

**The Plane-Wave
Detection Problem**

W. W. Symes

**CRPC-TR90035
May, 1989**

Center for Research on Parallel Computati
Rice University
P.O. Box 1892
Houston, TX 77251-1892



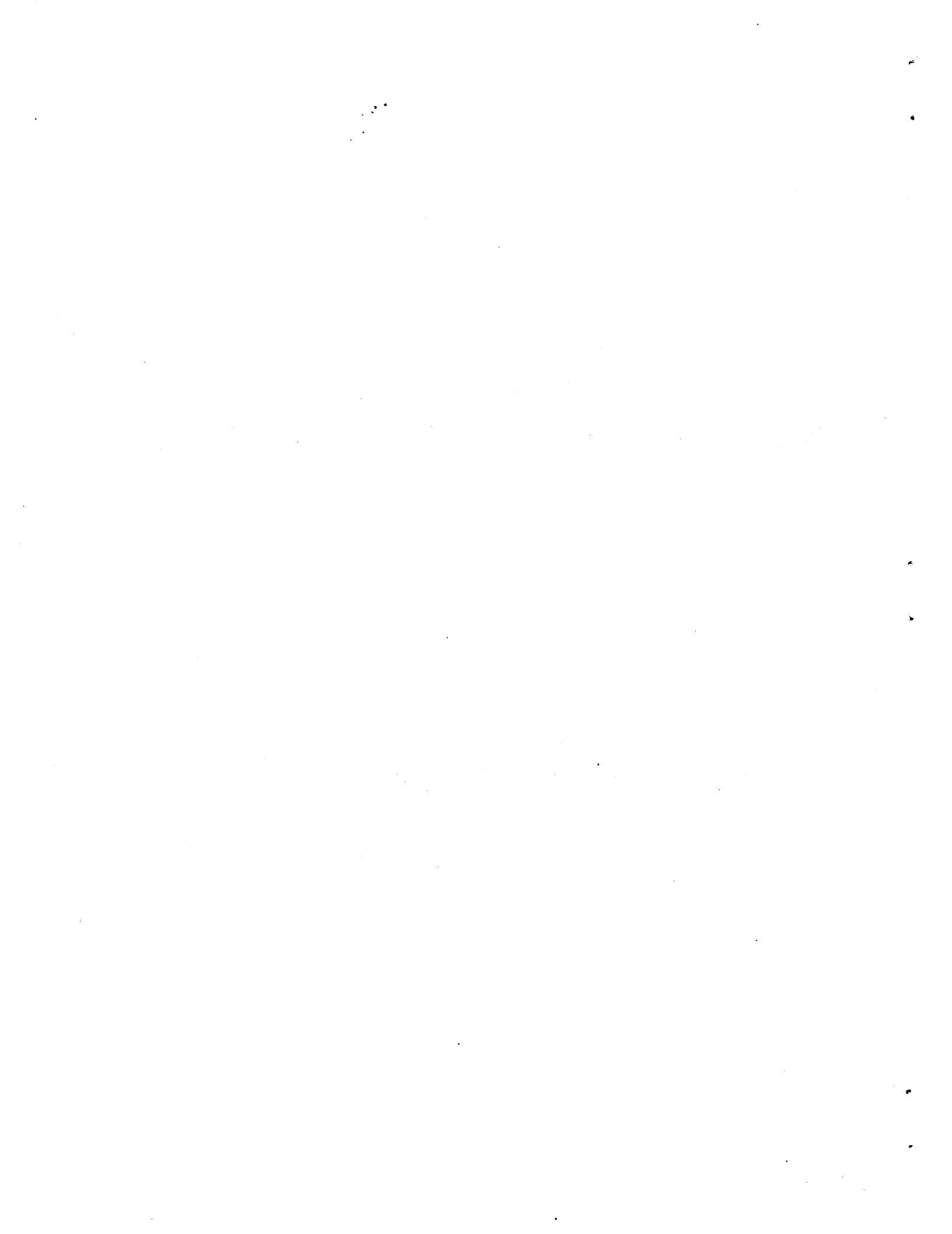
The Plane-Wave Detection Problem

William W. Symes
Department of Mathematical Sciences
Rice University
Houston, TX 77005
U.S.A

May 1989

Abstract

We study in some detail a simple nonlinear estimation problem, which shares several important features with some inverse problems in wave propagation. We consider the estimation of waveforms and incidence angles of transient plane waves from measurements along a line segment. We formulate this estimation problem as a nonlinear least-squares problem in several ways. We show that the "natural" formulation, *output least squares*, is severely ill-posed because of the extreme nonlinearity of the model/data relation. We suggest an alternate formulation, the *penalized coherency method*, and show that this alternative optimization problem is well-posed. We use throughout as our main analytical tool G. Chavent's theory of *quasi convex sets*.



0 Introduction

This paper presents a detailed study of a simple but nontrivial nonlinear least-squares problem in infinite dimensions. Its main purpose is to illustrate the following propositions:

- (i) that nonlinear least-squares problems may be ill-posed for strictly nonlinear reasons, i.e. they may have uniformly coercive quadratic (“linearized”) models but still exhibit arbitrarily unstable dependence of the solution on the data;
- (ii) that it is sometimes possible to replace such problems by well-posed least-squares problems having the “same” solutions for consistent data.

We call the problem studied here the “(plane-wave) detection problem.” It is a simple model for some inverse problems in wave propagation previously studied by the author (Symes [1988], Symes and Carazzone [1989]). Suppose that a scalar field in three-dimensional space-time is sampled at every point in the interval $[-1, 1]$ on the x -axis:

$$z(x, t) = U(x, 0, t) \quad -1 \leq x \leq 1 .$$

Suppose moreover that U is *a priori* known to be a plane wave moving at speed 1, except possibly for some noise:

$$U(x, y, t) \cong u(t - x \sin \theta - y \cos \theta) .$$

The *goal* of the plane wave detection problem (or detection problem, for short) is to estimate the waveform $u(t)$ and the incidence angle θ made with the y -axis, or equivalently its sine $s = \sin \theta$.

Our primary focus in this paper is on aspects of the detection problem affecting the feasibility and efficiency of computational methods. For example, because one believes that the noise in the measurement z is small in the mean-square sense, or for statistical reasons (e.g. Tarantola (1987)), one may naturally attempt to fit a prediction of z in the mean-square sense. If the waveform is $u(t)$ and the direction sine is s , one predicts the measurement

$$\phi_s[s, u] = u(t - sx).$$

The optimal choice of model $[s, u]$ is then the solution of the *output least-squares* problem

$$\min_{s, u} \|\phi_s[s, u] - z\|^2$$

where $\|\cdot\|$ denotes the L^2 norm on a suitable domain. This formulation may be attacked numerically after suitable discretization. Note that ϕ_s is linear in u but quite nonlinear in s .

Consider for a moment a more general class of problems, in which physical theory connects a set of model parameters $\{m\}$ to a set of data $\{z\}$ through a mapping: $z = \phi(m)$ (in the plane wave detection problem, e.g. $m \sim [s, u]$). Whether through solution of a least-squares problem or by some other means, one obtains an estimate $m = S[z]$ of the model from the data z . In this paper, we take the point of view that such an estimator S is satisfactory if it has the following properties:

- (i) if z is consistent, i.e. $z = \phi[m]$ for m in an *a priori* prescribed admissible set of models, then $m = S[z]$;
- (ii) S is locally Lipschitz continuous, and is well-defined on a neighborhood of the set of consistent data, in the sense of suitable metrics;
- (iii) S is computable by means of local (Newton-type) mathematical programming techniques.

