Smooth objective functionals for seismic velocity inversion

Christiaan C. Stolk and William W. Symes

Department of Computational and Applied Mathematics
Rice University
6100 Main St. – MS 134
Houston, TX 77005, USA
email: cstolk@caam.rice.edu, symes@caam.rice.edu

AMS Subject Classification: 35R30, 86A15

ABSTRACT

In seismic inverse scattering, the data are divided into subsets, each of which is used to reconstruct the medium discontinuities by linearized inversion. The reconstructions depend on an a priori unknown, smoothly varying background medium (velocity model). The semblance principle, which states that the images must agree, is the basis for the reconstruction of the background medium. Several estimators for the background medium have been proposed, based on optimization of different objective functionals. Use of local, gradient-based optimization methods requires that the functionals to be optimized are smooth. Such a smoothness requirement essentially implies that the objective functional is of *differential semblance* type. The proof involves a characterization of pseudodifferential operators as having L^2 continuous repeated commutators with order 1 pseudodifferential operators.

1. INTRODUCTION

In a seismic experiment, acoustic waves in the subsurface are generated by a source at the surface. At positions where there is a strong contrast in mechanical properties of the medium, part of the energy is reflected. The wavefield is recorded at the surface by an array of receivers. The experiment is repeated with varying source position. The data contains a wide frequency range, but low frequencies are absent, meaning that the target region is many wave lengths away from sources and receivers.

Geophysicists aim to construct an image of the subsurface from the reflections present in the data. For this purpose the data are modeled using a high-frequency linearization ("ray-Born approximation"), where the medium coefficient is written as the sum of a smooth background constituent and a singular or oscillatory perturbation (reflectivity), assumed to be small enough that linearization is accurate. The waves propagate according to geometrical optics in the background medium, and reflect at singularities of the reflectivity. Neither the background medium nor the reflectivity are known *a priori*; both are to be determined from the data. An additional complication is that the background medium is generally strongly inhomogeneous. This distinguishes the seismic inverse problem from other inverse scattering problems such as radar or ultrasonic medical imaging. While the predicted ("ray-Born") data depend linearly on the reflectivity component of the model, its dependence on the background medium is quite nonlinear.

Given a choice of background medium often a *set of images* of the singular part of the reflectivity can be obtained. The *semblance principle*, which states that the different images in this generally highly redundant set must agree, is used for determination of the background model (migration velocity analysis (Yilmaz, 1987)). To automate this procedure several optimization procedures have been proposed, in which a functional of the set of images having an extremum when the images are the same, is minimized or maximized (Al-Yahya, 1989; Symes and Carazzone, 1991; Toldi, 1989). The main examples, all quadratic in the data, are given in section 3. For an overview and more references see (Chauris and Noble, 2001).

Due to the large size of the problem local, gradient-based optimization methods are needed to keep the computational cost within reasonable limits. Therefore it is important that the objective function is smooth as a function of the background model. The inevitable presence of noise, which in principle can only be assumed to be square integrable (i.e. have finite energy), implies that the objective function really needs to be smooth as a function of background model and data jointly. In this paper we study the implication of such a smoothness requirement. Our main result is that, under reasonable additional assumptions explained below, essentially the only quadratic "semblance functional" with smooth dependence on the background medium is a quadratic form in the set of images defined by a pseudodifferential operator, with symbol depending smoothly on the background medium (Theorem 5). A similar result for the simpler plane wave detection problem (Symes, 1994) was established by (Kim and Symes, 1998). Such objective functionals have received the name "differential semblance" (Symes, 1986; Symes and Carazzone, 1991). Chauris and Noble have recently given numerical evidence supporting this claim (Chauris and Noble, 2001). In a numerical example they observed that the conventional semblance or "stacking power" function had local maxima far away from the correct maximum, while differential semblance only had a global extremum. Their results also suggest that both the density and sharpness of the local extrema of conventional semblance increase with the dominant wavenumber of the signal, i.e. the oscillation of conventional semblance is not bounded under general L^2 perturbation of the data, consistent with the results proven here.

In principle implications of C^k dependence can be derived from the arguments used to prove our statement. However, this appears to lead to many technicalities and will not be discussed here.

Before we give the precise formulation of the result in section 4, there are first two sections of an introductory nature. In section 2 we describe some well known results in seismic imaging due primarily to Beylkin. We then describe the class of semblance norms we consider, and show that the three main examples are in this class. The main result Theorem 5 and an outline of the proof is given in section 4. The proof is the subject of sections 5 and 6. In section 7 we further discuss some of the implications of our result as well as the relation with some other results in the literature. The appendix illustrates the necessity of one of the assumptions in section 4.

The main intermediate result is given in section 5. Pseudodifferential operators can be characterized as those operators on L^2 whose repeated commutators with first order pseudodifferential operators are bounded on L^2. In this section we derive a criterion for a set of pseudodifferential or differential operators to be sufficiently large, so that this characterization property is still true for repeated commutators with elements of the set.

2. SEISMIC IMAGING

We assume the medium occupies the half space $\mathbb{R}^{n-1} \times \mathbb{R}_+$ and is described by the soundspeed $c = c(x) \in C^\infty(\mathbb{R}^{n-1} \times \overline{\mathbb{R}_+})$. The data are a function of source position, receiver position and time (s, r, t) in open subset Y of $\mathbb{R}^{n-1} \times \mathbb{R}^{n-1} \times \mathbb{R}_+$, that gives the pressure field at a surface point r due to point source at s according to the acoustic equation. In practice of course $n = 2$ or 3, corresponding to sources and receivers along a line or covering a plane. We assume that the effects of the boundary, in particular surface waves and reflections, are suppressed, so that the modeling can be done using a medium that extends above the surface.

In the Born approximation the medium coefficient is written as the sum of a smooth background medium, and a perturbation $c = c_0 + \delta c$. The perturbation contains the discontinuities and is assumed to be small. The background medium is used to describe the propagation of the wavefield that is transmitted (not reflected at discontinuities). The single reflected data are modeled by the first order perturbation to the Green's function, evaluated at (s, r, t),

$$d(s, r, t) = \delta G(r, s, t), (s, r, t) \in Y.$$

It follows that the data are given by a linear map $F = F[c_0]$ acting on the singular medium perturbation (reflectivity) given by $f(x) = 2\frac{\delta c(x)}{c_0(x)^3}$. Using a geometrical optics approximation the kernel of F can be written as (Beylkin, 1985)

$$F(s, r, t; x) = \int A(x, s, r, \tau) e^{i\tau(t - T(x,s,r))} \, d\tau. \tag{1}$$

Here T is the two-way travel time function, that is the sum of two one-way travel times

$$T(x, s, r) = T_1(x, s) + T_1(x, r).$$

The latter give the travel time along a ray from x to s. We assume such rays are unique in the domain of interest (no caustics/multipathing). The ray can be determined by solving a Hamilton system.

Practical data processing involves a joint reconstruction of the background c_0 and the reflectivity f. This involves two main steps. One is the reconstruction of f given an estimate for the background medium c_0, which we discuss in this section. Second is the estimation of the background, using a set of reconstructions for f, to be discussed in the following section and the remainder of the paper.

It turns out that, since data are of dimension $2n-1$ and f is of dimension n, the data can be partitioned into n dimensional subsets, each of which can be used for reconstruction. Throughout this paper we will assume the data are partitioned into sets of constant offset $h = r - s$. The constant offset data are given by $d_h(r,t) = d(r-h,r,t)$, and F_h will be the restriction of F to constant $h = r - s$. Alternatively one can use constant source data subsets (done less in practice), for a which a similar analysis can be done.

Let $T_h(x,r) = T(x, r-h, r)$ and assume

$$\det \begin{pmatrix} \frac{\partial^2 T_h}{\partial x \partial r} & \frac{\partial T_h}{\partial x} \end{pmatrix} \neq 0. \tag{2}$$

In (Beylkin, 1985) it was observed that there is a partial (microlocal, see below) inverse

$$G_h(x;r,t) = \int \psi(r-h,r,t) B(x,h,r,\tau) \mathrm{e}^{\mathrm{i}\tau(T_h(x,r)-t)} \, \mathrm{d}\tau. \tag{3}$$

Here ψ is a C^∞ cutoff function, equal to one on a subset Y_1 of the acquisition set Y, vanishing outside a larger compact subset Y_0 of Y, and going smoothly from 1 to 0 in between.

