The Reflection Inverse Problem for Acoustic Waves

W. W. Symes

CRPC-TR90107
December, 1990
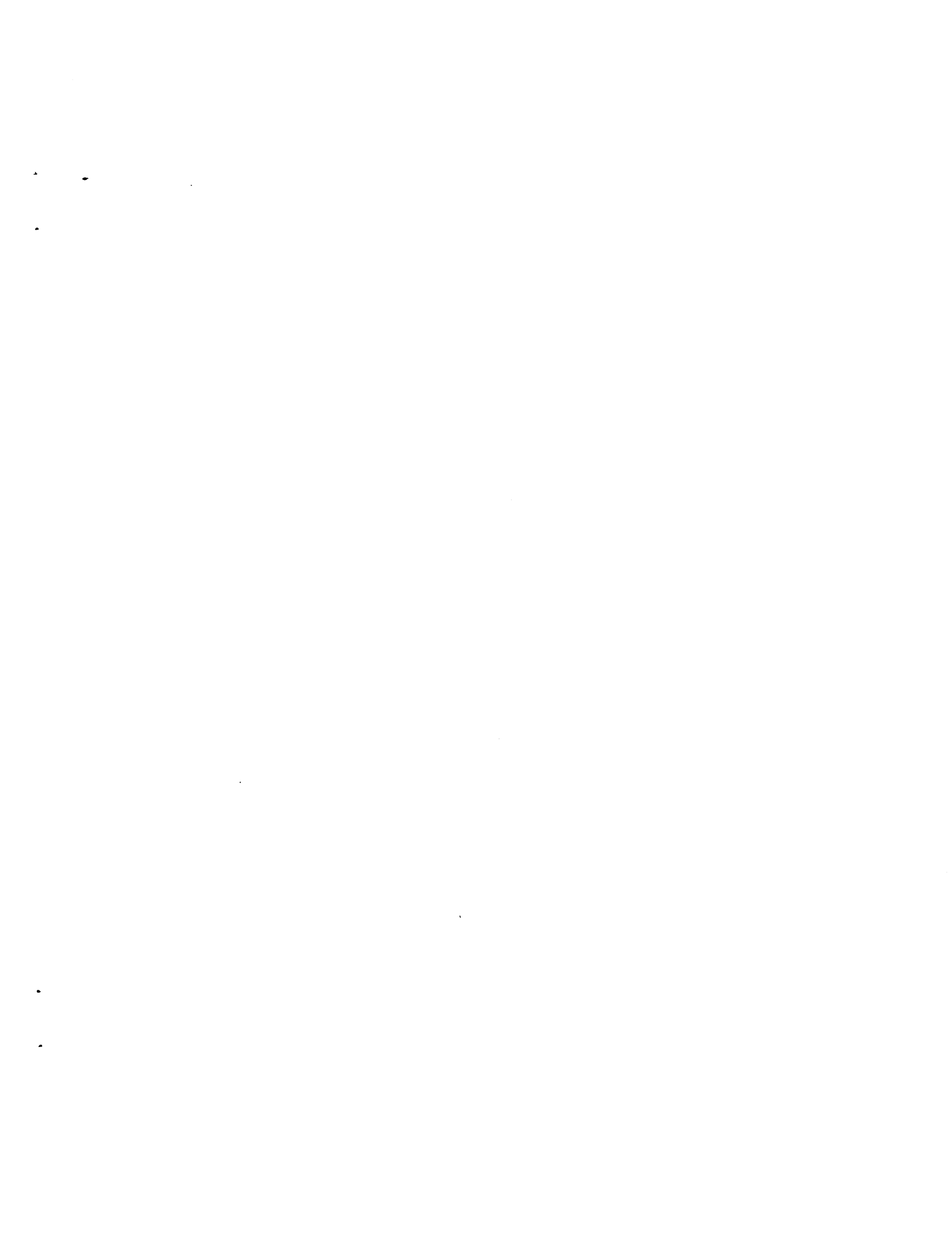
The Reflection Inverse Problem
for Acoustic Waves