Our main results are that the output-least-squares problem stated above *cannot* produce an estimator with these properties, and that a variant on output-least-squares *does* produce an estimator satisfying these conditions. We call the variant the *penalized coherency method* — the reasons for this terminology will be evident.

We will shortly describe the results in more detail, but first we offer a few comments concerning conditions (i)–(iii).

Condition (i) serves to connect the estimator and the estimation problem (i.e. with the “physics”). It is certainly very stringent — for example, we could replace (i) by the requirement that a family of estimators be given which arbitrarily well approximate a model from its (consistent) data set. Such families are produced by numerical methods, of course, and also by regularization of ill-posed estimation problems, for example (see Tihonov and Arsenin [1974]). It might even be satisfactory that a (linear or nonlinear) *projection* of the model be reproduced. We will stay mostly with the strong version of condition (i) for the sake of simplicity.

Condition (ii) could be paraphrased: “stability for low-noise data sets.” It guarantees that one is rewarded for efforts in the direction of

- more accurate data collection
- more accurate basic physical modeling.

It is motivated by the presumption that the theory is accurate, i.e. that models exist which predict experimental data up to small errors — small in the sense of the “suitable norms” mentioned in the statement. Repeated experiments then yield necessarily data sets with small differences. Obviously this necessary “small scatter” condition is experimentally verifiable (for a given choice of norm), at least in principle. Condition (ii) asserts that the model estimates should have differences of sizes proportional to the sizes of the data differences. Thus the smaller the measurement errors, the more unambiguous the model estimate.

Weaker notions of continuity could be used, but it is not clear that these would be as useful in practice as Lipschitz continuity, which is a qualitatively maximal notion of stability.

Note that nothing is said in conditions (i) and (ii) about the statistical nature of data or estimation noise — only its size is addressed. Statistical assumptions about the data would doubtless entail consequences for the model statistics. Rather ambitious attempts have appeared recently to characterize the solution of inverse problems *via* statistical notions (see especially Tarantola (1987)). These characterizations are usually difficult or impossible to apply in practical situations, whereas the much less demanding conditions (i) and (ii) can sometimes be verified.

We shall have little to say about the “high-noise” case, i.e. when either the data tends to be very inaccurate, or the model grossly incomplete, or

both. We shall assume that the data signal is close to that which would be produced by a plane-wave, and ask that the direction sine and waveform be estimated with comparable accuracy.

Condition (iii) is motivated by the role of this problem as a simple relative of a number of inverse problems in wave propagation, with which it shares central analytical properties. For these latter problems, the scalar parameter s is replaced by a vector in a high-dimensional space, and evaluation of the analogue of the model-to-data map ϕ (and of its derivatives) is very computation-intensive, even with present-day supercomputers. The vastly greater efficiency of smooth local optimization (quasi-Newton) methods, as compared to exhaustive or Monte-Carlo search, or to non-smooth (subgradient) techniques motivated (iii): these latter options are simply out of the question for the more complex problems for which plane wave detection is a model.

Of course, estimators with properties (i)—(iii) may or may not exist. Our principal result is the construction of a family of such estimators for the plane wave detection problem.

We begin in Section 1 with an analysis of the output least-squares problem, formulated in a suitably precise way. Our main tool throughout is the *quasi-convexity theory* of G. Chavent [1988] which gives sufficient conditions for a nonlinear least-squares problem to have a unique stable, global solution (minimum) which is also the unique local minimum in a prescribed neighborhood, and also estimates the size of the set of admissible data (a neighborhood of the set of consistent data) for which this uniqueness and stability property holds (nonuniqueness of the minimum is regarded as in-

stability). We show that the output least-squares problem never satisfies the conditions of Chavent's theory, and that moreover arbitrarily many local minima exist in arbitrarily small neighborhoods of a dense set of models *even for error-free data*. Therefore application of a local optimization algorithm to the output least squares formulation of our problem can *never* produce an estimator satisfying (i)—(ii) above. The essence of the difficulty is that, depending on choices of topology, either ϕ_δ is not differentiable, or $D\phi_\delta$ is not coercive.

We also study the "bandlimited" version of the output least-squares problem, in which ϕ_δ is replaced by

$$\begin{aligned}\phi_\delta(x, t) &= \int d\tau f(t - \tau)u(\tau - sx) \\ &= f * \phi_\delta(x, t)\end{aligned}$$

the convolution being done in t . If ω is a peak-frequency for the transfer function or wavelet f we show that the amount of noise permitted in the data (before uniqueness fails) is $O(1/\omega)$, and that the direction sine must also be known *a priori* to a precision of $O(1/\omega)$ in order for local optimization methods to yield unique, stable solutions.

Note that the range of ϕ_δ is characterized by the first-order hyperbolic equation: a function $\tilde{u}(x, t)$ satisfies $\tilde{u} = \phi_\delta[s, u]$ for some s, u iff

$$W[s, \tilde{u}] = \frac{\partial \tilde{u}}{\partial x} - s \frac{\partial \tilde{u}}{\partial t} = 0.$$

W is the *coherency operator*. In Section 2 we introduce an *augmented model space* consisting of direction sines s and *two-dimensional signals* \tilde{u} . We form

the map

$$\tilde{\phi}_{f,\sigma}[s, \tilde{u}] = \begin{pmatrix} f * \frac{\partial \tilde{u}}{\partial t} \\ \sigma W[s, \tilde{u}] \end{pmatrix}$$

which gives a *cover* of the output least-squares problem, in the following sense: if

$$\tilde{\phi}_{f,\sigma}[s, \tilde{u}] = \tilde{z} = \begin{pmatrix} z \\ 0 \end{pmatrix}$$

and z is in the range of $\tilde{\phi}_f$, then

$$u(t) = \int_{-1}^1 dx \frac{\partial \tilde{u}}{\partial t}(x, t + sx)$$

solves

$$\phi_f[s, u] = z.$$

Thus the constrained problem

$$\min_{[s, \tilde{u}]} \left\| f * \frac{\partial \tilde{u}}{\partial t} - z \right\| \quad \text{subj. } W[s, \tilde{u}] = 0$$

has the “same” solutions as the output least-squares problem for consistent data, (i.e. $z \in \text{Range } \phi_f$, but involves only differentiable maps. Since you can't get something for nothing something must be wrong. In Section 2 we study the level sets $W^{-1}(g)$, and show that for a dense set of g , these are *not submanifolds* of the model space. In particular, the feasible set $W^{-1}[0]$ for the constrained problem is not a submanifold, having “cusps” at a dense set of points. Thus analysis of the dependence of solutions on data *via* the implicit function theorem is impossible, as is the use of Lagrange multiplier methods to compute the solution.

We conjecture that *all* level sets of W fail to be submanifolds. This would provide an interesting infinite-dimensional counterpoint to Sard's theorem.