Singularities of distributions can be localized not only in position (cf. the singular support), but also with respect to direction, i.e. in the cotangent or phase space $T^*\mathbb{R}^n\backslash 0 = \mathbb{R}^n \times (\mathbb{R}^n\backslash 0)$, by the wave front set (Hörmander, 1983). Cotangent vectors will be denoted by Greek letters, ξ, σ, ρ, τ will be the cotangent vectors with x, s, r, t. The operators F and G_h are Fourier integral operators, for which the mapping of singularities is well understood (Duistermaat, 1996). The wave front sets $\mathrm{WF}'(F_h)$, describing the mapping of singularities from $T^*\mathbb{R}^n\backslash 0$ to $T^*\mathbb{R}^n\backslash 0$ (Duistermaat, 1996), is given by

$$\Lambda_h = \{(r, T_h(x,r), -\tau\frac{\partial T_h}{\partial r}, \tau; x, \tau\frac{\partial T_h}{\partial x}) \mid (r-h, r, T_h(x,r)) \in Y, \tau \in \mathbb{R}\backslash 0\}.$$

Only singularities of f whose position in the cotangent space is in the projections of Λ_h on the right (x,ξ) variables can be present in the data and can be reconstructed. Thus the reconstruction is only partial, only the singular (high-frequency) part of f, and then only the part of singularities that is "illuminated", is reconstructed.

The set Λ_h gives rise to two projection mappings. First from a subset of points (x,r,τ) to $T^*\mathbb{R}^n_x\backslash 0$ and second from the set of (x,r,τ) to a subset of $T^*\mathbb{R}^n_{(r,t)}\backslash 0$. By assumption (2) these mappings are locally invertible. We will assume they are also globally invertible, so that Λ_h corresponds to an invertible map between subsets of $T^*\mathbb{R}^n_x\backslash 0$ and $T^*\mathbb{R}^n_{(r,t)}\backslash 0$. Denote the projection on $T^*\mathbb{R}^n_x\backslash 0$ by π_{X}. This mapping leads to an h-family of order 0 symbols $\psi_{\rm X}(h,x,\xi)$ on $T^*\mathbb{R}^n_x\backslash 0$, through pull back off the cutoff ψ in (3) by the inverse of $\pi_{\rm X}$, $\psi_{\rm X} = (\pi_{\rm X,h}^{-1})^*\psi$. The reconstruction takes place in the sense that

$$G_hFf = \psi(h,x,D_x)f, \tag{4}$$

where $\psi(h,x,D_x)$ is an h-family of pseudodifferential operators in x with principal symbol $\psi_{\rm X}$. We will often write ψ where we mean an operator $\psi(h,x,D_x)$ with principal symbol given by $\psi_{\rm X}$.

We will assume the data are normalized to be in L^2 (by convolution in time variable $(1-D_t^2)^{-(n-1)/4}$) and that F, F_h and G_h are modified accordingly to be continuous between L^2 spaces. We define the operator G to be the map from data to the set of reconstructions $G : d \mapsto f(h,x) = (G_hd)(x)$. The operator G depends on which cutoff function ψ is chosen. We use the notation G_ψ for the operator defined by (3) using cutoff function ψ.

3. SEMBLANCE FUNCTIONALS

If the background medium is correctly chosen then the images should agree, as far as possible given that the inversion of the reflectivity is only partial. Comparing images for different values of the offset $h, \bar{h}$ this means that for some cutoff given by $\tilde{\psi}(x,\xi)$ (supported where both $\psi(h,x,\xi), \psi(\bar{h},x,\xi)$ are equal to one) we have

$$\tilde{\psi}(x,D_x)(G_hd - G_{\bar{h}}d) \in C^\infty.$$

Locally in h it follows from (4) that for the correct background medium the derivatives of G_hd with respect to the different components of h

$$\frac{\partial}{\partial h_i}G_hd,$$

are singular of the same order as G_hd, as opposed to one order more singular, which is in general the case when the background medium is not correctly chosen.

This *semblance principle* suggests a method to estimate c_0, namely by maximizing the similarity of the images $f(h,x)$ in some sense. In this paper we study quadratic similarity functionals of the form

$$J[c_0] = \|B\,N[c_0]G[c_0]d\|^2 \tag{5}$$

where $B : L^2(\mathbb{R}^{2n-1}) \to L^2(\mathbb{R}^{\tilde{m}};\mathbb{R}^K)$, and $N[c_0]$ is a pseudodifferential factor depending on the medium. Note that B doesn't depend on c_0, the c_0 dependence is through $N[c_0]G[c_0]$. The latter is of the same form of G, but possibly with a different amplitude due to N. Let $A = B^*B : L^2(\mathbb{R}^{2n-1}) \to L^2(\mathbb{R}^{2n-1})$. The expression for $J[c_0]$ can be written as

$$J[c_0] = \langle d, G[c_0]^*N[c_0]^*AN[c_0]G[c_0]d\rangle. \tag{6}$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product on $L^2(\mathbb{R}^{2n-1})$.

Example 1: For the *differential* semblance functional (Symes, 1986; Symes and Carazzone, 1991) the images are compared by taking the derivatives $\frac{\partial}{\partial h_i}$. A pseudodifferential factor, denoted by $(1 + D_t^2)^{-1/2}$, is added so that the operator is continuous $L^2(\mathbb{R}^{2n-1}) \to L^2(\mathbb{R}^{2n-1}; \mathbb{R}^m)$,

$$B_{\mathrm{DS}} = \left(\tfrac{\partial}{\partial h_1}(1 + D_t^2)^{-1/2} \ldots, \tfrac{\partial}{\partial h_{n-1}}(1 + D_t^2)^{-1/2}\right)^t .$$

Note that this only works because $\mathrm{WF}(\mathcal{R}(F)) \cap \{\tau = 0\} = \emptyset$. Because of the microlocal (high-frequency) nature of the inversion for f an additional cutoff around $\xi = 0$, or $\tau = 0$ is required (in practice small frequencies are usually not present in the data). This leads to a norm $J_{\mathrm{DS}}[c_0]$ that assumes its minimum at a correct choice of c_0.

Example 2: The semblance or "stacking power" functional is given by (5) with B given by

$$(B_{\mathrm{SP}} f)(x) = \int \phi(h) f(x, h)\, \mathrm{d}h, \tag{7}$$

where $\phi(h)$ is some cutoff. Velocity analysis by maximizing this functional has been investigated in the geophysical literature (Toldi, 1989). The motivation is that only when the data for different values of h has the same phase a significant contribution appears, otherwise (7) is small due to destructive interference. From (7) it follows that the corresponding operator A_{SP} is nonlocal.

Example 3: Given the background medium c_0, there is a microlocal least squares (partial) inverse for f using all the data. It can be written in the form

$$f_{\mathrm{LS}}[c_0, d] = \int K(h, x, D_x)(Gd)(h, x)\, \mathrm{d}h \tag{8}$$

Here K depends on c_0. The least-squares approach to find c_0 is to minimize

$$\|F f_{\mathrm{LS}}[c_0, d] - d\|^2$$

This amounts to minimizing

$$-2\langle F f_{\mathrm{LS}}[c_0, d], d\rangle + \|F f_{\mathrm{LS}}[c_0, d]\|^2. \tag{9}$$

Using (8) it follows that this amounts to minimizing (5) with B_{LS} of the form

$$(B_{\mathrm{LS}} f)(x) = \int \phi(h) f(x, h)\, \mathrm{d}h,$$

where $\phi(h)$ is some cutoff, and some suitable choice $N_{\mathrm{LS}}[c_0]$ for $N[c_0]$. As in the previous example the corresponding operator A_{LS} is nonlocal.

4. SMOOTHNESS IMPLIES *DIFFERENTIAL* SEMBLANCE

This section contains the statement of our main result. Under assumptions specified below, smooth dependence of a functional of the form (5) on the background medium, uniformly for any data in $L^2(Y)$, implies that the operator $A = B^*B$ is a pseudodifferential operator with symbol in $S^0_{1,0}$.

We assume that c_0 is in a finite dimensional submanifold C of $C^\infty(\overline{\mathbb{R}^{n-1} \times \mathbb{R}_+})$, that must satisfy certain requirements. The basic smoothness requirement is

$$\langle d, G^*_\psi N^* A N G_\psi d\rangle \text{ depends in } C^\infty \text{ fashion on } (c_0, d) \in C \times L^2(Y). \tag{10}$$

This is equivalent to the requirement that the dependence is C^∞ on c_0 uniformly in d. Observe that this requirement is stronger when C is larger, so that there is no loss of generality when C is assumed to be finite dimensional.