W.W. Symes*
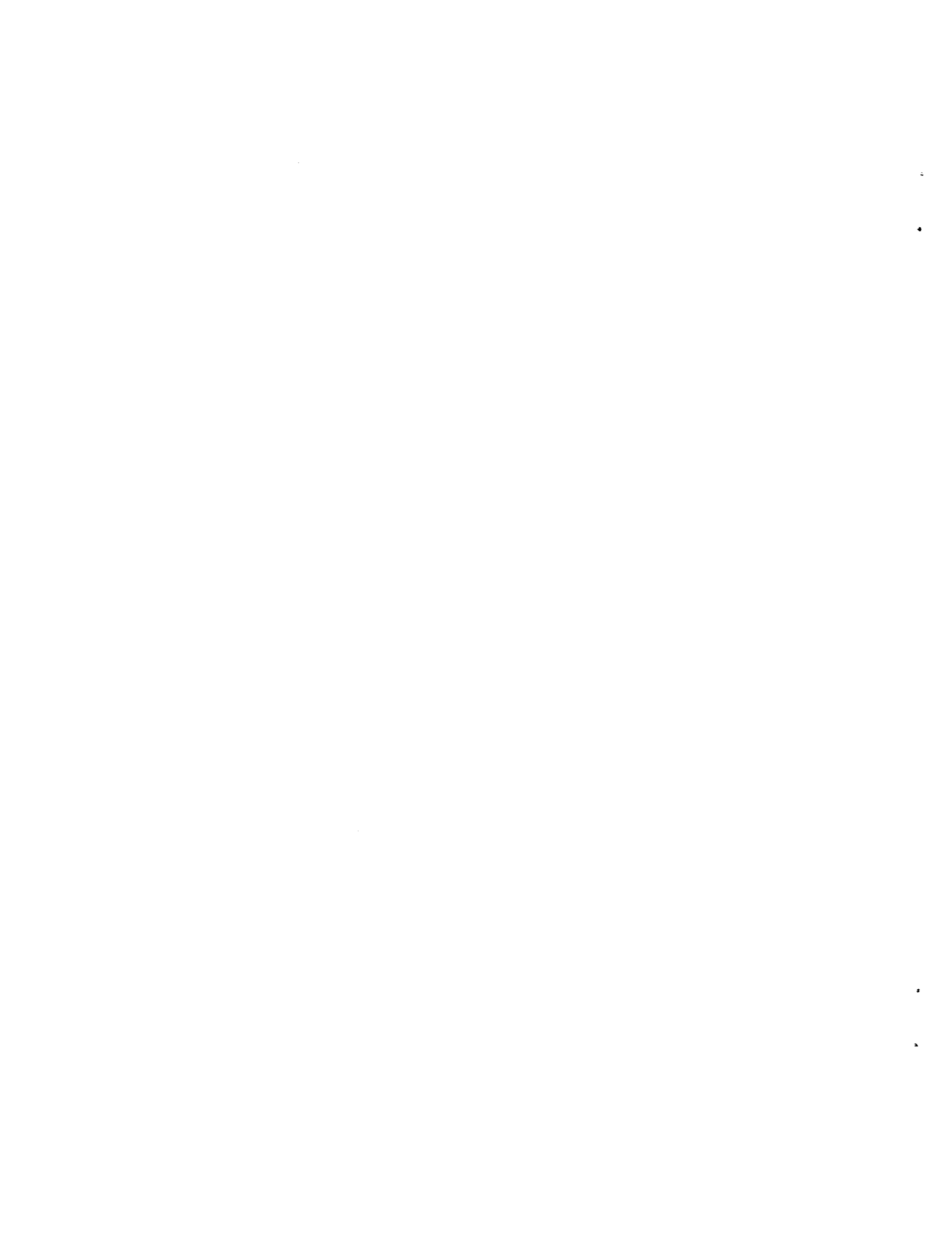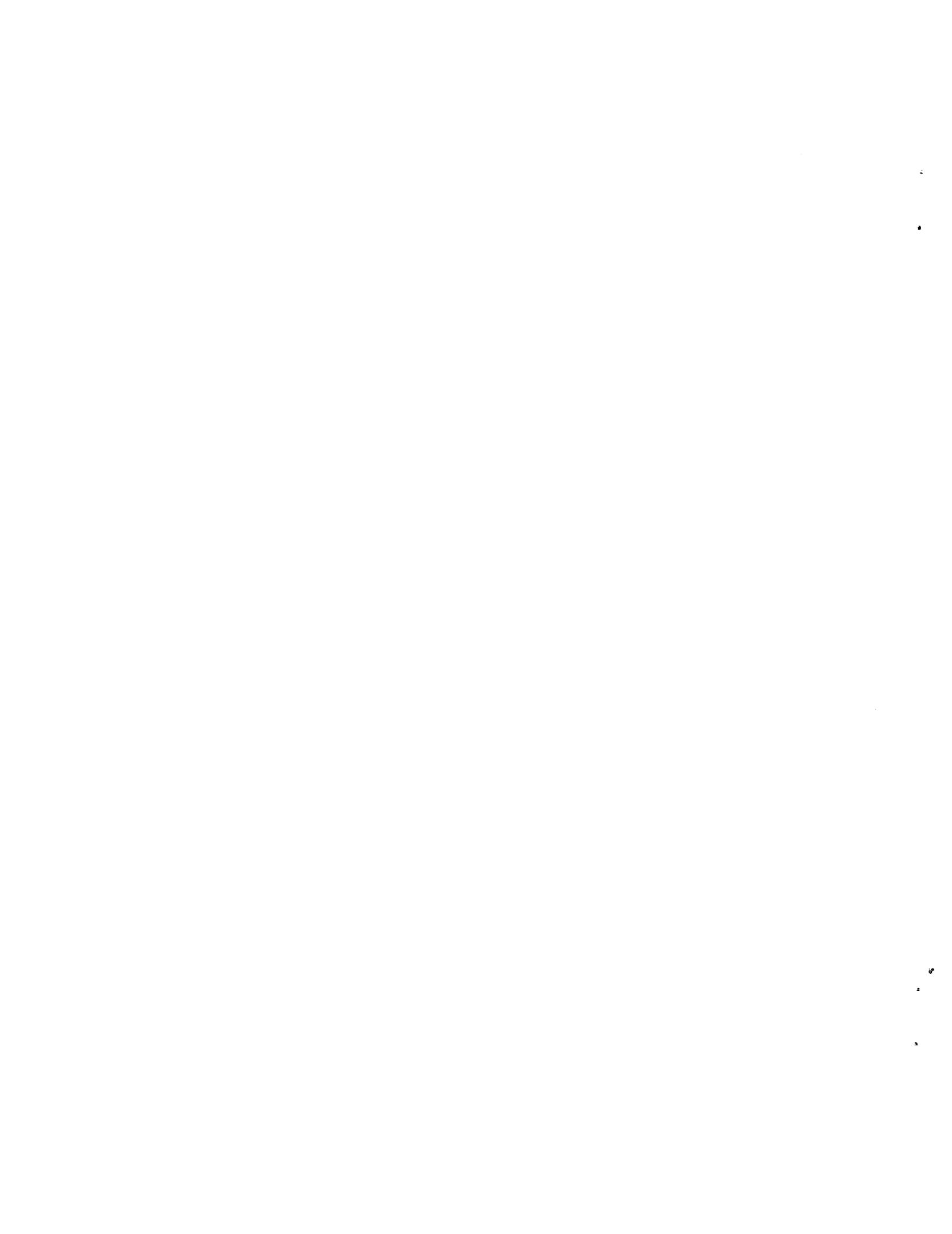
Abstract

We study an acoustic model of the inverse problem of reflection seismology. A straightforward best-fit formulation of this problem has been used by several investigators as the basis for numerical solution of this problem, but typically fails when coupled with local optimization algorithms. We give an explanation for this failure, and suggest a relaxation of the best-fit formulation which may be more amenable to quasi-Newton optimization. Our analysis relies on linearization and on a high-frequency (Fourier Integral) approximation to the scattered field.