In Section 3 we examine the *penalized coherency* least-squares problem

$$\min_{s, \tilde{u}} \|\tilde{\phi}_{\delta, \sigma}[s, \tilde{u}] - \tilde{z}\|^2$$

which also has the "same" solutions as output least-squares for consistent data (i.e. $\tilde{z} = (z, 0)^T$ with $z \in \text{Range}\phi_\delta$). We are able to show that a finite range of the penalty parameter σ exists for which this problem has a unique solution, stably dependent on the data, in a ball of positive radius in the parameter space. We give constructive estimates for the radius of this ball and for the amount of data noise permitted. In other words, the penalized coherency problem is *locally well-posed*, whereas the output least-squares problem is not. We are also able to show that a suitably constructed quasi-Newton method is assured of convergence to the global minimum.

An interesting aspect is that the range of values for the penalty parameter σ for which the penalized coherency problem is well-posed on a fixed domain is *finite*. One *cannot* drive this parameter to ∞ , as could be done to solve a smooth constrained problem, or in Tihonov regularization, without losing well-posedness.

Notation:

Norms in the L^2 based Sobolev spaces H^k will be denoted by $\|\cdot\|_k$ (thus $\|\cdot\|_0$ is the L^2 norm). Domains are either evident from context or specified explicitly. The ball of radius r in a Banach space E , centered at f , is $B_r(f, E)$; or $B_r(f)$ if E is supposed to be clear from context.

Acknowledgement. I take great pleasure in thanking Professor Guy Chavent for his insights into the mechanics of nonlinear least-squares, which were critical to the work reported here. I am grateful to the program CEREMADE at Universite de Paris IX (Dauphine) for its hospitality and support during the writing . This research was partially supported by the Office of Naval Research under grant N00014-89 -J1115 and by the National Science Foundation under grants DMS 86-03164 and 89-05878.

1 Output Least-Squares Detection

We begin this section by formulating the plane-wave detection problem precisely as a nonlinear least-squares problem. We find immediately that in the impulsive case, the mapping from direction (sine) and waveform to observed signal must be restricted to a domain of relatively smooth functions in order that it be of class C^2 . Having made this rather unpleasant restriction, we then find that the quasiconvexity theory of Chavent [1988] does not apply at all, in the impulsive case, and requires that we know the direction sine to "within a wavelength," in the bandlimited case, if we are to be assured of well-posedness. Moreover, we show, in the impulsive case, that multiple minima of the least-squares functional *inevitably appear*. Essentially, the output-least-squares formulation is useless for computational purposes, in the sense made precise in the introduction.

As mentioned before, we shall assume that

- (i) the waveform and signal are square-integrable;
- (ii) the support of the waveform u is known *a priori* to be contained in the unit interval $I = [0, 1]$, i.e. $u \in L^2(I) \subset L^2(\mathbb{R})$.

Note that the direction sine is necessarily contained in the interval $[-1, 1]$. The signal resulting from the waveform u at direction sine s is (for the impulsive case $f = \delta$):

$$\phi_\delta[s, u](x, t) = u(t - sx).$$

We may regard ϕ as mapping

$$\phi_\delta : [-1, 1] \times L^2[0, 1] \rightarrow L^2([-1, 1] \times \mathbb{R}).$$

The output-least-squares formulation of the plane-wave detection problem is:

given data

$$z \in L^2([-1, 1] \times \mathbb{R})$$

find

$$(s, u) \in [-1, 1] \times L^2[0, 1]$$

to minimize

$$J(s, u; z) = \|\phi_\delta(s, u) - z\|_0^2.$$

The following observations are trivial:

- (i) ϕ_δ , hence J , is continuous, but not locally uniformly continuous, hence *a fortiori* not differentiable, even once.
- (ii) In order that ϕ_δ be differentiable of class C^2 , it is necessary to restrict its domain as follows:

$$\phi_\delta : [-1, 1] \times H_0^2[0, 1] \rightarrow L^2([-1, 1] \times \mathbb{R}).$$

The least squares problem now appears to require regularization, *even if s is known exactly and the data is noise-free*. Thus resolution must be restricted, even though ϕ_δ is not a smoothing operator! In fact, the situation is much worse. To see how this is so, we recall Chavent's theory of *quasiconvex sets* [1988], which gives a systematic approach to the study of well-posedness for nonlinear least-squares problems.

To decide whether a set D in an (abstract) Hilbert space F is quasiconvex, one equips it with a collection \mathbb{P} of C^2 paths $P : [0, 1] \rightarrow D$ satisfying

(i) if $X, Y \in D$, there exists $P \in \mathbb{P}$ so that

$$P(0) = X, \quad P(1) = Y$$

(ii) \mathbb{P} is stable under restriction

(iii) a nondegeneracy condition on critical points.

The length $\delta(p)$ of $P \in \mathbb{P}$ is defined as usual:

$$\delta(P) = \int_0^1 \|\dot{P}\|$$

and the "kinematic radius of curvature" ρ by

$$\rho(\nu) = \begin{cases} \frac{\|\dot{P}(\nu)\|^2}{\|\ddot{P}(\nu)\|} & \text{if } \ddot{P}(\nu) \neq 0 \\ \infty & \text{if } \ddot{P}(\nu) = 0, \dot{P} \neq 0 \end{cases}$$

through the nondegeneracy condition otherwise.

If $\langle \dot{P}(\nu), \ddot{P}(\nu) \rangle = 0$, as occurs for example if ν is arc-length, $\rho(\nu)$ reduces to the usual radius of curvature, $\bar{\rho}(\nu)$; otherwise it is an underestimate.

A pair (D, \mathbb{P}) consisting of a set D and a collection of paths \mathbb{P} having the above properties ("pseudosegments") is called *quasiconvex* if there exist a neighborhood $V \supset D$ and a continuous function $\epsilon : V \rightarrow (\mathbb{R}^+ \setminus \{0\}) \cup \{+\infty\}$ satisfying:

(i) ϵ is uniformly positive on D

(ii) if $z \in V$, $P \in \mathbb{P}$ such that

$$\|P(j) - z\| < \text{dist}(z, D) + \epsilon(z), \quad j = 0, 1,$$

and

$$\alpha(\nu) = \begin{cases} \text{angle between } P(\nu) - z \text{ and} \\ \bar{P}(\nu) - \frac{\langle \bar{P}(\nu), \dot{P}(\nu) \rangle}{\|\dot{P}(\nu)\|} \frac{\dot{P}(\nu)}{\|\dot{P}(\nu)\|} \\ \text{if } \bar{P}(\nu) \neq 0, \\ \pm \frac{\pi}{2} \text{ else} \end{cases}$$

then

$$\sup_{\nu \in [0,1]} \frac{\|P(\nu) - z\|}{\bar{\rho}(\nu)} \cos \alpha(\nu) < 1. \quad (1.1)$$

Compare Chavent [1988] pp. 4-7, where the geometrical meaning of (1.1) is explained. In essence, (1.1) demands that paths in \mathbb{P} connecting points in D as near as possible to z never go too far away from z relative to their radii of curvature.