The requirement (10) by itself is not enough. We make two modifications. In the previous section we introduced a cutoff in the Fourier domain around

$$\xi = 0, \theta \neq 0,$$

where θ is the covector corresponding to h. If χ is such a cutoff (which we will assume to be selfadjoint), then we will study $\chi A \chi$ (we will use the notation χ both for the symbol $\chi(\xi, \theta)$ and the corresponding operator). We will assume that when (ξ, θ) is in the support of $\chi(\xi, \theta)$, and $\|(\xi, \theta)\| > 1$, then

$$\xi > C\|\theta\|,$$

for some $C > 0$. The cutoff χ is chosen independent of c_0. Only $\xi \neq 0$ corresponds to reconstructed discontinuities, so no relevant signal is removed. This cutoff is in addition to the built in cutoff ψ discussed in section 2. Thus we consider the functional

$$J[c_0] = \langle d, G^*_\psi[c_0] N^*[c_0]\, \chi A \chi\, N[c_0] G_\psi[c_0] d\rangle. \tag{11}$$

In the appendix an example is given which shows that equation (10) with the cutoff χ is not sufficient to guarantee that $\psi\chi A\chi\psi$ is pseudodifferential. In addition we require smooth behavior under translation of data T_h in h direction

$$\langle d, G^*_\psi N^* T_{-h} \chi A \chi T_h N G_\psi d\rangle \text{ depends in } C^\infty \text{ fashion on } (c, d, h) \in C \times L^2(Y) \times H_0, \tag{12}$$

where H_0 is a neighborhood of 0 in $\mathbb{R}^{n-1}$. An equivalent requirement is that the dependence on $(c, h) \in C \times H_0$, is C^∞ uniformly in d. The smooth dependence on translations in h is satisfied for all the examples in section 3.

The key observation to prove our result is that (12) implies continuity of the maps

$$\mathrm{Ad}(P_1) \ldots \mathrm{Ad}(P_k)(\psi\chi A\chi\psi) : L^2 \to L^2, \tag{13}$$

for P_j in a sufficiently large set $\mathcal{S}$ of order 1 pseudodifferential operators. Here $\mathrm{Ad}(P)A = [P, A] = PA - AP$. It is known that equation (13) can be used to characterize pseudodifferential operators among all operators on L^2 according to the next lemma.

Lemma 4: *(H.O. Cordes, see (Taylor, 1981, p. 175)) Let* $A : \mathcal{D}'(\mathbb{R}^n) \to \mathcal{D}'(\mathbb{R}^n)$ *be compactly supported, and suppose that, given any* $P_j \in \mathrm{OP}\,\mathcal{S}_0$, *where* $\mathcal{S}_0 := \{x_i\xi_j\}_{i,j=1,\dots,n} \cup \{\xi_i\}_{i=1,\dots,n}$, *you have (13). Then* $A \in \mathrm{OP}\,S^0_{1,0}$.

In this paper we show that A is pseudodifferential is implied by (13) when the P_i are in a different set $\mathcal{S}$ arising from the derivative w.r.t. c_0 of G and translations T_h. This yields our main result, the following theorem:

Theorem 5: *Let* $J[c_0]$ *be of the form (11). There is a smooth finite dimensional submanifold* C *of* $C^\infty(\mathbb{R}^{n-1} \times \overline{\mathbb{R}_+})$ *containing* $c_0 = 1$, *such that if the smoothness requirement (12) for* $J[c_0]$ *is satisfied for any* ψ *of the type described above at* $c_0 = 1$ *and* $h = 0$, *then for any such* ψ

$$\psi\chi A\chi\psi \in OPS^0_{1,0}. \tag{14}$$

Thus smoothness of $J[c_0]$ restricts our choice of the operator A strongly. The examples 2 and 3 do not satisfy the smoothness requirement. Since A must be pseudodifferential it remains only to choose the symbol of A. The symbol of A should have minimum at $\theta = 0$. This naturally leads to the differential semblance norm discussed above where the symbol $A(x, h, \xi, \theta)$ is given by

$$A(x, h, \xi, \theta) = \frac{\|\theta\|^2}{\|(\xi, \theta)\|^2},$$

(or with possibly a different choice of denominator with essentially the same behavior).

The next section contains our criterion for the set $\mathcal{S}$ so that (13) implies that $\psi\chi A\chi\psi$ is pseudodifferential. Section 6 contains the proof of Theorem 5.

5. CHARACTERIZATION OF $\mathrm{OPS}^0_{1,0}$ USING REPEATED COMMUTATORS

In this section we consider operators

$$A : L^2(\mathbb{R}^{m+n}) \to L^2(\mathbb{R}^{m+n}).$$

We will use the notation $(x, y) \in \mathbb{R}^m \times \mathbb{R}^n$ for the coordinates in this section, i.e. y instead of x, x instead of h. We assume A satisfies the assumptions

$$A \text{ is compactly supported,} \tag{15}$$

$$\mathrm{WF}'(A) \cap \{(x, y, \xi, \eta; \bar{x}, \bar{y}, \bar{\xi}, \bar{\eta}) \mid \eta = 0 \text{ or } \bar{\eta} = 0\} = \emptyset. \tag{16}$$

The latter condition can be enforced by applying pseudodifferential cutoffs.

We assume there is a set $\mathcal{S}_0$ containing symbols $L(x, y, \eta)$ and symbols ξ_i, such that

$$\mathrm{Ad}(P_k(x, y, D_x, D_y)) \dots \mathrm{Ad}(P_1(x, y, D_x, D_y)A \text{ is continuous } L^2(\mathbb{R}^{m+n}) \to L^2(\mathbb{R}^{m+n}), \tag{17}$$

where $\mathrm{Ad}(P)A = [P, A]$, for all $P_1, \dots, P_k \in \mathcal{S}_0$. We show that, when $\mathcal{S}_0$ satisfies certain assumptions, then (17) is in fact valid for all P_i in a much larger set. The $L(x, y, \eta)$ are independent of ξ and therefore the corresponding operators are pseudodifferential only in y, they are smooth x-families of pseudodifferential operators. Using assumption (16) we can still obtain the desired estimates. In the text we will not always make the distinction between pseudodifferential operators and x-families of pseudodifferential operators, but it will be clear from the notation what is meant.

First we assume $\mathcal{S}_0$ contains n_V symbols $V_i(x, y, \eta)$, $i = 1, \dots n_V$, such that the vector of principal symbols $\vec{v}(x, y, \eta) := (v_1(x, y, \eta), \dots, v_L(x, y, \eta))$, is uniformly nonzero,

$$\|\vec{v}(x, y, \eta)\| \geq C\|\eta\|, \tag{18}$$

for all $(x, y, \eta) \in \mathbb{R}^{m+n} \times \mathbb{R}^n$. It follows from (17) and this assumption that

$$A \text{ is continuous } L^2(\mathbb{R}^m, H^k(\mathbb{R}^n)) \to L^2(\mathbb{R}^m, H^k(\mathbb{R}^n))$$

for any $k \in \mathbb{Z}$. By (16) we have

$$\|\xi\|^2 \leq c\|\eta\|^2$$

on the support of the Fourier transform $\mathcal{F}Au$. Therefore it follows that in fact

$$A \text{ is continuous } H^k(\mathbb{R}^{m+n}) \to H^k(\mathbb{R}^{m+n}).$$

We assume $\mathcal{S}_0$ contains in addition a set of n_W symbols of the form

$$\sum_{j=1}^{n_V} W_{ij}(x, y, \eta) V_j(x, y, \eta) = \vec{W}_i(x, y, \eta) \cdot \vec{V}(x, y, \eta),$$

$i = 1, \dots, n_W$. Here the W_{ij} are of order 0 (only the principal parts of the V_i and the W_{ij} will be of interest). In case $n_V = 1$ we make the following assumption about the vectors $\vec{w}_i(x, y, \eta)$ of principal symbols

$$\begin{gathered}\text{the map } \mathbb{R}^{m+n} \times S^{n-1} \to \mathbb{R}^{n_W} : \\ (x, y, \eta) \mapsto (w_1(x, y, \eta), \dots, w_{n_W}(x, y, \eta)) \text{ is an embedding.}\end{gathered} \tag{19}$$