1 Introduction

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

$$\frac{1}{v^2}\frac{\partial^2 p}{\partial t^2} - \nabla^2 p = f$$

with appropriate side conditions. Regarding the source term f as known, we study the dependence of p on v, with the aim of understanding the *inverse* problem, i.e. the inference of v from a sampling of p at receiver locations. When both sources (i.e. the support of f) and receivers are separated from the target heterogeneities in v by a hyperplane, it's natural to call this inference the *reflection* inverse problem.

The study of such problems originated in exploration geophysics, and much insight can be gleaned form the literature of that subject.

It is natural to approach this inverse problem by developing an objective function, the minimization of which gives an estimate of the unknown coefficient v. In this paper we study two choices of objective function. The first, a straightforward mean-square data error measure, has been the most common choice for numerical work, but leads to intractable computations: the objective function is essentially non-smooth, and non-convex. We explain
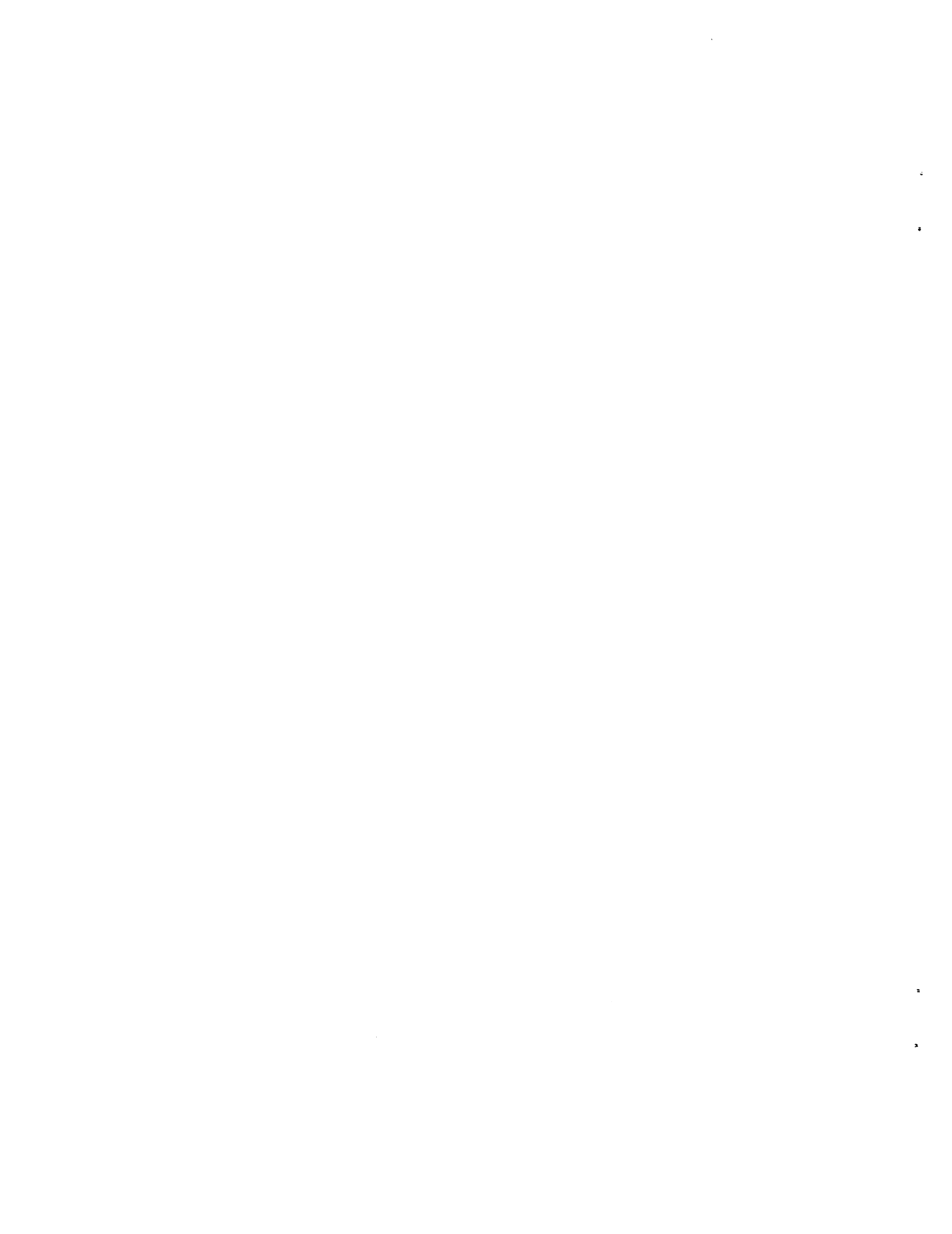
this pathology, and show how enlargement of the model space and addition of suitable constraints may lead to a smooth optimization problem, with qualitatively better-conditioned quadratic approximations and at least local convergence near consistent minima.

The arguments in this paper are for the most part formal, and based on some drastic approximations. Detailed mathematical treatment and extensive numerical experiments for the special case of plane-wave source and layered medium (i.e. v depending only on one coordinate) appear in Santosa and Symes [12], Symes [13], [16] and Symes and Carazzone [17], [18]. Some further discussion on point sources and general media, as discussed here, appears in Symes [14] [15]. These latter references include numerical experiments with the (second) objective functional introduced below.

2 The Model

For the purposes of this paper, the *data* of the inverse problem is the *trace* of p on the hypersurface $\{x_n = 0\}$, multiplied by a smooth function of compact support. The source f is of point support in the space variables, and we allow it to move in the plane $\{x = (x', x_n) : x_n = z_s\}$:

$$f(x, t) = F(t)\delta(x - x_s) \qquad x_{s,n} = z_s \ .$$

Thus p depends on x_s as well. We assume that the source is quasi-impulsive, i.e. $F(t) = \delta(t) +$ a smooth perturbation. Finally, we assume that v is constant, $v = 0$, in the half-space $\{x : x_n < z_b\}$, $0 < z_s < z_b$.

These assumptions violate a number of limitations of sampling and bandwidth important in the treatment of real-world data, e.g. in reflection seismology, but in our judgement leave the mathematical heart of the problem intact.

