans Les Equations Aux Derivees Partielles. Les Presses de l'Universite de Montreal. 2^e edition, 1967.

14. K. MOSEGARD and A. TARANTOLA. Monte Carlo analysis of geophysical inverse problems. In *Proc. 61st Annual International Meeting and Exposition.* Society of Exploration Geophysicists, 1991.

15. J.A.SCALES. Tomographic inversion via the conjugate gradient method. *Geophysics*, 52:179-185, 1987.

16. J.A.SCALES, M.L.SMITH, and T.L.FISCHER. Global optimization methods for highly nonlinear inverse problems. In G.Cohen, L.Halpern, and P.Joly, editors, *Mathematical and Numerical Aspects of Wave Propagation Phenomena*, pages 434-444. SIAM, Philadelphia, 1991.

17. M.K.SEN and P.STOFFA. Nonlinear multiparameter optimization using genetic algorithms: Inversion of plane wave seismograms.*Geophysics*, 56:1794-1810,1991.

18. M.K.SEN and P.STOFFA. Nonlinear one-dimensional seismic waveform inversion using simulated annealing. *Geophysics*,56:1624-1636, 1991.

19. H.SONG and W.W.SYMES. On crosswell traveltime tomography. *TRIP Annual Report*, 1994.

20. W.SYMES. A differential semblance algorithm for the inverse problem of reflection seismology.*Computers and Math. with Appl.*,22(4/5):147-178,1991a.

21. W.SYMES. Layered velocity inversion: a model problem from reflection seismology. *SIAM J. Math. Anal.*,22:680-716,1991b.

22. W.SYMES. Non-interactive estimation of the Marmousi velocity model by differential semblance optimization: Initial trials. In Grau and R. Versteeg, editors, *The Marmousi Experience: Proceedings of the EAEG Workshop on Practical Aspects of Inversion,* The Hague, 1991c.

23. W.SYMES. A differential semblance criterion for inversion of multioffset seismic reflection data.J. Geophys. Res., 1992a. to appear.

24. W.SYMES and J. CARAZZONE. Velocity inversion by differential semblance optimization. *Geophysics*, 56(5):654-663,1991.

25. W.SYMES and M.KERN. Inversion of reflection Seismograms by Differential semblance Analysis: Algorithm Structure and Synthetic Examples. TR92-3, Department of Computational and Applied Mathematics, Rice University,Houston, TX, 1992. To appear in *Geophysical Prospecting.*

26. W.SYMES. The plane wave detection problem.TR92-22, Department of Computational and Applied Mathematics, Rice University,Houston, TX, 1992b. To appear in *Inverse Problems.*