In case $n_V > 1$ there is the following generalization of this condition

$$\begin{gathered}\text{for all } (x, y, \eta, \bar{x}, \bar{y}, \bar{\eta}) \in \mathbb{R}^{m+n} \times S^{n-1} \times \mathbb{R}^{m+n} \times S^{n-1}, \\ \mathrm{span}(\{\vec{w}_1(x, y, \eta) - \vec{w}_1(\bar{x}, \bar{y}, \bar{\eta}), \dots, \vec{w}_{n_W}(x, y, \eta) - \vec{w}_{n_W}(\bar{x}, \bar{y}, \bar{\eta})\}) = \mathbb{R}^{n_V},\end{gathered} \tag{20a}$$

$$\begin{gathered}\text{for all } (x, y, \eta) \in \mathbb{R}^{m+n} \times S^{n-1}, \\ \mathrm{span}(\{\tfrac{\partial}{\partial(x,y,\eta)}\vec{w}_1(x, y, \eta), \dots, \tfrac{\partial}{\partial(x,y,\eta)}\vec{w}_{n_W}(x, y, \eta)\}) = \mathbb{R}^{n_V(m+2n-1)},\end{gathered} \tag{20b}$$

$$\text{the above conditions are valid uniformly.} \tag{20c}$$

In addition we assume $\mathcal{S}_0$ contains the set

$$\{\xi_i\}_{i=1,\dots,m}$$

Define $\mathcal{S}$ by

$$\mathcal{S} = S^1_{1,0}(\mathbb{R}^{m+n}, \mathbb{R}^n) \cup \{\xi_i\}_{i=1,\ldots,m}, \tag{21}$$

(where $S^1_{1,0}(\mathbb{R}^{m+n}, \mathbb{R}^n)$ is the space of symbols of the form $L(x, y, \eta)$).

Lemma 6: *Suppose that*

$$\mathcal{S}_0 \supset \{V_1, \ldots, V_{n_V}, \vec{W}_1 \cdot \vec{V}, \ldots, \vec{W}_{n_W} \cdot \vec{V}, \xi_1, \ldots, \xi_m\}, \tag{22}$$

where the V_i and the $\vec{W}_i \cdot V$ satisfy (18) and (20). Suppose $A : L^2 \to L^2$ satisfies (15), (16). Assume that (17) holds for all $P_1, \ldots, P_k$ in $\mathcal{S}_0$. Then (17) holds for any k order 1 pseudodifferential operators or x-families of pseudodifferential operators $P_1, \ldots, P_k$ with principal symbols in $\mathcal{S}$.

Proof. Let P_i, $i = 1, \ldots, k$ be in $\mathcal{S}$. We show that in fact $\mathrm{Ad}(P_k) \ldots \mathrm{Ad}(P_1)A$ can be written as an infinite sum

$$\mathrm{Ad}(P_k) \ldots \mathrm{Ad}(P_1)A = \sum_{l=0}^{k} \sum_{i_1,\ldots,i_l} \sum_{j_1,j_2 \in \mathbb{N}} \beta_{i_1,\ldots,i_l,j_1,j_2} Q_{1,j_1} \mathrm{Ad}(L_{i_k}) \ldots \mathrm{Ad}(L_{i_1}) A Q_{2,j_2}, \tag{23}$$

where the L_i are in $\mathcal{S}_0$, the Q's are order zero pseudodifferential operators, uniformly bounded on L^2 and the sequence $|\beta_{i_1,\ldots,i_l,j_1,j_2}|$ is rapidly decreasing as $(j_1, j_2) \to \infty$.

We first show the case $k = 1$, denote $P = P_1$. Clearly the statement is true when $P = \xi_i$ for some i, so we consider the case that P has symbol $P(x, y, \eta)$. There is a vector $\vec{P}$ of order 0 pseudodifferential operators, s.t. $P = \vec{P} \cdot \vec{V}$, modulo lower order terms. We have

$$[A, P] = [A, \vec{P} \cdot \vec{V}] = \vec{P} \cdot [A, \vec{V}] + [A, \vec{P}] \cdot \vec{V}.$$

The first term in the last expression is already of the desired form. We show that the second term can be written as an infinite sum involving terms $[A, \vec{W}_i] \cdot \vec{V}$, and therefore in the form (23) (for $k = 1$). The Schwarz kernel of $[\vec{P}, A] \cdot V$ is given (modulo lower order terms of the form $Q_1 A Q_2$, with Q_1, Q_2 pseudodifferential of order 0) by

$$\begin{aligned} (2\pi)^{-2n} \int_{\mathbb{R}^{4n}} & (\vec{P}(x, y, \eta) - \vec{P}(\bar{x}, \bar{z}, \bar{\eta}))\vec{v}(\bar{x}, \bar{z}, \bar{\eta}) \\ & \times A(x, z, \bar{x}, \bar{z}) \mathrm{e}^{\mathrm{i}\langle y - z, \eta\rangle + \mathrm{i}\langle \bar{z} - \bar{y}, \bar{\eta}\rangle} \, \mathrm{d}z \, \mathrm{d}\bar{z} \, \mathrm{d}\eta \, \mathrm{d}\bar{\eta}, \end{aligned} \tag{24}$$

where $A(x, z, \bar{x}, \bar{z})$ denotes the kernel of A.

Using the Malgrange preparation theorem (Hörmander, 1983, theorem 7.5.7) we show in a few steps that there are smooth functions $q_i(x, y, \eta, \bar{x}, \bar{y}, \bar{\eta})$ such that

$$(\vec{P}(x, y, \eta) - \vec{P}(\bar{x}, \bar{y}, \bar{\eta})) = \sum_{i=1}^{n_W} q_i(x, y, \eta, \bar{x}, \bar{y}, \bar{\eta})(\vec{w}_i(x, y, \eta) - \vec{w}_i(\bar{x}, \bar{y}, \bar{\eta})). \tag{25}$$

The q_i are homogeneous of order 0 separately in η and $\bar{\eta}$. To get to the result (25) first observe that, by the Malgrange preparation theorem there are $\tilde{\gamma}_{i;jk}(x, y, \alpha, \bar{x}, \bar{y}, \bar{\alpha})$ such that

$$\vec{w}_i(x, y, \alpha) - \vec{w}_i(\bar{x}, \bar{y}, \bar{\alpha}) = \sum_{j=1}^{n_V} \sum_{k=1}^{m+2n-1} \tilde{\gamma}_{i;jk} \vec{e}_j((x, y, \alpha)_k - (\bar{x}, \bar{y}, \bar{\alpha})_k).$$

By assumption (20b) the $\tilde{\gamma}_{i;jk}(x,y,\alpha,\bar{x},\bar{y},\bar{\alpha})$ are the coefficients of a matrix (with row index i, and j,k together the column index), that has maximal rank at $(x,y,\alpha)=(\bar{x},\bar{y},\bar{\alpha})$. By assumption (20a) the same is true when $(x,y,\alpha)\neq(\bar{x},\bar{y},\bar{\alpha})$. Together with assumption (20c) it follows that there is $\gamma_{jk;i}$ such that

$$\vec{e}_j(\,(x,y,\alpha)_k-(\bar{x},\bar{y},\bar{\alpha})_k\,)=\sum_{i=1}^{n_W}\gamma_{jk;i}(\vec{w}_i(x,y,\alpha)-\vec{w}_i(\bar{x},\bar{y},\bar{\alpha})) \tag{26}$$

Again using the Malgrange preparation theorem there are $\tilde{q}_{i;jk}$ such that

$$(\vec{P}(x,y,\eta)-\vec{P}(\bar{x},\bar{y},\bar{\eta}))=\sum_{j=1}^{n_V}\sum_{k=1}^{m+2n-1}\tilde{q}_{i;jk}\vec{e}_j(\,(x,y,\alpha)_k-(\bar{x},\bar{y},\bar{\alpha})_k\,). \tag{27}$$

Using (26) in (27) yields (25).

Since the q_i are smooth they can be expanded as a sum with coefficients $\beta_{i;abcd}$ that decay rapidly when $(a,b,c,d)\to\infty$

$$q_i(x,y,\eta,\bar{x},\bar{y},\bar{\eta})=\sum_{a,b,c,d\in\mathbb{N}}\beta_{i,abcd}b_a(x,y)\omega_b(\eta)b_c(\bar{x},\bar{y})\omega_d(\bar{\eta}). \tag{28}$$

Here the b_a can for instance be chosen a Fourier basis of $L^2(B)$, where B is a sufficiently large cube (recall that A is compactly supported), and ω_a can be chosen to be the spherical harmonics on the $n-1$ sphere. For an example of the expansion of symbols for pseudodifferential operators, see e.g. (Taylor, 1991, proposition 1.1.A).