The importance of quasiconvexity stems from the following result (Chavent [1988] Theorem 2.9): if one can equip D with a collection \mathbb{P} of pseudo segments and a neighborhood V so that the above conditions hold, then the projection from V onto D is unique, and other local minima of the distance function occur at distances larger than $\text{dist}(z, D) + \epsilon(z)$. The application to least-squares problems consists in taking $D = \phi(C)$, where $\phi : C \rightarrow F$ is defined on a set C in another Banach space E , and is of class C^2 . The minimization of $\|\phi(x) - z\|_F^2$ over $x \in C$ yields $\phi(x) =$ projection of z on $\phi(C)$.

In this connection we prove

Theorem 1.1 *Let ϕ_δ be the impulsive signal map defined above, and set*

$$C = [-1, 1] \times (H_0^2[0, 1] \cap B_1(0; H^1[0, 1])).$$

Let Π be any collection of C^2 paths in C satisfying

- (i) $\mathbb{P} = \{\phi \cdot \pi : \pi \in \Pi\}$ *is a collection of pseudosegments of $\phi(C)$*
- (ii) *For any tangent vector $[\delta s, \delta u] \in \mathbb{R} \times H_0^2[0, 1]$, there exists $\pi \in \Pi$ so that $\dot{\pi}(0) = [\delta s, \delta u]$*

Then $(\pi(C), \mathbb{P})$ is not quasiconvex.

Proof. In view of (1.1) it is required to produce

- (i) $z \in L^2([-1, 1] \times \mathbb{R}) \setminus \phi(C)$ at arbitrarily small distance to $\phi(C)$;
- (ii) a path $P \in \mathbb{P}$ for which the angle between $P(0) - z$ and $\bar{P}(0)$ is bounded away from $\pi/2$, and $\bar{p}(0)$ is arbitrarily small relative to $\|P(0) - z\|_0$.

Since any target vector is supposed to be fit by a path from \mathbb{P} , we select $[s, u] \in C$ with $s = 0$ and u to be chosen momentarily, and the tangent vector $(1, 0)$. Thus $\pi \in \Pi$ is given by

$$\pi(\nu) = (\nu + \nu^2 s(\nu), u + \nu^2 \tau(\nu))$$

where $s : [0, 1] \rightarrow [\mathbb{R}]$ and $\tau : [0, 1] \rightarrow H_0^2[0, 1]$ are C^2 . In the sequel it will be important to allow u to range over an H^1 -bounded set. Since $\pi(\nu)$ lies in

the H^1 -bounded set C , it follows that $r(\nu)$ will range over an H^1 -bounded set. Thus

$$\begin{aligned}
P(\nu)(x, t) &= \phi_\delta \cdot \pi(\nu)(x, t) \\
&= u(t - (\nu + \nu^2 s(\nu))x) + \nu^2 r(\nu; t - (\nu + \nu^2 s(\nu))x) \\
\dot{P}(\nu)(x, t) &= -(1 + 2\nu s(\nu) + \nu^2 \dot{s}(\nu))x \frac{\partial u}{\partial t}(t - (\nu + \nu^2 s(\nu))x) \\
&\quad + 2\nu r(\nu; t - (\nu + \nu^2 s(\nu))x) + \nu^2 \dot{r}(\nu; t - \nu + \nu^2 s(\nu))x \\
&\quad - \nu^2 (1 + 2\nu s(\nu) + \nu^2 \dot{s}(\nu))x \frac{\partial r}{\partial t}(\nu; t - (\nu + \nu^2 s(\nu))x) \\
\ddot{P}(0)(x, t) &= -(2s(0))x \frac{\partial u}{\partial t}(t) + x^2 \frac{\partial^2 u}{\partial t^2}(t) + 2r(0; t).
\end{aligned}$$

The arc length $\bar{\nu}$ is given by

$$\bar{\nu} = \int_0^\nu \|\dot{P}\|_0$$

so that

$$\begin{aligned}
\bar{\nu}(\nu) &= \nu \left\| \frac{\partial u}{\partial t} \right\|_0 + O(\nu^2) \\
\nu(\bar{\nu}) &= \bar{\nu} \left\| \frac{\partial u}{\partial t} \right\|_0^{-1} + O(\bar{\nu}^2)
\end{aligned}$$

Thus we can assume that ν is arc-length, and π is parameterized as before, but with r ranging over an L^2 -bounded set as u ranges over an H^1 -bounded set.

Now set, for arbitrary $\mu > 0$

$$z(x, t) = \phi_\delta(0, u)(x, t) + \mu \ddot{P}(0)(x, t) / \|\ddot{P}(0)\|_0.$$

Evidently $z \notin \phi(C)$ (unless $u \equiv 0$, which we exclude), and

$$\|P(0) - z\|_0 = \mu.$$

We will eventually want the entire curve $P(\nu)$ to sit inside the ball $B_{2\mu}(z)$. We simply take a sufficiently short initial segment. It is important that the path length be uniform as ν varies over a bounded set in H^1 . As noted above, r then varies over a bounded set in L^2 . On the other hand,

$$\|P(\nu) - P(0)\|_0 \leq \text{const}(\sqrt{\nu}\|u\|_1 + \nu^2\|r\|_0)$$

so it is possible to choose a subinterval $[0, \nu]$ uniformly in u (over a bounded set in H^1), so that the corresponding segment of P is contained in $B_{2\mu}(z)$. Since Π is stable under restriction, the inequality in (1.1) stands or falls independently of the restriction as well.

A lower bound on $\|\tilde{P}(0)\|_0$ follows from a trivial consequence of the equivalence of norms in \mathbb{R}^n :

Lemma 1.1 *For some $\lambda > 0$, all $a, b, c \in \mathbb{R}$*

$$\int_{-1}^1 dx |ax^2 + bx + c|^2 \geq \lambda^2(a^2 + b^2 + c^2).$$

Thus

$$\begin{aligned} \|\tilde{P}(0)\|_0^2 &= \int_{-1}^1 dx \int dt \left| -(2s(0)x \frac{\partial u}{\partial t}(t) + x^2 \frac{\partial^2 u}{\partial t^2}(t) + 2r(0, t)) \right|^2 \\ &\geq \lambda^2 \int dt \left\{ (2s(0))^2 \left| \frac{\partial u}{\partial t}(t) \right|^2 + \left| \frac{\partial^2 u}{\partial t^2}(t) \right|^2 \right\} \\ &\geq \lambda^2 \left\| \frac{\partial^2 u}{\partial t^2} \right\|_0^2. \end{aligned}$$

Since we supposed that the parameter ν is arc-length,

$$\bar{\rho}(0)^{-1} = \frac{\|\tilde{P}(0)\|_0}{\|\dot{P}(0)\|_0^2}.$$

Finally the angle $\alpha(0) = 0$, since we choose $z = P(0) + \mu\tilde{P}(0)$. Thus

$$\begin{aligned} \frac{\|P(0) - z\|}{\bar{\rho}(0)} \cos \alpha(0) &= \mu \frac{\|\tilde{P}(0)\|_0^0}{\|\dot{P}(0)\|_0^2} \\ &\geq \lambda^2 \mu \frac{\|\frac{\partial^2 u}{\partial t^2}\|_0}{\|\frac{\partial u}{\partial t}\|_0^2}. \end{aligned}$$

Many choices of u will cause this ratio to $\rightarrow \infty$, destroying quasiconvexity.