Thus finally

$$S[v] := p \mid_{x_n=0}$$

is the data of the inverse problem. Since our study is modeled on reflection seismology, we call S the *seismogram*.

The properties of the map $v \mapsto S$ are at present only poorly understood, so we study instead an approximation, obtained by splitting $v = v_b + v_r$ into a *smooth background velocity* v_b and a *rough* or *oscillatory perturbation* v_r. Using regular first-order perturbation theory we write

$$S \cong S_b + S_r$$

where S_b is the background seismogram, i.e. the trace $p \mid_{x_n=0}$ with $v = v_b$, and S_r is the perturbation due to v_r. A great deal of numerical evidence indicates that this approximation is quite accurate so long as v_r is oscillatory. For a theoretical study in the one-dimensional case see Lewis [7]. If v_b is sufficiently smooth, which we assume, then S_b consists of the direct wave, plus possibly refractions. We limit our attention to reflections here; so we assume that S_b has been subtracted out (a nontrivial step in practice!) and identify S with S_r.

S_r is thus the sampling at the receiver locations of the pressure field perturbation δp, which solves

$$\frac{1}{v_b^2} \frac{\partial^2 \delta p}{\partial t^2} - \nabla^2 \delta p = \frac{2v_r}{v_b} \nabla^2 p$$

plus appropriate side conditions. Note the appearance of the *reflectivity* $r = v_r/v_b$; in the sequal we shall use it rather than v_r to represent the rough part of the model.

As is well-known (Cohen and Bleistein [3], Beylkin [1], Rakesh [11]) S can be approximated rather effectively as an oscillatory integral (Fourier Integral Operator) of the form

$$S[v_b]r(x_s; x_r, t)$$
$$\cong \quad f * \int d\xi A[v_s](x_s, x_r, t, \xi) e^{i\phi[v_s](x_s, x_r, t, \xi)} \cdot \hat{r}(\xi) \; .$$

The notation is chosen to emphasize the following points:

(1) The seismogram is a function of the source parameter x_s, the receiver coordinate x_r, and time t;

(2) The *amplitude* or *symbol A* and the *phase* ϕ also depend on a wave vector ξ of the same dimensionality as the space coordinates.

(3) A and ϕ depend further on a point source coordinate x and are convolved in x and t against the source distribution $f(x, t, x_s)$.

(4) The seismogram S, the symbol A, and the phase function ϕ depend functionally on v_b.

(5) The seismogram S depends linearly on the reflectivity r.

For space dimension n, the symbol A behaves for large $|\xi|$ like

$$A_0(x, \xi/|\xi|)|\xi|^{\frac{n-1}{2}}$$

for a suitable smooth function A_0 which is non-zero over a sector in $\xi/|\xi|$ determined by the bicharacteristic geometry (hence by v_b). The phase function ϕ is positively homogeneous of degree 1 in ξ.

3   Least Squares Inversion

With the conventions established so far, we can state a simple version of the least-squares inversion problem:

> *Find v_b, r to minimize*
>
> $$J_{LS}[v_b, r; S_{\text{data}}] = \frac{1}{2} \int \int \int dx_s dx_r dt \, |S[v_b] \cdot r - S_{\text{data}}|^2$$

Here we understand the integral sign to mean integral or sum, as appropriate.

In a typical reflection seismic model in 2D, v_b might be represented by a few tens of parameters, while r requires perhaps $10^5 - 10^6$ parameters for a useful degree of resolution. Thus the least-squares problem is computationally very large, and efficient minimization algorithms are required. By far the most efficient numerical optimization techniques are the descent methods related to Newton's method — when they work. These iterations take steps predicted by the linearized model/data relation so rely for their effectiveness on a close relation between the cost function and its quadratic approximation. Accordingly, we now examine (somewhat formally) the response of S to perturbations in v_b and r.

From the oscillatory integral expression above the perturbation of S due to a change δv_b in v_b is

$$\delta_{v_b} S = \int d\xi (i\delta\phi \cdot A + \delta A)e^{i\phi}\hat{r} \ .$$

This is an oscillatory integral of the same form as that approximating S, with a different symbol. In fact $\delta\phi$ is also homogeneous of order 1, exactly as is ϕ. Therefore the symbol in the above integral grows as $|\xi|^{n+\frac{1}{2}}$ as $|\xi| \to \infty$, i.e. at a more rapid rate than A. It follows that for at least some oscillatory r, smooth δv_b

$$|\delta_{v_b} S| >> |S| \ .$$

Taking this reasoning one step further, one sees immediately that

$$|\delta_{v_b}^2 S| >> |\delta_{v_b} S| \ ,$$

that is, that S is very nonlinear in v_b.

To appreciate the consequences for the behaviour of the cost function J_{LS}, we introduce the factorization

$$\delta_{v_b} S = S Q_1$$

which follows from the calculus of Fourier integral operators (e.g. Duistermaat [4]). Here $Q_1[v_b, \delta v_b]$ is a pseudodifferential operator, i.e. an oscillatory integral of the form

$$Q_1\phi(x) = \int d\xi e^{ix\cdot\xi} q_1(x,\xi)\hat{\phi}(\xi) \ .$$

The symbol $q_1 = q_1[v_b, \delta v_b]$ is of order 1 (i.e. grows like $|\xi|$ for large $|\xi|$) and depends smoothly on v_b, linearly on δv_b. Moreover, Q_1 is essentially skew-adjoint: there is another pseudodiffereential operator Q_0 of order zero (i.e. whose symbol q_0 is bounded as $|\xi| \to \infty$) so that

$$Q_1 + Q_1^T = Q_0$$

where the superscript "T" denotes the transpose, or formal adjoint.

Note in general that for a pseudodifferential operator B with symbol b depending smoothly on a parameter α:

$$B[\alpha]u(x) = \int d\xi b(x,\xi,\alpha)e^{ix\cdot\xi}\hat{u}(\xi)$$

the perturbation of B with respect to α is a pseudodifferential operator of the same order, i.e., whose symbol has the same order of growth as $|\xi| \to 0$ as does b:

$$\delta_\alpha B[\alpha]u(x) = \int d\xi \delta_\alpha b(x,\xi,\alpha)e^{i\xi\cdot x}\hat{u}(\xi) \ .$$

This feature of pseudodifferential operators is due to the fixed nature of the phase: it is always $\xi \cdot x$, independent of the parameters on which the symbol depends. Note the contrast with the behaviour of oscillatory integrals such as S, i.e. Fourier integral operators with phases depending nontrivially on parameters (v_b in the case of S).