Inserting (25) and (28) into (24) we find that the operator given by (24) is given by

$$\sum_{i=1}^{n_W}\sum_{a,b,c,d\in\mathbb{N}}\beta_{i;abcd}\big(b_a(x,y)\vec{W}_i(x,y,D_y)\omega_b(D_y)A\,b_c(x,y)\vec{V}(x,y,D_y)\omega_d(D_y)$$
$$-\,b_a(x,y)\omega_b(D_y)A\,b_c(x,y)\vec{W}_i(x,y,D_y)\cdot\vec{V}(x,y,D_y)\omega_d(D_y)\big). \tag{29}$$

Expression (29) can be written as

$$\begin{aligned}\sum_{i=1}^{n_W}\sum_{a,b,c,d\in\mathbb{N}}&\beta_{i;abcd}\big(b_a(x,y)\omega_b(D_y)[\vec{W}_i(x,y,D_y),A]\vec{V}(x,y,D_y)b_c(x,y)\omega_d(D_y)\\&+b_a(x,y)[\vec{W}_i(x,y,D_y),\omega_b(D_y)]A\,b_c(x,y)\vec{V}(x,y,D_y)\omega_d(D_y)\\&-b_a(x,y)\omega_b(D_y)A\,[b_c(x,y),\vec{W}_i(x,y,D_y)\cdot\vec{V}(x,y,D_y)]\omega_d(D_y)\big).\end{aligned} \tag{30}$$

We discuss the sum of the second and third terms in this expression. It is shown in (Taylor, 1991, (3.6.35)) that, if $\tilde{P}=\tilde{P}(x,y,\xi,\eta)\in S^0$,

$$\|[b_c(x,y),\tilde{P}(x,y,D_x,D_y)]\|_{H^{-1}\to L^2}\leq C_{\tilde{P}}\|b_c\|_{\mathrm{Lip}^1}.$$

It is easy to see that this is also true for an x-family of symbols $\tilde{P}(x,y,\eta)$. This leads to a bound that is polynomial in c. Therefore, after rescaling the coefficient and the operator $[b_c(x,y),\tilde{P}(x,y,D_y)]$, the infinite sum of the third terms in (30) is of the form (23) (with $k=0$), in particular it is continuous on L^2. It is observed in (Taylor, 1991, (4.1.7)) that, if $\tilde{P}\in S^1$,

$$\|[\tilde{P}(y,D_y),\omega_c(D_y)]\|_{L^2\to L^2}\le C_{\tilde{P}}\langle c\rangle^K,$$

i.e. there is a polynomial bound in c on the commutator. Therefore the infinite sum of the second terms in (30) is also of the form (23) (with $k=0$) and is continuous on L^2. It follows after identifying the Q's, and renumbering to conform with different notation, that (29) is given by (23) for $k=1$.

The result for $k>1$ can be derived by induction, using the same argument as above. E.g. for $k=2$ the argument is applied with A replace by the sum (30). The additional commutator terms that arise are polynomially bounded in the indices a,b,c,d, and therefore lead to more terms in (23) with finite sum. □

Remarks The addition of the ξ_i in the set of symbols (22) is done because the lemma is used in this setting below. However, they could be omitted from (22) and (21). The lemma is somewhat more general than needed, because below only the case $n_V=1$ is used. However, the case $n_V>1$ is needed to consider an extension to differential operators. Such an extension is not hard to make. The key point is that the Malgrange preparation theorem can be used to construct the elements of $\mathcal{S}$. In the case of differential operators let the V_i be linearly independent vector fields that at each point (x,y) span $\mathbb{R}^n$, and the $W_{ij}=W_{ij}(x,y)$ are expansion coefficients $W_i=\vec{W}_i(x,y)\cdot\vec{V}$. Then $\mathcal{S}$ is modified to the set vector fields $\sum_i u_i(x,y)\frac{\partial}{\partial y_i}$ (together with $\frac{\partial}{\partial x_i}$) and (20) is modified to be conditions for the Malgrange preparation theorem in this case, as follows

for all $(x,y,\bar{x},\bar{y})\in\mathbb{R}^{m+n}\times\mathbb{R}^{m+n}$,
$\operatorname{span}(\{\vec{w}_1(x,y)-\vec{w}_1(\bar{x},\bar{y}),\ldots,\vec{w}_{n_W}(x,y)-\vec{w}_{n_W}(\bar{x},\bar{y})\})=\mathbb{R}^{n_V}$,
for all $(x,y)\in\mathbb{R}^{m+n}$, $\operatorname{span}(\{\frac{\partial}{\partial(x,y)}\vec{w}_1(x,y),\ldots,\frac{\partial}{\partial(x,y)}\vec{w}_{n_W}(x,y)\})=\mathbb{R}^{n_V(m+n)}$,

the above conditions are valid uniformly.

6. PROOF OF THEOREM 5

Proof. We show that (12) implies continuity of the map

$$\mathrm{Ad}(P_1)\ldots\mathrm{Ad}(P_k)(\psi\chi A\chi\psi):L^2(\mathbb{R}^{2n-1})\to L^2(\mathbb{R}^{2n-1}),\tag{31}$$

where the P_i are pseudodifferential operators of order 1 with any principal symbols θ_i or $P_i(x,h,\xi)$ in the set $\mathcal{S}$ of (21). We show this by induction on k, using up to k-th order derivatives in (12). Therefore assume (31) is proved up to order $k-1$.

First we show that (31) is valid when the P_i are in a smaller, finite, set, obtained by taking the derivative of (11). Define the operator H_ψ by

$$H_\psi=T_aNG_\psi$$

The operator H_ψ is again a Fourier integral operator similar to the operator G. The pseudodifferential factor N leads to a modified amplitude, and the translation must be taken into account. Its kernel can be written in the form

$$H_\psi(x,\bar{h};r,t,h) = \delta(\bar{h}-a-h)\int \tilde{A}(x,h,r,t)\psi(r-h,r,t)\mathrm{e}^{\mathrm{i}\tau\phi(x,h,r,t)}\,\mathrm{d}\tau$$

where $\phi = T(x,r-h,r;c_0)-t$ (here we indicated explicitly that T depends on the background medium c_0).

Consider a finite dimensional space of background mediums C, and let γ be local coordinates on C. The travel time T, the amplitude $\tilde{A}$ and hence the operator H_ψ now become a function of γ. Differentiating H_ψ with respect to γ, we find

$$\frac{\partial H_\psi}{\partial \gamma_i}(x,\bar{h};r,t,h) = \delta(\bar{h}-a-h)\int \big(\mathrm{i}\tau\frac{\partial\phi}{\partial\gamma}\tilde{A}\psi + \frac{\partial}{\partial\gamma}(\psi a)\big)\mathrm{e}^{\mathrm{i}\phi}\,\mathrm{d}\tau. \tag{32}$$

The first term of the sum is of order 1, due to the factor τ from differentiating the exponent. Up to principal level the multiplication by $\mathrm{i}\tau\frac{\partial\phi}{\partial\gamma}$ can be seen as the application of a pseudodifferential operator L_i to H_ψ. The remainder $\frac{\partial H_\psi}{\partial\gamma_i} - L_i H_\psi$ in general involves derivatives of ψ, and can therefore not necessarily be written as a pseudodifferential operator acting on H_ψ. However, let $\tilde{\psi}$ be a different cutoff function that satisfies $\tilde{\psi} > 0$ on $\operatorname{supp}\psi$. The lower order terms can be written as a pseudodifferential operator acting on $H_{\tilde{\psi}}$. Hence the derivative $\frac{\partial H_\psi}{\partial\gamma_i}$ can be written as

$$\frac{\partial H_\psi[\gamma]}{\partial\gamma_i} = L_i H_\psi[c] + R_i(\psi,\tilde{\psi})H_{\tilde{\psi}}$$

where $L_i \in \mathrm{OPS}^1_{1,0}$ has principal symbol

$$L_i(x,\tau\frac{\partial T_h}{\partial x}) = \mathrm{i}\tau\frac{\partial\phi}{\partial\gamma_i}(x,h,r,t) \tag{33}$$

(defined on a neighborhood of $\operatorname{supp}\tilde{\psi}$ in the cotangent space $T^*_x\mathbb{R}^n$). We will define the L_i (first order ψDO) to be *anti-selfadjoint* (meaning $L_i^* = -L_i$; note that the principal symbol is imaginary, so that this only involves the modification of lower order terms) and $R_i(\psi,\tilde{\psi})$ (a zero order ψDO) to be selfadjoint. Clearly for the adjoint operator H_ψ^* we have

$$\frac{\partial}{\partial\gamma_i}H_\psi^* = -H_\psi^* L_i\psi + H_{\tilde{\psi}}^* R_i(\psi,\tilde{\psi}).$$

The derivative with respect to the translation parameter a is also of this form. We let the L_i,R_i, $i = \dim C + 1,\ldots,\dim C + n - 1$ correspond to translations and we have $R = 0$ and $L_{i+\dim C} = \frac{\partial}{\partial a_i}$.