For example, choose u

$$u(t) = u_0(t) + \delta u(t)$$

with $u_0 \in H_0^2[0, 1]$ arbitrary and

$$\delta u(t) = \frac{1}{\omega} x(t) \sin \omega t, \quad x \in C_0^\infty(0, 1).$$

Then $\|u\|_0$ and $\|\frac{\partial u}{\partial t}\|_0$ are uniformly bounded in ω , but $\|\frac{\partial^2 u}{\partial t^2}\|_0 \rightarrow \infty$ as $\omega \rightarrow \infty$. Moreover $\phi_\delta(0, u) \rightarrow \phi_\delta(0, u_0)$ in $L^2([-1, 1] \times \mathbb{R})$ as $\omega \rightarrow \infty$.

Suppose that $\phi_\delta(C)$ were quasiconvex. Since $\epsilon_0 := \epsilon(\phi_\delta(0, u_0)) > 0$ (recall ϵ must be uniformly positive on D), there exists $r_0 > 0$ so that $B_{r_0}(\phi_\delta(0, u_0)) \subset V$ and $\epsilon(z) \geq r_0$ for $z \in B_{r_0}(\phi_\delta(0, u_0))$. Thus for all $z \in B_{r_0}(\phi_\delta(0, u_0)) \setminus \phi(C)$, $\|\phi_\delta(0, u_0) - z\|_0 < \text{dist}(z, \phi(C)) + \epsilon(z)$. Now choose ω so large that $\|\phi_\delta(0, u_0) - \phi_\delta(0, u)\|_0 < \frac{r_0}{4}$. Then $z \in B_{r_0}(\phi_\delta(0, u_0))$ whence $\|P(0) - z\|_0 < \epsilon(z) < \text{dist}(z, \phi(C)) + \epsilon(z)$.

As noted above, we can assume that the entire curve P lies inside the ball $B_{2\mu}(z) = B_{r_0}(z)$. Since $\epsilon(z) > r_0$, it is certainly the case that $P(0); P(1)$ lie

inside the ball of radius $\text{dist}(z, \phi(C)) + \epsilon(z)$, so P is of the class of curve tested in the definition (1.1). All of this is *uniform in ω* , so we can let $\omega \rightarrow \infty$; according to the preceding inequalities, at some point the inequality in (1.1) fails, contradicting quasiconvexity. q.e.d.

This result might appear slightly disappointing; since $\|D^2\phi_\delta\|$ is not bounded over the set C defined in the statement, one would *expect* the curvature to tend to infinity along some sequence of paths. To clear up the misapprehension that this expected behavior explains the failure of quasiconvexity, we give the

Theorem 1.1' *Let the hypotheses be the same as in Theorem 1.1, except take*

$$C = [-1, 1] \times B_1(0; H_0^2[0, 1]).$$

Then the same conclusion holds.

Proof. Follows the same outline. In constructing the path $u(\nu)$, make the alternate choices

$$u_0 \equiv 0, \quad \delta u(t) = \omega^{-2} \chi(t) \sin \omega t$$

with $\chi \in C_0^\infty(0, 1)$. The uniform estimates on r follow as before. q.e.d.

Since quasiconvexity is merely a sufficient condition for the uniqueness and stability of the projection, it remains to determine whether the impulsive least-squares problem has multiple local minima. In fact, it does:

Theorem 1.2 *Let N be any natural number, $r > 0$. There exists $R_N > 0$, $\delta > 0$ so that*

$$C = [-1, 1] \times (B_r(0; L^2[0, 1]) \cap B_{R_N}(0; H_0^2[0, 1])) \subset E_2$$

satisfies: for any neighborhood V of $\phi(C)$, there exists $z \in V$ for which

$$[s, u] \in E_2 \longmapsto \|\phi[s, u] - z\|_F$$

has N distinct local minima interior to C . Moreover, some pair $[s_i, u_i]$, $i = 1, 2$ of these minimizers satisfies $\|[s_1, u_1] - [s_2, u_2]\|_E \geq \delta$.

The proof is rather tedious direct construction which is relegated to an Appendix. This construction is unilluminating except that

- the directions in which quasiconvexity fails, as in the proof of Theorem 1.1, are also good directions in which to look for multiple minima;
- $R_N \rightarrow \infty$ as $N \rightarrow \infty$, ie. to get more local minima we must demand less regularity;
- distinct local minima exist *even for consistent (noise-free) data*.

Unfortunately, at the moment we cannot simply appeal to a converse to Chavent's theorem.

We turn now to the bandlimited problem, that is

$$\phi_\delta[s, u] = f * \phi_\delta[s, u]$$

the convolution being in t only. In fact we shall consider a family of such problems:

$$f_\alpha(t) = \frac{1}{\alpha} f_1\left(\frac{t}{\alpha}\right), \quad 0 < \alpha \leq 1.$$

We assume that $f_1 \in C_0^\infty(\mathbb{R})$, $\text{supp } f \subset \mathbb{R}^+$, and select $0 < \Omega_\ell < \Omega_h$ and $p > 0$ so that

$$|\hat{f}_1(\omega)| \geq p, \quad \Omega_\ell \leq |\omega| \leq \Omega_h.$$

Thus $[\Omega_\ell, \Omega_h]$ is the “passband” for f_1 . According to the work of Landau, Slepian, and Pollak [], if $\Omega_h - \Omega_\ell$ is sufficiently large, there exists $W \in L^2[0, 1]$ with

$$\int_{\Omega_\ell \leq |\omega| \leq \Omega_h} |\hat{W}|^2 \geq 0.8 \|W\|_0^2 .$$

Since $\hat{f}_\alpha(\omega) = \hat{f}_1(\alpha\omega)$, the passband for f_α is $[-\Omega_\ell/\alpha, \Omega_h/\alpha]$ with the same constant p , so we can assume that the passband for f_1 is sufficiently large that such W exists. *A fortiori*, for $0 \leq \alpha \leq 1$ there exists $W_\alpha \in L^2[0, 1]$ with

$$\int_{\frac{\Omega_\ell}{\alpha} \leq |\omega| \leq \frac{\Omega_h}{\alpha}} |\hat{W}_\alpha|^2 \geq 0.8 \|W_\alpha\|_0^2 .$$

Since f_α is smooth, we can regard

$$\phi_\alpha[s, u] := f_\alpha * \phi_\delta[s, u]$$

as a C^2 map:

$$\phi_\alpha : [-1, 1] \times L^2[0, 1] \rightarrow L^2([-1, 1] \times \mathbb{R})$$

i.e. without introducing additional smoothness on u .

A neighborhood V of D is *cylindrical* if $V = D + U$, where U is a neighborhood of $0 \in F$. Cylindrical neighborhoods are natural in the application to least-squares problems: they represent *uniform estimates of error*.