With the standard notations

$$\langle \psi, \phi \rangle = \int \bar{\psi}\phi$$

$$\|\psi\| = \langle \psi, \psi \rangle^{\frac{1}{2}}$$

(the integration being over appropriate variables), we can write

$$J_{LS}\left[v_b, r, S_{\text{data}}\right] = \frac{1}{2}\left\|S[v_b]r - S_{\text{data}}\right\|^2$$

$$
\begin{aligned}
\delta_{v_b}J_{LS} &= \langle \delta_{v_b}S[v_b]r, S[v_b]r - S_{\text{data}}\rangle \\
&= \langle S[v_b]Q_1[v_b, \delta_{v_b}]r, S[v_b]r - S_{\text{data}}\rangle \\
&= \langle r, Q_1^T S^T \left[Sr - S_{\text{data}}\right]\rangle .
\end{aligned}
$$

(We have dropped arguments as convenient to make the structure of the expressions clearer). Similarly, the perturbation of J_{LS} with respect to r is given by

$$\delta_r J_{LS} = \langle \delta r, S^T \left[Sr - S_{\text{data}}\right]\rangle .$$

Differentiating once more, we find the following expressions for the blocks of the Hessian operator:

$$\delta_{v_b, v_b}^2 J_{LS} =$$
$$\langle r, \delta_{v_b}Q_1^T S^T [Sr - S_{\text{data}}] + Q_1^T(Q_1^T S^T S + S^T S Q_1)r\rangle$$
$$\delta_{v_b, r}^2 J_{LS} =$$
$$\langle \delta r, Q_1^T S^T S r + S^T S Q_1 r - Q_1^T S^T S_{\text{data}}\rangle$$
$$\delta_{r, r}^2 J_{LS} =$$
$$\langle \delta r, S^T S \delta r\rangle = \|S\delta r\|^2 .$$

Recall that Q_1 is of order 1, hence enhances high-frequency content. Again according to the calculus of oscillatory integrals, $Q_1^T Q_1^T$ is of order 2, hence enhances high-frequency components even more strongly. Thus one might well expect that

$$\left|\delta_{v_b, v_b}^2 J_{LS}\right| >> \left|\delta_{r, r}^2 J_{LS}\right|$$

and this is indeed the case for oscillatory $r, \delta r$ of the same magnitude and smooth $v_b, \delta v_b$. Thus the Hessian is extremely ill-conditioned.

Moreover, the growth rate of J_{LS} as one moves v_b away from the minimizer is many times the overall size of J_{LS} itself. Therefore the growth cannot be sustained over a large change in v_b, and J_{LS} saturates. Consequently J_{LS} tends to be very non-convex, with a very small region of convexity near the global optimum model. See Symes and Carazzone [17], Figure 4 for an actual picture of J_{LS} illustrating these features.

The highly non-quadratic nature of J_{LS} explains the great difficulty of recovery of v_b by least-squares inversion reported frequently in the literature (Gauthier *et al.* [5], Kolb *et al.* [9], Mora [8], for example — see also Santosa and Symes [12]). One can say with confidence that extraction of v_b by means of least-squares inversion and local, Newton-type optimization is *impossible*, unless the initial estimate if v_b is so accurate as to render further refinement almost pointless.

Random or systematic search has been suggested as an alternative technique (e.g. Cao *et al.*. [2]). Such methods may work well when the background velocity may be represented by a few parameters in a known way. In general, severely parsimonious parameterization is likely to introduce unjustified bias, and to fail to sample the model space sufficiently to well-approximate the optimal v_b. On the other hand, refined parameterization generates impossibly large search tasks.

In sum, estimation of v_b *via* the least-squares principle is unlikely to yield useful results in general, or reliable inversion methods.

4 The Coherency Method

Our resolution of the difficulty outlined in the preceding paragraphs begins with two observations:

(i) For *fixed* v_b, J_{LS} is perfectly convex — in fact, quadratic!

(ii) If the set of shot parameter values $\{x_s\}$ reduces to a *singleton*, e.g. only one point source record is used, the minimum value of J_{LS} is essentially independent of v_b.

That is, the inversion of a single shot record is feasible, and constrains only r, *not* v_b. Since this task is practical, it suggests the expedient of viewing r as a function of the shot parameter x_s

$$r = r(x, x_s) \,.$$

Of course, if S_{data} is noise free,

$$S_{\text{data}} = S[v_b^*]r^*$$

then $r(x, x_s) \equiv r^*(x)$ is amongst the minimizers of

$$\int \int dx_r dt \left| S[v_b^*]r(\cdot, x_s) - S_{\text{data}}(\cdot, x_s) \right|^2$$

and has the addition property of *coherence*, or independence of x_s, which we can express as

$$\frac{\partial r}{\partial x_s} \equiv 0 \,.$$

Only coherent reflectivity estimates have any ultimate meaning, since there is only one earth!

The above two conditions can be combined into a single cost functional, for instance:

$$\|S[v_b]r - S_{\text{data}}\|^2 + \sigma^2 \|\partial r/\partial x_s\|^2$$

where r is now allowed to depend explicitly on x_s — with such dependence penalized by the second term, weighted by a parameter σ^2.

This functional is quadratic in r, so the minimization with respect to r presents no difficulties, in principle. On the other hand, as a functional of *both* v_b and r, it is still quite non-convex, for the same reasons as before. Together these two observations suggest *elimination* of r: that is, we define a functional of v_b only by

$$J_{CM}\,[v_b; S_{\text{data}}] =$$
$$\min_r \frac{1}{2} \left\{ \|S[v_b] \cdot r - S_{\text{data}}\|^2 + \sigma^2 \left\| \frac{\partial r}{\partial x_s} \right\|^2 \right\} \,.$$

It is a remarkable fact that this functional is smooth — in fact, nearly quadratic — in its dependence on v_b, despite its rather close relation with the least squares functional! We also conjecture that it is strongly convex for near-consistent data S_{data} and proper choice of σ^2, over a large subset of background velocity models. We are able to give a proof in the plane/wave layered medium case [13].