Using this we first show that

$$\frac{\partial^k}{\partial\gamma_{i_1}\ldots\partial\gamma_{i_k}}H_\psi^*\chi A\chi H_\psi = H_\psi^*\,\mathrm{Ad}(L_{i_1})\ldots\mathrm{Ad}(L_{i_k})\chi A\chi H_\psi + \text{sum of lower order terms,} \tag{34}$$

where the lower order terms are of the form

$$H_{\tilde{\psi}}^* Q_1^* \operatorname{Ad}(P_l) \ldots \operatorname{Ad}(P_1) \chi A \chi Q_2 H_{\tilde{\psi}} + \text{adjoint} \tag{35}$$

where $l < k$, the P_i are anti selfadjoint first order, and the Q_i are of order 0, and $\tilde{\psi}$ is some cutoff that satisfies $\tilde{\psi} \geq 0$ on $\operatorname{supp} \psi$. Equation (34) is proven by induction. Consider the derivative of (35) with respect to γ_i

$$\begin{aligned} & H_{\tilde{\psi}}^* Q_1^* \operatorname{Ad}(L_i) \operatorname{Ad}(P_l) \ldots \operatorname{Ad}(P_1) \chi A \chi Q_2 H_{\tilde{\psi}} \\ & \quad + H_{\tilde{\psi}}^* Q_3^* \operatorname{Ad}(P_l) \ldots \operatorname{Ad}(P_1) \chi A \chi Q_2 H_{\tilde{\psi}} + H_{\tilde{\psi}}^* Q_1^* \operatorname{Ad}(P_l) \ldots \operatorname{Ad}(P_1) \chi A \chi Q_4 H_{\hat{\psi}} \\ & \quad\quad + 2 \operatorname{Re} \sum_{j=1}^{l} H_{\tilde{\psi}}^* Q_1^* \operatorname{Ad}(P_l) \ldots \operatorname{Ad}(P_{j+1}) \operatorname{Ad}(\frac{\partial P_j}{\partial \gamma_i}) \operatorname{Ad}(P_{j-1}) \ldots \operatorname{Ad}(P_1) \chi A \chi Q_2 H_{\tilde{\psi}}. \end{aligned}$$

It follows by induction that equation (34) is satisfied. This equation is not quite in the right form, equation (31) involves the operator $\psi \chi A \chi \psi$, and not $\chi A \chi$ by itself. The operator H_ψ can be written

$$H_\psi = \psi(h, x, D) H_{\tilde{\psi}} + \text{lower order terms.}$$

Commuting the cutoffs ψ for H_ψ and H_ψ^* we find that (34) is equal to an expression of the form

$$\begin{aligned} \frac{\partial^k}{\partial \gamma_{i_1} \ldots \partial \gamma_{i_k}} H_\psi^* \chi A \chi H_\psi = {} & H_{\tilde{\psi}}^* \operatorname{Ad}(L_{i_1}) \ldots \operatorname{Ad}(L_{i_k}) \psi \chi A \chi \psi H_{\tilde{\psi}} \\ & + \text{sum of lower order terms as in (35).} \end{aligned} \tag{36}$$

By the induction hypothesis the lower order terms in (36) are continuous $L^2 \to L^2$. By the smoothness assumption the whole of (36) is continuous $L^2 \to L^2$. It follows that

$$\operatorname{Ad}(L_{i_1}) \ldots \operatorname{Ad}(L_{i_k})(\psi \chi A \chi \psi) \text{ is continuous } L^2 \to L^2. \tag{37}$$

To prove (31) from (37) it remains to show that the L_i form a large enough set, so that Lemma 6 applies. We must show that, for some choice of the manifold of background velocities C, the set of symbols L_i contains a set of the form (22), where the V_i and the W_{ij} are as described. We define the inverse velocity, or medium slowness $\nu = c_0^{-1}$. From (33) it follows that we must consider the medium perturbation $\delta T(x, r, s)$ of the travel time $T(x, r, s)$, when a smooth medium perturbation $\delta\nu$ is added to ν. We consider perturbations around $\nu = 1$. In this medium all the rays are straight lines and the travel time is given by the Euclidean length. One can show that in this medium the travel time perturbation for a single ray segment given by $\mathbf{x}(t)$, $a < t < b$ is given by the line integral (Nolet, 1987)

$$\delta T = \int_a^b \delta\nu(\mathbf{x}(t)) \, \mathrm{d}t. \tag{38}$$

It follows that a medium perturbation $\delta_0\nu := 1$ leads to $\delta_0 T = T$. It follows that we may choose $V = L_1$, given by $\delta_0 T$ (that is $n_V = 1$, and V is a scalar operator), and that then (18) is satisfied. In the next lemma we show that there are a set of medium perturbations $\delta_i\nu$, $i = 1, \dots, k$ so that the corresponding variations in travel time $\delta_i T$ lead to operators of the form $W_i V$, with the W_i satisyfing (19).

Lemma 7: *Let $\nu = 1$. Let $\alpha, \beta > 0$. There are $\delta_i\nu \in C^\infty(\overline{\mathbb{R}^n_+})$, $i = 1, \dots, 4n-3$, such that the map*

$$\{x \in \mathbb{R}^n \mid x_n > \alpha\} \times \{(r,s) \in \mathbb{R}^{n-1} \times \mathbb{R}^{n-1} \mid (r-s) \in]\beta, \infty[^{n-1}\} \to \mathbb{R}^{4n-3} : \\ (x,r,s) \mapsto \left(\frac{\delta_1 T}{T}, \dots, \frac{\delta_{4n-3} T}{T} \right) \quad (39)$$

is an embedding.

Proof. Let $\phi(x_n)$ be positive with support in $[0, \alpha_0]$, $0 < \alpha_0 < \alpha$. Let

$$\begin{aligned}
\delta_1 \nu &= x_n, & \\
\delta_{1+i} \nu &= x_i, & i &= 1, \dots, n-1, \\
\delta_{n+i} \nu &= x_i \phi(x_n), & i &= 1, \dots, n-1, \\
\delta_{2n-1+i} \nu &= x_i^2, & i &= 1, \dots, n-1, \\
\delta_{3n-2+i} \nu &= x_i^3, & i &= 1, \dots, n-1.
\end{aligned}$$

The corresponding travel time perturbations follow from (38). We show that the map $(x,r,s) \mapsto \left(\frac{\delta_1 T}{T}, \dots, \frac{\delta_{4n-3} T}{T} \right)$ is injective, i.e. that $\left(\frac{\delta_1 T}{T}, \dots, \frac{\delta_{4n-3} T}{T} \right)$ determines (x,s,r). Denote $x = (x', x_n)$, $x' = (x_1, \dots, x_{n-1})$. Let u be a parameter that increases linearly from 0 to 1 along source and receiver segment, i.e. for the source ray $\mathbf{x}_\mathrm{s} = (s,0) + u(s - x', x_n)$, $\mathbf{x}_\mathrm{r}$ is defined similarly. Let $T_\mathrm{s} = \sqrt{(s-x')^2 + x_n^2}$, and similarly for T_r, then $T = T_\mathrm{s} + T_\mathrm{r}$ and

$$\delta T = \int_0^1 \left[T_\mathrm{s} \delta\nu(\mathbf{x}_\mathrm{s}(u)) + T_\mathrm{r} \delta\nu(\mathbf{x}_r(u)) \right] \mathrm{d}u. \quad (40)$$

It follows that

$$\frac{\delta_1 T}{T} = \frac{x_n}{2}, \quad (41)$$

whence $\frac{\delta_1 T}{T}$ determines x_n. It further follows that

$$\frac{\delta_{1+i} T}{T} = \frac{T_\mathrm{s}}{2T} s_i + \frac{T_\mathrm{r}}{2T} r_i + \tfrac{1}{2} x_i, \qquad i = 1, \dots, n-1.$$

The line through x and the center of mass of the figure consisting of source and receiver ray is given by

$$\mathbf{x}_\mathrm{m}(u) = \frac{T_\mathrm{s}}{T} \mathbf{x}_\mathrm{s}(u) + \frac{T_\mathrm{r}}{T} \mathbf{x}_\mathrm{r}(u) = (1-u) \left(\frac{T_\mathrm{s}}{T}(s,0) + \frac{T_\mathrm{r}}{T}(r,0) \right) + ux.$$