Theorem 1.3 *Let $C = [-1, 1] \times B_1(0) \subset [-1, 1] \times L^2[0, 1]$. Assume that $(\phi_\alpha(C), \mathbb{P})$ is quasiconvex, $\mathbb{P} = \phi \circ \Pi$, and Π has the “arbitrary tangents” property stated in Theorem 1.1. Then a cylindrical neighborhood V of $\phi_\alpha(C)$ with the property stated in the definition of quasiconvexity must satisfy*

$$V \subset \phi_\alpha(C) + B_{K\alpha}(0)$$

where K depends only on f_1 . Moreover, the function $\epsilon(z)$ in the definition must satisfy

$$\epsilon(z) \leq K\alpha, \quad z \in V.$$

Remark. Since the conclusion of Chavent's theorem is that local minima other than the projection occur at points ϕ with $\|\phi - z\|_0 \geq \text{dist}(z, \phi(C)) + \epsilon(z)$, in order to exclude "spurious" local minima it is necessary to restrict $[s, u]$ to a set for which $\|[s, u] - z\|_0 < \text{dist}(z, \phi_\alpha(C)) + \epsilon(z) \leq 2K\alpha$.

Proof. We construct the same path $\bar{u}(\nu)$ as in the proof of Theorem 1.1 (up to the choice of initial point $\pi(0) = u$). We obtain

$$\begin{aligned} \dot{P}(0) &= \frac{\partial f_\alpha}{\partial t} * u \\ \ddot{P}(0) &= 2s(0)x \frac{\partial f_\alpha}{\partial t} * u + x^2 \frac{\partial^2 f_\alpha}{\partial t^2} * u + 2f_\alpha * r(0, \cdot). \end{aligned}$$

We do not assume that ν is arc-length this time, as we have no control over higher derivatives of r . Instead, we compute the acceleration vector

$$a(0) := \frac{\ddot{P}(0)}{\|\dot{P}(0)\|_0^2} - \frac{\dot{P}(0)}{\|\dot{P}(0)\|_0} \left\langle \frac{\dot{P}(0)}{\|\dot{P}(0)\|_0}, \frac{\ddot{P}(0)}{\|\dot{P}(0)\|_0^2} \right\rangle.$$

Using

$$\begin{aligned} \|\dot{P}(0)\|_0 &= \left\| \frac{\partial f_\alpha}{\partial t} * u \right\|_0 \\ \left\langle x \frac{\partial f_\alpha}{\partial t} * u, x^2 \frac{\partial^2 f_\alpha}{\partial t^2} * u \right\rangle &= 0 \end{aligned}$$

we obtain for the acceleration vector

$$a(0) = \left\| \frac{\partial f_\alpha}{\partial t} * u \right\|_0^{-2} \left\{ \ddot{P}(0) - \frac{\dot{P}(0)}{\|\dot{P}(0)\|} \left(2s(0) \left\| \frac{\partial f_\alpha}{\partial t} * u \right\|_0^2 + 2 \langle f_\alpha * r(0, \cdot), \frac{\partial f_\alpha}{\partial t} * u \rangle_{L^2(\mathbb{R})} \right) \right\}.$$

Exactly as before we obtain

$$\|\bar{P}(0)\|_0^2 \geq \lambda^2 \left\| \frac{\partial^2 f_\alpha}{\partial t^2} * u \right\|_0$$

whence

$$\begin{aligned} \frac{1}{\bar{\rho}(0)} &= \|a(0)\| \\ &\geq \lambda \frac{\left\| \frac{\partial^2 f_\alpha}{\partial t^2} * u \right\|_0}{\left\| \frac{\partial f_\alpha}{\partial t} * u \right\|_0^2} \\ &\quad - \left(1 + \frac{\|f_\alpha * r(0, \cdot)\|_0}{\left\| \frac{\partial f_\alpha}{\partial t} * u \right\|_0^2} \right). \end{aligned}$$

Note that

$$\begin{aligned} \left\| \frac{\partial f_\alpha}{\partial t} \right\|_{L^1(\mathbb{R})} &= \alpha \left\| \frac{\partial f_1}{\partial t} \right\|_{L^1(\mathbb{R})} =: \alpha F > 0 \\ \left\| \frac{\partial f_\alpha}{\partial t} * u \right\|_0 &\geq \alpha F \|u\|_0. \end{aligned}$$

Also note that since $r(0, \cdot)$ is restricted to a bounded set in L^2 , for some $G \geq 0$

$$\|f_\alpha * r(0, \cdot)\|_0 \leq G.$$

Finally,

$$\left\| \frac{\partial^2 f_\alpha}{\partial t^2} \right\|_0^2 \geq \frac{\rho^2 \Omega_\ell^2}{\alpha^2} \int d\omega |\hat{u}(\omega)|^2, \quad \frac{\Omega_\ell}{\alpha} \leq |\omega| \leq \frac{\Omega_u}{\alpha}.$$

Choose $u = \alpha W_\alpha$ as before the statement of the theorem. We may clearly also require $\|W_\alpha\|_0 = .5$, say. Then

$$\frac{1}{\bar{\rho}(0)} \geq 2\lambda \frac{\rho \Omega_\ell}{F \alpha} - \left(1 + 2 \frac{G}{F} \right).$$

Thus $\bar{\rho}(0) \rightarrow 0$ as $\alpha \rightarrow 0$: in fact

$$\bar{\rho}(0) \leq K\alpha .$$

Once again setting

$$z_\alpha = P(0) + \mu \frac{a(0)}{\|a(0)\|}$$

we see that the inequality in (1.1) can only hold if $\mu = O(\alpha)$. The rest of the proof goes as in the proof of Theorem 1.1. q.e.d.

One use of the quasiconvexity estimates is to specify the size of the domain to which the model must be limited in order that only one local minimum of the mean-square residual be present — i.e. the size of well-posedness domains. Such an estimate follows from Theorem 1.3 *via*

Lemma 1.2 *There exists $\lambda > 0$ so that for $f \in C_0^\infty(\mathbb{R})$, $s \in [-1, 1]$, $u \in L^2[0, 1]$,*

$$\|D\phi_f[s, u][\delta s, \delta u]\|_0^2 \geq \lambda^2 \left(\delta s^2 \left\| \frac{\partial f}{\partial t} * u \right\|_0^2 + \|f * \delta u\|_0^2 \right) .$$

Proof. Note that $\phi_f[s, u] = f * \phi_s[s, u]$ and

$$D\phi_f[s, u][\delta s, \delta u] = x\delta s \frac{\partial}{\partial t} \phi_f(s, u) + \phi_f(s, \delta u) .$$

Now apply Lemma 1.1. q.e.d.

Theorem 1.4 *Suppose $\phi_\alpha(C)$ is quasiconvex when equipped with the set \mathbb{P} of images of linear segments, i.e.*

$$\mathbb{P} = \{ \nu \mapsto (s_1(1 - \nu) + s_2\nu, u_1(1 - \nu) + u_2\nu) : s \in [-1, 1], u \in B_1(0) \} .$$

Let $V_\alpha, \epsilon_\alpha$ be as in the definition of quasiconvexity, and suppose that V_α is cylindrical. Then if $(s_i, u_i), i = 1, 2$, satisfy

$$\|\phi_\alpha(s_i, u_i) - z\|_0 \leq \text{dist}(z, \phi_\alpha(C)) + \epsilon(z), \quad i = 1, 2,$$

then

$$(s_2 - s_1) \leq \frac{K\alpha}{\lambda} \left(\max \left\| \frac{\partial f_\alpha}{\partial t} * u_1 \right\|^{-1}, \left\| \frac{\partial f_\alpha}{\partial t} * u_2 \right\|^{-1} \right).$$

In particular, if $u_1 = u_2 = u_\alpha$ as in the proof of Theorem 1.3, then

$$|s_2 - s_1| < 2 \frac{K\alpha}{\lambda p}.$$

Proof. Follows directly from Lemma 1.2, Theorem 1.3 and the defining estimate for u_α . q.e.d.