Minimization of J_{CM} over a smooth class of background velocities v_b is the *coherency optimization problem*. Note that for noise-free data, J_{CM} attains the value 0 for $v_b = v_b^*$, which is clearly its global minimum, and that this minimum is reached by setting $r = r^*$ on the right-hand side. That is, the global minimum is achieved at the correct velocity — and, implicitly, at the correct reflectivity.

In the remainder of this section, we will outline the reasons for the smoothness of J_{CM}, and our reasons for thinking that J_{CM} might be minimized quite efficiently. We give only the formal skeletons of arguments here; precise statements and proofs will be presented elsewhere.

Before starting we take care of a few technical details. The first is that the normal operator

$$S^T S$$

is a pseudodifferential operator of order $n-1$ if the source is impulsive, $f(x,t) = \delta(x-x_s)\delta(t)$, under some ray-geometric restrictions (no caustics in the incident wavefront). This is another immediate consequence of the FIO calculus (Duistermaat) [4]), and is mentioned explicitly in Beylkin [1], Rakesh [11] for example. As shown in Percell's thesis [10], this conclusion is *false* when caustics are present in the incident wave-front — a generic occurrence in heterogeneous media. It is possible to recover the pseudodifferential nature of $S^T S$ by modifying the definition of S. Without going into details, we assume that this has been done. This operator is *elliptic*, i.e. acts as an invertible Fourier multiplier at high spatial frequencies, over a conic sector of wave vectors (the "reflection aperture") determined by the relative positions of sources and receivers and the ray geometry of the background velocity field. Outside of the reflection aperture, which varies with location in the subsurface, $S^T S$ suppresses high-frequency components (these correspond to "off-cable" reflections). Therefore the high-frequency components of r outside the inversion aperture must be constrained *a priori* in solving equations involving $S^T S$.

To accomplish this goal in a well-scaled way, we first modify the definition of S: we assume that the source has point-support, and in its time dependence is a low frequency perturbation of the $(\frac{3-n}{2})$-th derivative of $\delta(t)$: thus

$$f(x,t) = \text{const.} \quad \begin{cases} \delta(x - x_s)t_+^{-\frac{3}{2}} , & n = 2 \\ \delta(x - x_s)\delta(t) , & n = 3 \end{cases} \\ + f_0(x,t)$$

(the distribution $t_+^{-\lambda}$ is defined in Gel'fand and Shilov [6], for example), where f_0 is a smooth function. This amounts to assuming that f, while bandlimited below, behaves as $t_+^{-\frac{3}{2}}$ ($n = 2$) or $\delta(t)$ ($n = 3$) across the upper part of the passband of the seismic signals. Practically, this assumption is realized by preprocessing the data to re-scale it in the frequency domain.

With this modification, $S^T S$ is a pseudodifferential operator of order 2: specifically,

$$S^T S \, r(x_s, x) = \int d\xi \, b(x_s, x, \xi)e^{ix\cdot\xi}\hat{r}(x_s, \xi)$$

where $b \sim b_0(x_s, x, \xi/|\xi|)|\xi|^2$ within the reflection aperture, ~ 0 outside of it, as $|\xi| \to \infty$. Thus $S^T S$ is a family of pseudodifferential operators in x, parameterized by x_s, of order 2. It is a slight technical headache that such a family of operators is *not* a pseudodifferential operator in x *and* x_s; however this is not an essential complication (e.g., Taylor [19], Appendix) and we shall ignore it here.
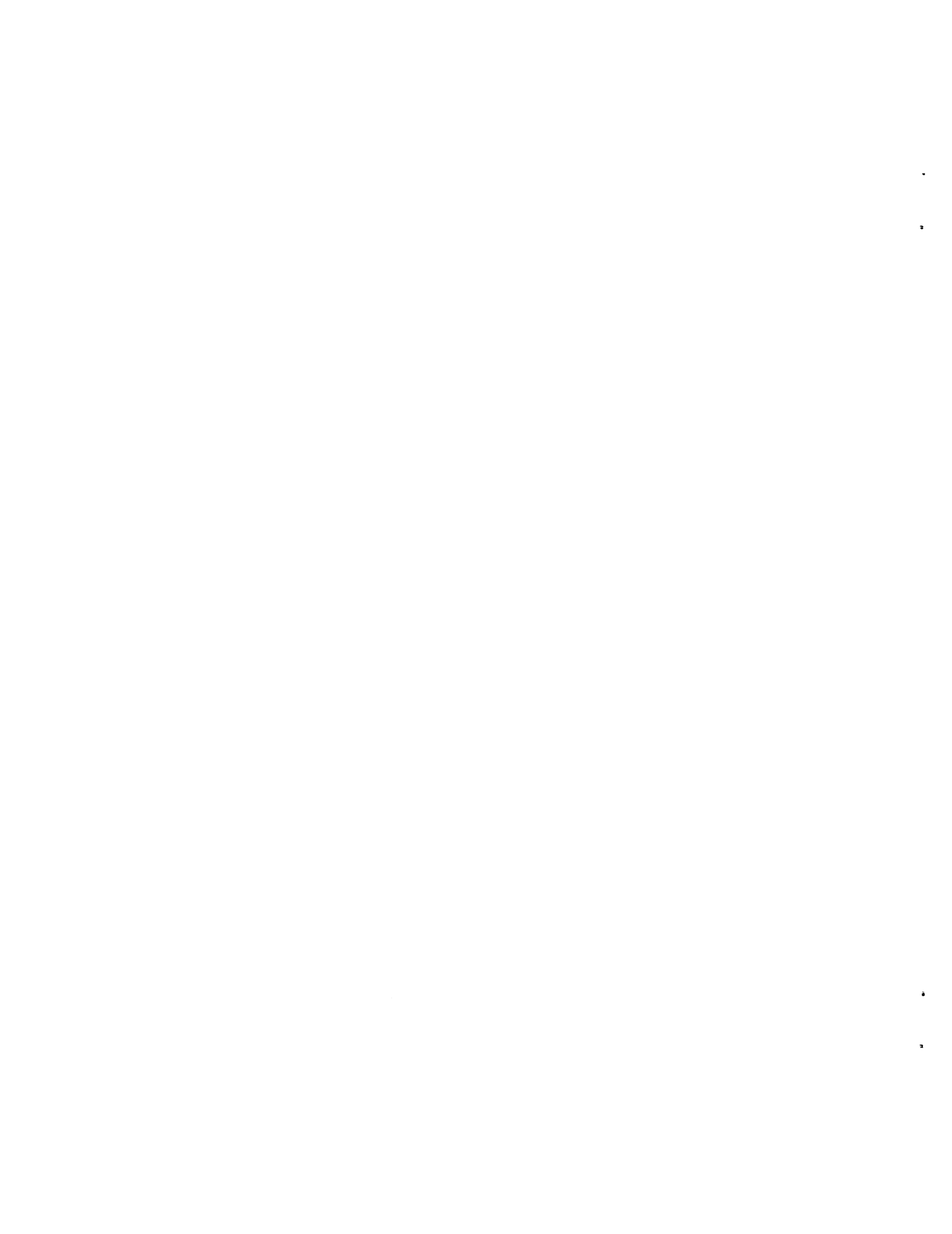
We chose a regularizing operator R, pseudodifferential of order 2 in x and depending parametrically on x_s, so that