The value $\mathbf{x}_{\mathrm{m}}(1/2)$ is determined by the $\frac{\delta_i T}{T}$, $i = 1, \ldots, n$. We have

$$\frac{\delta_{n+i}T}{T} = \left(\frac{T_{\mathrm{s}}}{2T}s_i + \frac{T_{\mathrm{r}}}{2T}r_i\right)\int_0^1 \phi(ux_n)(1-u)\,\mathrm{d}u + x_i\int_0^1 \phi(ux_n)u\,\mathrm{d}u, \quad i = 1, \ldots, n-1. \tag{42}$$

Because x_n is determined from (41), the integrals $\int_0^1 \phi(ux_n)\,\mathrm{d}u$ and $\int_0^1 \phi(ux_n)u\,\mathrm{d}u$ are known. It follows that (42) determines another point on the line $u \mapsto \mathbf{x}_{\mathrm{m}}(u)$, and therefore fixes x and $m_1 \in \mathbb{R}^{n-1}$ defined by

$$m_{1,i} := \mathbf{x}_{\mathrm{m},i}(0) = \frac{T_{\mathrm{s}}}{T}s_i + \frac{T_{\mathrm{r}}}{T}r_i. \tag{43}$$

We have

$$\begin{aligned}\frac{\delta_{2n-1+i}T}{T} &= \frac{T_{\mathrm{s}}}{T}\int ((1-u)s_i + ux_i)^2\,\mathrm{d}u \\ &\quad + \frac{T_{\mathrm{r}}}{T}\int ((1-u)r_i + ux_i)^2\,\mathrm{d}u, \qquad i = 1, \ldots, n-1.\end{aligned} \tag{44}$$

Because x is known the values

$$m_{2,i} := \frac{T_{\mathrm{s}}}{T}s_i^2 + \frac{T_{\mathrm{r}}}{T}r_i^2, \qquad i = 1, \ldots, n-1. \tag{45}$$

can be found from (44). Similarly from $\frac{\delta_{3n-2+i}}{T}$, $i = 1, \ldots, n-1$, the values

$$m_{3,i} := \frac{T_{\mathrm{s}}}{T}s_i^3 + \frac{T_{\mathrm{r}}}{T}r_i^3, \qquad i = 1, \ldots, n-1. \tag{46}$$

can be found. For each i, equations (43), (45) and (46), form a set three equations with three unknowns r_i, s_i and $\frac{T_{\mathrm{s}}}{T} = 1 - \frac{T_{\mathrm{r}}}{T}$. By applying a change in coordinates we may assume that $m_{1,i} = 0, m_{2,i} = 1$. It follows (since by assumption $r_i > s_i$) that for each i there is a unique solution for r_i, s_i and $\frac{T_{\mathrm{s}}}{T}$ given by

$$\begin{aligned} r_i &= \tfrac{1}{2}\left(m_{3,i} + \sqrt{4 + m_{3,i}^2}\right), \\ s_i &= \tfrac{1}{2}\left(m_{3,i} - \sqrt{4 + m_{3,i}^2}\right), \\ \frac{T_s}{T} &= \tfrac{1}{2} - \frac{m_{3,i}}{2\sqrt{4 + m_{3,i}^2}}. \end{aligned}$$

By tracing the steps in the determination of (x, r, s) from the $\frac{\delta_i T}{T}$, $i = 1, \ldots, 3n-2$ it follows that in fact the map (39) is immersive and proper. This shows that the map (39) is an embedding. □

The submanifold C is chosen of dimension $4n-2$ such that the $\delta_i\nu$, $i = 0, \ldots, 4n-3$ of the lemma are tangent to the submanifold at $\nu = 1$. The operators L_i so far are only defined microlocally on the set (x, ξ) that is related to data through the theory of section 2. They can be modified microlocally outside this set, since $\psi\chi A\chi\psi$, is smoothing there, and we also know that for any operator with symbol supported outside this set the commutator is smoothing, hence continuous on L^2. In particular we may assume $V = L_1$ is extended to be nonzero outside this set, and that there are additional W_i, supported outside this set, so that (19) is valid for all (x, ξ), with x in the support of $\psi\chi A\chi\psi$.

With the set of L_i modified in this way we can apply Lemma 6. This completes the induction and shows that (31) holds. In particular it holds for symbols P_j in (with the notation θ for the covector with h)

$$\mathcal{S}_1 = \{\xi_j\}_{j=1,\dots,n} \cup \{h_i\xi_j\}_{i=1,\dots,n-1,j=1,\dots,n} \cup \{x_i\xi_j\}_{i,j=1,\dots,n} \cup \{\theta_j\}_{j=1,\dots,n-1} \tag{47}$$

We can now complete the proof of the theorem by applying arguments similar to the ones used in the proof of the Cordes lemma in (Taylor, 1981, p. 175). Suppose $B = \mathrm{Ad}(P_k)\dots\mathrm{Ad}(P_1)\psi\chi A\psi\chi$. We show by induction that all such B are continuous $H^k(\mathbb{R}^{2n-1}) \to H^k(\mathbb{R}^{2n-1})$. Assume the statement is proven for $k-1 \geq 0$, let $u \in H^k(\mathbb{R}^{2n-1})$. Then we have to show that $D_i B u \in H^{k-1}$ and $Bu \in H^{k-1}$. The latter is automatic, the former can be shown by commuting D_i to the right.

Consider

$$B_{\alpha\beta} = Ad(h,x)^\alpha Ad(D_h, D_x)^\beta(\psi\chi A\chi\psi).$$

From (31) and (47) it follows that $B_{\alpha\beta}D_x^\gamma$ is continuous $L^2(\mathbb{R}^{2n-1}) \to L^2(\mathbb{R}^{2n-1})$, if $|\gamma| = |\alpha|$. It follows that

$$B_{\alpha\beta} \text{ is continuous } L^2(\mathbb{R}^{2n-1}) \to L^2(\mathbb{R}^{n-1}, H^{|\alpha|}(\mathbb{R}^n)).$$

Due to the cutoff χ there is C such that the Fourier transform $\mathcal{F}(B_{\alpha\beta}u)$ decays rapidly towards infinity for (ξ, θ) in the set given by

$$\|\theta\|^2 > C\|\xi\|^2.$$

It follows that in fact $B_{\alpha\beta}$ is continuous $L^2(\mathbb{R}^{2n-1}) \to H^{|\alpha|}(\mathbb{R}^{2n-1})$. Using a similar argument as above it also follows that

$$B_{\alpha\beta} \text{ is continuous } H^k(\mathbb{R}^{2n-1}) \to H^{k+|\alpha|}(\mathbb{R}^{2n-1}). \tag{48}$$

For the remainder of this proof introduce the notation $y = (x, h)$, η the covector. Define $a(y, \eta) = \mathrm{e}^{-\mathrm{i}y\cdot\eta}A(\mathrm{e}^{\mathrm{i}y\cdot\eta})$. Then for $u \in \mathcal{E}'$, $Au = (2\pi)^{-2n+1}\int a(y,\eta)\mathrm{e}^{\mathrm{i}y\cdot\eta}\hat{u}(\eta)\,\mathrm{d}\eta$. We need $a(y, \eta) \in S^0_{1,0}$. Now $D^\alpha_\eta D^\beta_x a(y, \eta) = \mathrm{e}^{-\mathrm{i}y\cdot\eta}B_{\alpha\beta}(\mathrm{e}^{\mathrm{i}y\cdot\eta})$, and (48) implies

$$\|B_{\alpha\beta}(\mathrm{e}^{\mathrm{i}y\cdot\eta})\|_{H^k} \leq c\|\mathrm{e}^{\mathrm{i}y\cdot\eta}\phi(x)\|_{H^{k-|\alpha|}}, \tag{49}$$

where $\phi \in C_0^\infty(\mathbb{R}^{2n-1})$ is a cutoff. For $|\alpha| = 0, |\beta| = \lfloor n/2 \rfloor + 1$, $k = 0$, with Sobolev's embedding theorem, (49) yields

$$|a(x, \xi)| \leq C_0. \tag{50}$$

To estimate $D_\eta^\alpha D_y^\beta a(y,\eta)$, which is the symbol $b_{\alpha\beta}(y,\eta)$ of $B_{\alpha\beta}$ note that $b_{\alpha\beta}(y,\eta)(1+|\eta|^2)^{|\alpha|/2}$ is the symbol of $E_{\alpha\beta} = B_{\alpha\beta}(1+\sum D_{x_i}^2 + \sum D_{h_i}^2)^{|\alpha|/2}$. It follows that $E_{\alpha\beta}$ satisfies the condition (31) with P_i in (47). So the arguments used to show (50) apply. The operator $E_{\alpha\beta}$ has bounded symbol, or equivalently

$$D_\eta^\alpha D_y^\beta a(y,\eta) \le C(1+|\eta|)^{-|\alpha|}.$$

This proves the theorem. □

7. DISCUSSION

We conclude with a brief discussion of the practical implications of our result and the relation with other results in the literature.