Remark 1.2. Since α is a wavelength scale, a crude paraphrase of this result is: In order to ensure well-posedness with quasi-convexity, the range of s must be restricted to an interval about the optimum s of length proportional to a wavelength.

Remark 1.3. The estimate in Lemma 1.2 reveals that the linearization of ϕ_α is actually well-conditioned, uniformly of α , in the correct norms. This is most clearly seen by examining the "passband" parts of $u, \delta u$:

$$\begin{aligned} & \|D\phi_\alpha[s, u_\alpha][\delta s, \delta u]\|_0^2 \\ & \geq \lambda^2 \left(\delta s^2 \int_{\frac{\Omega_L}{\alpha} \leq |\omega| \leq \frac{\Omega_H}{\alpha}} d\omega |\omega \hat{f}_\alpha(\omega) \hat{u}_\alpha(\omega)|^2 + \int_{\frac{\Omega_L}{\alpha} \leq |\omega| \leq \frac{\Omega_H}{\alpha}} d\omega |\hat{f}_\alpha(\omega) \widehat{\delta u}(\omega)|^2 \right) \\ & \geq \lambda^2 p(0.4) \delta s^2 + \frac{1}{\Omega_u^2} \int_{\frac{\Omega_L}{\alpha} \leq |\omega| \leq \frac{\Omega_H}{\alpha}} |\widehat{\delta u}(\omega)|^2. \end{aligned}$$

On the other hand, a simple estimate gives

$$\|D\phi_\alpha[s, u_\alpha](\delta s, \delta u)\|_0^2$$

$$\begin{aligned}
&\leq \Lambda^2 (\delta s^2 + \|f_\alpha * \delta u\|^2) \\
&\leq \Lambda^2 (\delta s^2 + \|f_1\|_{L^1(\mathbb{R})}^2 \|\delta u\|_0^2) .
\end{aligned}$$

Consequently, if $u, \delta u$ are restricted to a subspace of $L^2[0, 1]$ consisting of functions obeying

$$\|\delta u\|_0 \leq \text{const} \int_{\frac{\Omega_\alpha}{\alpha} \leq |\omega| \leq \frac{\Omega_\alpha}{\alpha}} d\omega |\widehat{\delta u}(\omega)|$$

with “const.” independent of α , then the condition number of the restriction of $D\phi_\alpha$ is bounded independently of α . Such restriction is obviously necessary in any case, to insure that the “stability” part of the well-posedness definition is satisfied. Also, u_α in the proof may be chosen to be a member of such a subspace. Thus the ill-posedness evident in this nonlinear least squares problem is strictly nonlinear in nature.

In fact, replacing f with δ , one sees that $D\phi_\delta[s, u]$ is well-conditioned as an operator $E \rightarrow F$ as u ranges over any bounded set in H^1 . As is evident from the proof of Theorem 1.2, the ill-posedness then comes from the H^1 -unboundedness of $D^2\phi_\delta(u)$.

2 The Constrained Coherency Problem

We introduce a Banach space

$$\tilde{E} \subset [-1, 1] \times H^1([-1, 1] \times \mathbb{R})$$

of “expanded” models for the detection

$$\tilde{\phi}_f : \tilde{E} \rightarrow (L^2([-1, 1] \times \mathbb{R}))^2$$

given by

$$\tilde{\phi}_f[s, \tilde{u}] = \begin{pmatrix} f * \frac{\partial \tilde{u}}{\partial t} \\ \frac{\partial \tilde{u}}{\partial x} - s \frac{\partial \tilde{u}}{\partial t} \end{pmatrix}.$$

We first show the “equivalence” of the “coherency” problem

$$\tilde{\phi}_f[s, \tilde{u}] = \begin{pmatrix} z \\ 0 \end{pmatrix}$$

and the problem

$$\phi_f[s, u] = z$$

by exhibiting a 1-1 correspondence between their solutions. In fact, the coherency problem is a “cover” of the problem discussed in the last section, in a sense which we shall make precise elsewhere.

The *coherency operator* is the second member of $\tilde{\phi}_f$:

$$W[s, \tilde{u}] := \frac{\partial \tilde{u}}{\partial x} + s \frac{\partial \tilde{u}}{\partial t}.$$

Given suitable domains, $\tilde{\phi}_f$ and W are both C^2 (in fact, C^∞) maps. This smoothness is independent of the regularity of f — in fact, the regularity estimates are uniform as f ranges over bounded sets of measures — in contrast to the situation of ϕ_f .

Thus the constrained least-squares problem

$$\min_{[s, \tilde{u}]} \left\| f * \frac{\partial \tilde{u}}{\partial u} - z \right\|_0^2 \quad \text{sub } W[s, \tilde{u}] = 0$$

involves only smooth maps and has the “same” solutions as the output-least-squares problem. Since the two problems are thus equivalent, this constrained coherency problem must inherit the pathology of output least-squares, in some form. In fact we show that for smooth g , the level sets $W^{-1}(g)$ are *not* submanifolds of \tilde{E} . $W^{-1}(g)$ fails to be a submanifold for many non-smooth $g \in L^2([-1, 1] \times \mathbb{R})$ as well; we conjecture that $W^{-1}(g)$ is *never* a submanifold of \tilde{E} . If true, this conjecture would be an interesting counterpart to Sard’s theorem.

In any case, the feasible set $W^{-1}(0)$ is not a submanifold of \tilde{E} . Immediate consequences are:

- (i) The implicit function theorem does not imply stability of the solutions of the constrained coherency problem with respect to data perturbations;
- (ii) The method of Lagrange multipliers, upon which are based most modern, efficient constrained optimization algorithms, is unavailable.

The first task is to define \tilde{E} precisely. A simple choice is

$$\tilde{E} = \mathbb{R} \times \left\{ u \in H^1([-1, 1] \times [-1, 3]) : u(x, -1) \equiv 0, \right. \\ \left. -1 \leq x \leq 1 \text{ and } u(0, t) \equiv 0, -1 \leq t \leq 0 \right\} .$$

We equip \tilde{E} with the norm $\|s, u\|_{\tilde{E}} := \|u\|_1$, so that \tilde{E} is a Banach space. It is easy to verify that $\tilde{\phi}_f$, defined as above, yields a C^2 map

$$\tilde{\phi}_f : \tilde{C} \rightarrow \tilde{F} = L^2([-1, 1] \times [-1, 3])^2$$

where $\tilde{C} = \{[s, u] \in \tilde{E} : -1 \leq s \leq 1\}$.

Remark 2.1. $F = L^2([-1, 1] \times [-1, 3])$ could as well have been taken for the data space of the output-least-squares problem, without changing any of the conclusions of §1.

For technical convenience, assume that $\text{supp } f \subset [0, 1]$ as well. Then $\text{supp } \phi_f[s, u] \subset \{(x, t) : -sx \leq t \leq 2 - sx\} \subset [0, 1] \times [-1, 3]$.