27. A.TARANTOLA. Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 51:1893-1903,1986.

28. A.TARANTOLA. *Inverse Problem Theory.* Elsevier, 1987.

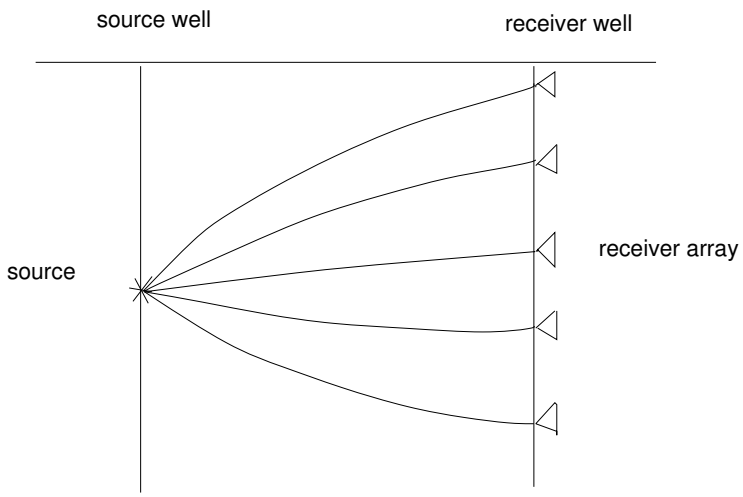

FIGURE 1: Experiment visualization

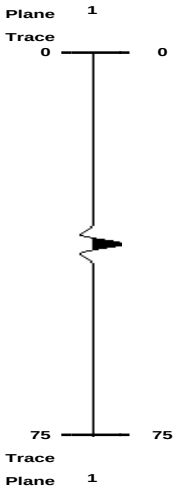

FIGURE 1: Ricker source wavelet

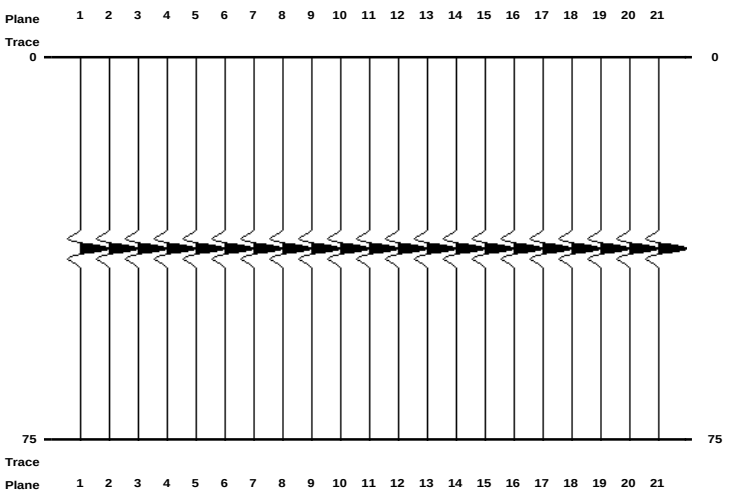

FIGURE 2: Receiver-reparametrized source wavelet

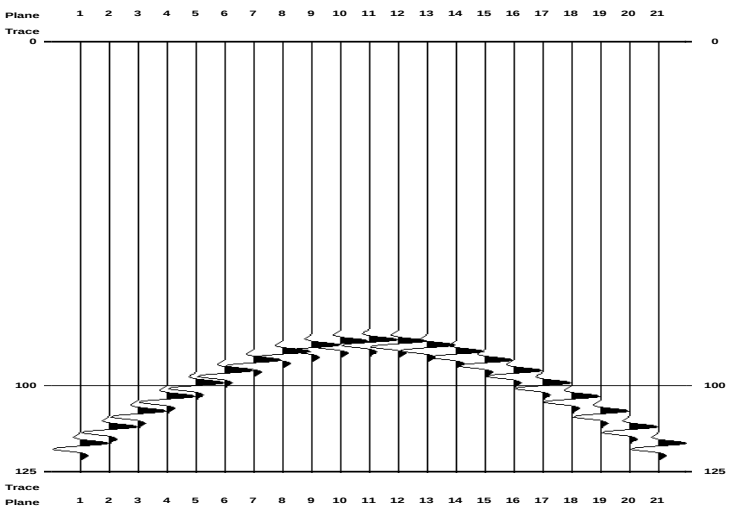