Seismic data are the primary source of information in present day exploration for oil and gas. In the processing of these data velocity estimation is an important bottleneck. Current practice is largely manual and costly. This is a strong motivation to study automated methods that treat the problem as an inverse problem. Here we study methods that in principle use *all* the data, as opposed to methods that use picking, i.e. extracting arrival time data of a relatively small number of strong arrivals (which is often difficult or requires some human intervention.)

As mentioned in the introduction, the low frequencies are generally absent in seismic data while at the same time very little is known a priori about the large scale structure of the medium. This is a key reason that conventional least squares inversion is generally unpractical for seismic problems. For the non-linear problem (full wave form inversion) it has been shown that, in the absence of low frequencies, gradient based optimization schemes require good initial estimates that are generally unavailable (Gauthier et al., 1986; Kolb et al., 1986). Farther away from the correct model the least squares functional exhibits oscillatory dependence on the velocity model with many local minima. Global search methods (Sen and Stoffa, 1991; Scales et al., 1991) are still far too expensive for multidimensional seismic data processing. For the linearized problem discussed in this paper, similar behavior was observed for the semblance and least squares functionals introduced in section 3 (Chauris and Noble, 2001; Landa et al., 1989). As the frequency content increases the classical semblance functional becomes increasingly non-smooth. On the contrary the differential semblance functional introduced in (Symes, 1986; Symes and Carazzone, 1991) (see Section 3) is smooth for arbitrary high frequency content. Smoothness of the differential semblance function and good convergence were also observed in numerical experiments (although other obstacles can still be encountered.)

The importance of smoothness of the objective functional for the behavior of optimization algorithms naturally leads to the question studied in this paper, i.e. whether there exist other smooth quadratic objective functionals in addition to differential semblance. We restricted ourselves to a large class of bilinear functionals, of which the main examples given in section 3 are members. Our main result Theorem 5 shows that from this class only the pseudodifferential bilinear forms are smooth, defined as being C^∞. It follows that the differential semblance functional is essentially the only asymptotically smooth optimization functional with correct minima – one could in fact call it the *differentiable semblance* functional.

Appendix A. APPENDIX

In this appendix we show by an example that (10) does not imply that $\psi A\psi \in \mathrm{OPS}^0_{1,0}$. Let $A = H(h_1 - a)\phi_1(h)\phi_2(x)$, where H is the Heaviside function, h_1 is the first component of h, $a > 0$, ϕ_1 is smooth cutoff with support that contains a neighborhood of $h = (a, 0, \ldots, 0)$, and ϕ_2 is a second smooth cutoff. For this operator A (10) is satisfied. We show that $\psi A\psi$ is not pseudodifferential. Given the small support of A, we may assume that $\psi = \psi(D_h, D_x)$. Let $g \in C_0^\infty(H)$ be supported around $(a, 0, \ldots, 0)$. Let $f \in L^2(X)$ and assume that $\mathrm{WF}(f) \cap \mathrm{supp}(\phi_2) \times \mathbb{R}^n \backslash 0$ is nonempty. Then $\mathrm{WF}(fg) = \{(0, \xi, h, x) \mid h \in \mathrm{supp}(g), (x, \xi) \in \mathrm{WF}(f)\}$. Let

$$k(h, x) = \left[\tfrac{\partial}{\partial h_1}, \psi A\psi\right] f(x)g(h). \tag{A-1}$$

If u is a distribution on $\mathbb{R}^{2n-1}$ we denote by $\Sigma(u)$ the conic subset of of $\mathbb{R}^{2n-1}$ such that the Fourier transform $\hat{u}$ does not decay in those directions, to be precise a direction (ξ_0, θ_0) is not in $\Sigma(u)$ if $\hat{u}(\lambda\xi, \lambda\theta)$ decays faster than λ^{-N} for any N and all (ξ, θ) in a neighborhood of (ξ_0, θ_0) (Hörmander, 1983, chapter 8). The set $\Sigma(u)$ is the projection of the wave front set on the cotangent variable. We show that

$$\Sigma(k) \nsubseteq \Sigma(fg). \tag{A-2}$$

The singular contribution in equation (A-1) comes from the derivative of the Heaviside function. This term is given by

$$\hat{k}(\theta, \xi) = \psi(\theta, \xi) \int \psi(\theta', \xi)(\hat{\phi}_1 * \hat{g})(\theta')(\hat{\phi}_2 * \hat{f})(\xi)\, \mathrm{d}\theta'.$$

It follows that $\Sigma(k)$ contains elements with nonzero θ unlike $\Sigma(fg)$. Therefore A is not pseudodifferential.

Acknowledgements

We wish to thank Michael Taylor for pointing out the result of Cordes described in his 1981 book, which was the starting point for the work reported here. This work was supported in part by The Rice Inversion Project. Part of this work was done while the first author was at The Mathematical Sciences Research Institute in Berkeley, California , supported by the NSF under grant DMS-9810361.

REFERENCES

Al-Yahya, K. M. (1989). Velocity analysis by iterative profile migration. *Geophysics*, 54(06):718–729.

Beylkin, G. (1985). Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform. *J. Math. Phys.*, 26(1):99–108.

Chauris, H. and Noble, M. (2001). Two-dimensional velocity macro model estimation from seismic reflection data by local differential semblance optimization: applications to synthetic and real data sets. *Geophys. J. Int.*, pages 14–26.

Duistermaat, J. J. (1996). *Fourier Integral Operators.* Birkhäuser, Boston.

Gauthier, O., Virieux, J., and Tarantola, A. (1986). Two-dimensional nonlinear inversion of seismic waveforms: Numerical results. *Geophysics*, 51(7):1387–1403.

Hörmander, L. (1983). *The Analysis of Linear Partial Differential Operators*, volume 1. Springer-Verlag, Berlin.

Kim, S. and Symes, W. W. (1998). Smooth detectors of linear phase. *Inverse Problems*, 14(1):101–112.

Kolb, P., Collino, F., and Lailly, P. (1986). Prestack inversion of a 1d medium. *Proceedings of the IEEE*, 74(3):498–508.

Landa, E., Beydoun, W. B., and Tarantola, A. (1989). Reference velocity model estimation from prestack waveforms; coherency optimization by simulated annealing. *Geophysics*, 54(8):984–990.

Nolet, G. (1987). *Seismic Tomography.* D. Reidel & Co., Dordrecht, Holland.

Scales, J. A., Smith, M. L., and Fischer, T. L. (1991). Global optimization methods for highly nonlinear inverse problems. In Cohen, G., Halpern, L., and Joly, P., editors, *Mathematical and Numerical Aspects of Wave Propagation*, pages 434–444. SIAM, Philadelphia.

Sen, M. K. and Stoffa, P. L. (1991). Nonlinear one-dimensional seismic waveform inversion using simulated annealing. *Geophysics*, 56(10):1624–1638.

Symes, W. W. (1986). Stability and instability results for inverse problems in several-dimensional wave propagation. In Glowinski, R. and Lions, J., editors, *Proc. 7th International Conference on Computing Methods in Applied Science and Engineering.* North-Holland.

Symes, W. W. (1994). The plane-wave detection problem. *Inverse Problems*, 10(6):1361–1391.

Symes, W. W. and Carazzone, J. (1991). Velocity inversion by differential semblance optimization. *Geophysics*, 56(5):654–663.

Taylor, M. E. (1981). *Pseudodifferential Operators.* Princeton University Press, Princeton, New Jersey.

Taylor, M. E. (1991). *Pseudodifferential Operators and Nonlinear PDE*, volume 100 of *Progress in Mathematics.* Birkhäuser, Boston.

Toldi, J. L. (1989). Velocity analysis without picking. *Geophysics*, 54(02):191–199.

Yilmaz, O. (1987). Seismic data processing. In *Investigations in Geophysics No. 2.* Society of Exploration Geophysicists, Tulsa.