$$S^T S + \lambda^2 R$$

is *elliptic* for each x_s as long as $\lambda^2 > 0$. A simple choice is

$$R = I - \nabla_x^2 .$$

This choice is suboptimal, as it also affects the components within the reflection aperture, but for small λ^2 this is probably of little consequence. It will be important in the sequel to write $R = C^T C$, with C a pseudodifferential operator of order 1. This is certainly possible for the simple choice just given with $C = (I - \nabla_x^2)^{\frac{1}{2}}$.

Having disposed of these preliminaries we recall that the coherency method functional J_{CM} is defined by minimizing over r the (regularized) quadratic

$$\frac{1}{2}\left\{ \|Sr - S_{\text{data}}\|^2 + \lambda^2 \langle r, Rr \rangle + \sigma^2 \left\|\frac{\partial r}{\partial x_s}\right\|^2 \right\} .$$

Our goal (and an important step in the proof that J_{CM} is smooth) is to show that the derivative δJ_{CM} of J_{CM} with respect to v_b is of the same size (roughly) as J_{CM} itself. Note the contrast with the behaviour of the least-squares functional J_{LS} described in the last section.

A minimizer of the above quadratic is a solution of the normal equations

$$Nr := [S^T S + \lambda^2 R - \sigma^2 \, \partial^2 / \partial x_s^2]r = S^T S_{\text{data}} .$$

The operator N is (essentially) an elliptic pseudodifferential operator of order 2 in x, x_s. Standard techniques show that N is invertible, under reasonable restrictions on r, and that r depends stably on S_{data} in suitable norms.

Since $S = S[v_b]$, the solution of the normal equations also depends on v_b: $r = r[v_b, S_{\text{data}}]$ also. The dependence of r on v_b is quite erratic — this is another consequence of our analysis of the least-squares problem in the last section.

With these conventions we can calculate the derivative of J_{CM} with respect to v_b:

$$
\begin{aligned}
\delta J_{CM} &= \langle \delta S \cdot r + S \cdot \delta r, Sr - S_{\text{data}} \rangle \\
&\quad + \lambda^2 \langle \delta r, Rr \rangle + \sigma^2 \langle \frac{\partial}{\partial x_s} \delta r, \frac{\partial}{\partial x_s} r \rangle \\
&= \langle \delta S \cdot r, Sr - S_{\text{data}} \rangle \\
&\quad + \langle \delta r, \left(S^T S + \lambda^2 R - \sigma^2 \frac{\partial^2}{\partial x_s^2} \right) r - S^T S_{\text{data}} \rangle \\
&= \langle \delta S \cdot r, Sr - S_{\text{data}} \rangle \\
&\quad + \langle \delta r, Nr - S^T S_{\text{data}} \rangle .
\end{aligned}
$$

Here δr is the (implicit) derivative of $r = r[v_b, S_{\text{data}}]$ with respect to v_b. This could be computed by differentiating the normal equations, but fortunately this effort is unnecessary: because of the normal equations the second term drops out. Thus

$$\delta J_{CM} = \langle \delta Sr, Sr - S_{\text{data}} \rangle .$$

Recalling the factorizations

$$\delta S = SQ_1 \quad \text{and} \quad R = C^T C$$

we calculate

$$
\begin{aligned}
\langle \delta Sr, Sr - S_{\text{data}} \rangle &= \\
& \langle Q_1 r, S^T(Sr - S_{\text{data}}) \rangle \\
&= \langle Q_1 r, (\lambda^2 R - \sigma^2 \, \partial^2/\partial x_s^2) r \rangle \\
& (\text{normal equations again!}) \\
&= \langle Q_1 r, \lambda^2 C^T C + \sigma^2 (\partial/\partial x_s)^T (\partial/\partial x_s) r \rangle \\
&= \lambda^2 \langle C Q_1 r, Cr \rangle + \sigma^2 \langle \partial/\partial x_s \, Q_1 r, \partial/\partial x_s r \rangle \\
&= \lambda^2 \langle Q_1 Cr, Cr \rangle + \sigma^2 \langle Q_1 \, \partial r/\partial x_s, \partial r/\partial x_s \rangle \\
& + \lambda^2 \langle [C, Q_1] r, Cr \rangle + \sigma^2 \langle [\partial/\partial x_s, Q_1] r, \partial r/\partial x_s \rangle .
\end{aligned}
$$