FIGURE 4: Seismogram

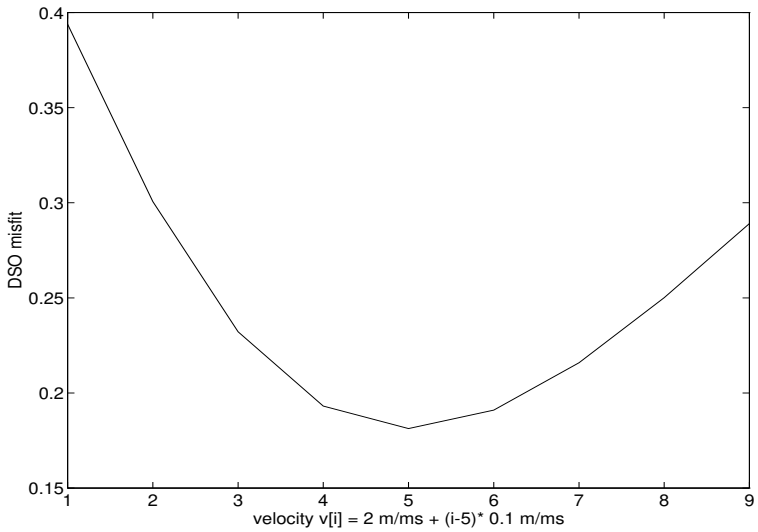

FIGURE 5: DSO objective function

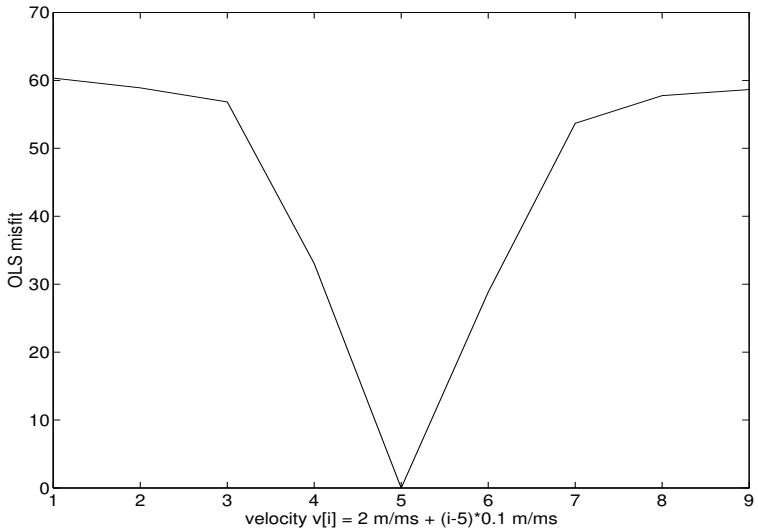

FIGURE 6: Output least squares objective function

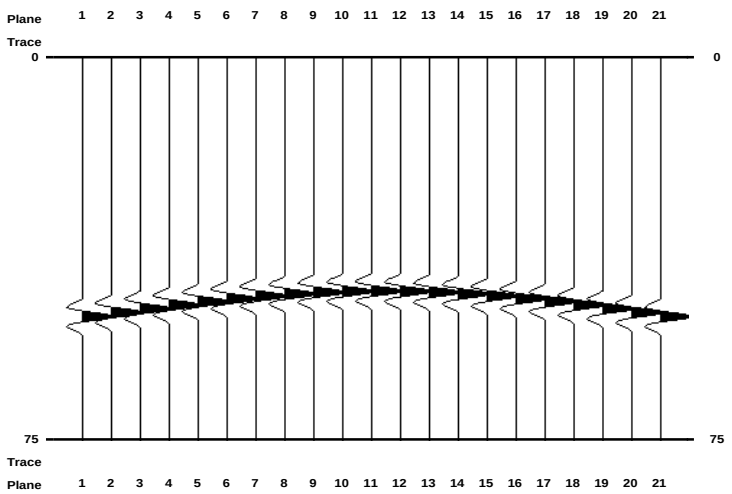

FIGURE 7: Estimation of the source wavelet with a wrong velocity model

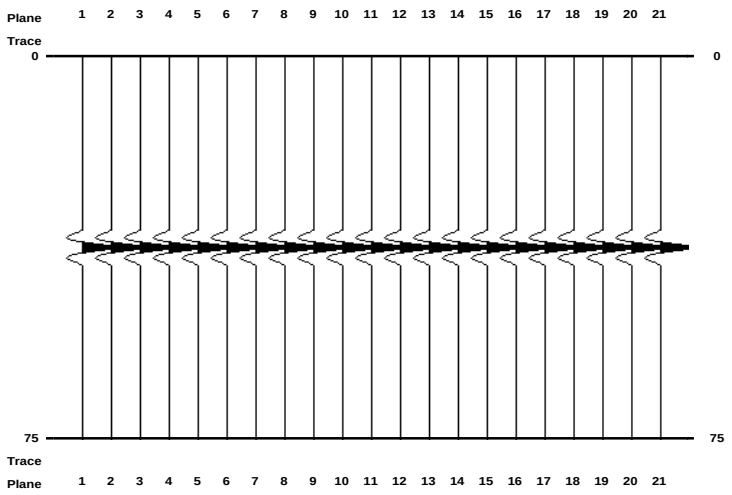

FIGURE 8: Estimation of the source wavelet with the correct velocity model