Here we have used the (standard) notation for the *commutator* of two operators:

$$[A, B] = AB - BA .$$

The calculus of pseudodifferential operators shows that the commutator of a pair of operators of orders p and q is an operator of order $p + q - 1$. That is, both

$$[C, Q_1] \quad \text{and} \quad [\partial/\partial x_s, Q_1]$$

are of order 1. On the other hand, recall that Q_1 is essentially skew-adjoint: using the notations of the last section

$$
\begin{aligned}
\langle Q_1 Cr, Cr \rangle &= \frac{1}{2} \langle (Q_1 + Q_1^T) Cr, Cr \rangle \\
&= \frac{1}{2} \langle Q_0 Cr, Cr \rangle
\end{aligned}
$$

where Q_0 is a pseudodifferential operator order zero. Similarly,

$$\langle Q_1 \, \partial r/\partial x_s, \partial r/\partial x_s \rangle = \frac{1}{2} \langle Q_0 \, \partial r/\partial x_s, \partial r/\partial x_s \rangle .$$

The upshot is the formula

$$
\begin{aligned}
\delta J_{CM} &= \langle r, [\lambda^2 (C^T [C, Q_1] + \frac{1}{2} C^T Q_0 C) \\
& - \sigma^2 (\partial/\partial x_s [\partial/\partial x_s, Q_1] + \frac{1}{2} \partial/\partial x_s Q_0 \, \partial/\partial x_s)] r \rangle
\end{aligned}
$$

and the operator in brackets on the right-hand side is of order 2. On the other hand, J_{CM} itself can be written as

$$
\begin{aligned}
J_{CM} &= \frac{1}{2} \langle r, [S^T S + \lambda^2 R - \sigma^2 \, \partial^2/\partial x_s] r - 2 S^T S_{\text{data}} \rangle + \frac{1}{2} \| S^T S_{\text{data}} \|^2 \\
&= \frac{1}{2} \langle r, Nr \rangle - \langle r, S^T S_{\text{data}} \rangle + \frac{1}{2} \| S^T S_{\text{data}} \|^2
\end{aligned}
$$

the quadratic part of which also involves an operator of order 2. Therefore δJ_{CM} should be of roughly the same size as J_{CM}, as claimed.

Note the difference with the least-square functional: because the r appearing in the definition of J_{CM} solves the normal equations, we were able to reduce the order in frequency of the symbols appearing in δJ_{CM}, *via* symmetry considerations.

Further analysis along these lines shows that the formal Hessian $\delta^2 J_{CM}$ is also given by a quadratic form in r defined by a symmetric pseudodifferential operator of order 2. This is not, of course, a proof that $\delta^2 J_{CM}$ is a well-conditioned quadratic form in δv_b, but it is certainly a step in that direction.

More discussion, including an outline of the proof that J_{CM} is smooth and a practical calculation of its gradient, may be found in Symes [14], which also contains some numerical investigations of J_{CM}. Examples of the gradient calculation, and an initial attempt at velocity inversion by optimization of J_{CM}, appear in Symes [15].

References

[1] G. BEYLKIN. Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized radon transform. *J. Math. Phys.*, 26:99–108, 1985.

[2] D. CAO, S. SINGH, and A. TARANTOLA. Simultaneous inversion for background velocity and impedance maps. *Geophysics*, 55:458–469, 1990.

[3] J.K. COHEN and N. BLEISTEIN. An inverse method for determining small variations in propagation speed. *SIAM J. Appl. Math.*, 32:784–799, 1987.

[4] J. DUISTERMAAT. Fourier integral operators,. Lecture notes, Courant Institute, New York, 1973.

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

[6] I.M. GEL'FAND and G.E. SHILOV. *Generalized Functions*, volume I. Academic Press, New York, 1958.

[7] R.M. LEWIS. *Source-Velocity Identification for a Layered Model of Reflection Seismology*. PhD thesis, Department of Mathematical Sciences, Rice University, Houston Texas, U.S.A, 1989.

[8] P. MORA. Nonlinear 2-d elastic inversion of real data. In *Proc. 57th Annual International Meeting*, pages 430–432, New Orleans, 1987. Society of Exploration Geophysicists. Expanded abstract. Also *Geophysics* 52.

[9] KOLB P., F. COLLINO, and P. LAILLY. Prestack inversion of a 1d medium. In *IEEE 74*, pages 498–506, 1986.

[10] C. PERCELL. The effect of caustics in acoustics inverse scattering experiments. Technical Report 89-3, Department of Mathematical Sciences, Rice University, Houston, Texas, U.S.A, 1989.

[11] RAKESH. A linearized inverse problem for the wave equation. *Comm. on P.D.E.*, 13:573–601., 1988.

[12] F. SANTOSA and W. SYMES. An analysis of least-squares velocity inversion. Geophysical Monograph 4, Soc. of Expl. Geophys., Tulsa, 1989.

[13] W. SYMES. Velocity inversion by coherency optimization. Technical Report 88-4, Department of Mathematical Sciences, Rice University, Houston, TX, 1988.

[14] W. SYMES. Multi-offset inversion by differential semblance optimization. Technical Report 90-38, Department of Mathematical Sciences, Rice University, Houston, TX, 1990.

[15] W. SYMES. Non-interactive estimation of the Marmousi velocity model by differential semblance optimization: Initial trials. Technical Report 90-36, Department of Mathematical Sciences, Rice University, Houston, TX, 1990. to appear in The Marmousi Experience: Proceedings of the EAEG Workshop on Practical Aspects of Inversion, ed. G. Gran and R. Versteeg, EAEG, The Hague, 1991.

[16] W. SYMES. Velocity inversion: a case study in infinite-dimensional optimization. *Math. Prog.*, 48:71–102, 1990.

[17] W. SYMES and J. CARAZZONE. Velocity inversion by coherency optimization. Technical Report 89-8, Department of Mathematical Sciences, Rice University, Houston, TX, 1989. (to appear in Proc. of Workshop in Geophysical Inversion, ed. J.B. Bednar, SIAM).

[18] W. SYMES and J. CARAZZONE. Velocity inversion by differential semblance optimization. *Geophysics*, in press, 1990.

[19] M. TAYLOR. Reflection of singularities of solutions to systems of differential equations. *Comm. on Pure and Applied Math.*, 28:457–478, 1975.