Theorem 2.1

(i) Suppose that $u \in E$, $s \in [-1, 1]$ and $z = \phi_f[s, u]$. Then there exists at least one $[s, \tilde{u}] \in \tilde{C}$ for which $\tilde{\phi}_f[s, \tilde{u}] = (z, 0)^T$. Moreover, if f^* is injective on E , then \tilde{u} is unique.

(ii) Conversely, suppose that $[s, \tilde{u}] \in \tilde{C}$, $\tilde{\phi}_f[s, \tilde{u}] = (z, 0)$, $z \in \text{Range } \phi_f[s, \cdot]$. Then $z = \phi_f(s, u)$, where u is given by

$$u(t) = \frac{1}{2} \int_{-1}^1 dx \frac{\partial \tilde{u}}{\partial t}(t + sx).$$

Proof.

(i) Take for \tilde{u} the function defined by

$$\tilde{u}(x, t) = \int_0^t dt' \phi_\delta[s, u](x, t').$$

It is easily verified that $(s, \tilde{u}) \in \tilde{C}$ and $\tilde{\phi}_f[s, \tilde{u}] = (z, 0)^T$.

(ii) In particular, $W[s, \tilde{u}] = 0$, which implies that

$$\tilde{u}(x, t) = \tilde{u}(0, t - sx) \quad p.p.$$

and that the trace $\tilde{u}(0, t) \in H^1[-1, 3]$. Note that

$$u(t) = \frac{1}{2} \int_{-1}^1 dx \frac{\partial \tilde{u}}{\partial t}(t + sx) = \frac{\partial \tilde{u}}{\partial t}(0, t)$$

and that $(\phi_f[s, u], 0) = \tilde{\phi}_f s, \tilde{u}$.

Thus $\phi_f(s, u) = z$. Since $z \in \text{Range } \phi_f[s, \cdot]$, it follows that $z(x, t) = 0$ for $t < -sx$ and $t > 2 - sx$. Now extend u by zero outside of $[-1, 3]$. By assumption $z(0, t) = f * w(t)$ for some $w \in L^2[0, 1]$. Now u, w , and f have compact support, which implies that $u = w$, *p.p.*, i.e. $u = 0$ *p.p.* for $t < 0$, $t > 1$. q.e.d.

Theorem 2.2 Suppose that $g \in H_0^2([-1, 1] \times [0, 3])$. Then $W^{-1}(g)$ is not a submanifold of \tilde{E} .

Proof. We claim that $W^{-1}(g)$ is parameterized by $s \in [-1, 1]$ and by $v_0 \in H^1(-\infty, 3]$, $v_0(t) = 0$ if $t \leq 0$. Then $[s, V[s, v_0]] \in W^{-1}(g)$, where

$$V[s, v_0](x, t) = v(x, t) = v_0(t - sx) + \int_0^x dx' g(t - s(x - x'), x').$$

Indeed since $g \in H^2$, the preceding formula evidently defines a member v of $H^1([-1, 1] \times [-1, 3])$, solving $W[s, v] = g$, $v \equiv 0$ on $\{t = -1\}$ and $v(0, t) \equiv 0$, $-1 \leq t \leq 0$. Thus for each $s \in [-1, 1]$, $v_0 \in H^1(-\infty, 3]$, $\text{supp } v_0 \subset [0, 3]$, $V[s, v_0] \in W^{-1}(g)$. Conversely, if $v \in W^{-1}(g)$, the hyperbolicity of the equation

$$\left(\frac{\partial}{\partial x} + s \frac{\partial}{\partial t} \right) \frac{\partial v}{\partial t} = \frac{\partial g}{\partial t}$$

implies (via the simplest energy estimate) that

$$\frac{\partial v}{\partial t}(0, \cdot) \in L^2_{\text{loc}}, \quad \text{supp } \frac{\partial v}{\partial t} \subset [0, 3],$$

so that the trace $v(0, \cdot)$ is actually in H^1 . It follows easily that $v = V[s, v_0]$ with $v_0 = v(0, \cdot)$.

Next we attempt to construct a curve in $W^{-1}(g)$ with prescribed tangent vector $(\delta s, \delta v)$ at (s, v) . Necessarily $(\delta s, \delta v) \in \ker DW(s, v)$, i.e.

$$\delta s \frac{\partial v}{\partial t} + \left(\frac{\partial}{\partial x} + s \frac{\partial}{\partial t} \right) \delta v = 0.$$

As before we conclude that

$$\delta v(x, t) = \delta v_0(t - sx) - \delta s \int_0^x dx' \frac{\partial v}{\partial t}(t - s(x - x'), x').$$

We claim that

$$(s(\nu), v(\nu)) \equiv (s + \nu \delta s, V(s + \nu \delta s, v_0 + \nu \delta v_0))$$

will do the trick, *provided that* in addition to other requirements, $v_0 \in H^3_{\text{loc}}$, $\delta v_0 \in H^2_{\text{loc}}$. We also claim that these requirements are *necessary* if $\delta s \neq 0$. Indeed $\dot{s}(0) = \delta s$, while formally

$$\begin{aligned} \dot{v}(0)(x, t) &= \delta v_0(t - sx) - x \delta s v'_0(t - sx) \\ &\quad - \int_0^x dx' \delta s (x - x') \frac{\partial v}{\partial t}(t - s(x - x'), x'). \end{aligned}$$

Using $v = V[s, v_0]$ and the expression for δv , we find indeed that $\dot{v}(0) = \delta v$.

To show that in fact

$$\|v + \nu \delta v - v(\nu)\|_1 = 0(\nu)$$

turns out to require one extra degree of smoothness for both v_0 and δv_0 . We leave the details to the reader.

Next we claim that if $v_0 \in H_{\text{loc}}^1 \setminus H_{\text{loc}}^2$, supported as before, and $v = V[s, v_0]$, then there are *no curves* through (s, v) in $W^{-1}(g)$ tangent to any vector of the form $(\delta s, *)$ with $\delta s \neq 0$. In fact, this is already clear from the above expression for $\dot{v}(0)$: while correctly defining a solution of the p.d.e. for elements of $\ker DW[s, v]$, it gives a member of H_{loc}^1 only if $v_0 \in H_{\text{loc}}^2$, unless $\delta s = 0$.

Thus if $v = V[s, v_0]$, $v_0 \in H_{\text{loc}}^1 \setminus H_{\text{loc}}^2$, then any C^1 curves in $W^{-1}(g)$ through (s, v) can have tangent only of the form $(0, \delta v)$. Fix such v , and construct a sequence $v_0^\epsilon \in H_{\text{loc}}^3(-\infty, 3]$, $\text{supp } v_0^\epsilon \subset [0, 3]$, $v_0^\epsilon \rightarrow v_0$ as $\epsilon \rightarrow 0$ in $H^1[-1, 3]$. Evidently $V[s, v_0^\epsilon] = v^\epsilon \rightarrow v$ in $H^1([-1, 1] \times [-1, 3])$ (note: fixed s !). Pick $\delta s \neq 0$, and construct C^1 curves $(s(\nu), v_1^\epsilon(\nu))$ tangent to $(\delta s, 0)$ at (s, v^ϵ) , as before.

Now suppose $W^{-1}(g)$ were a submanifold. Then $T_{(s,v)}W^{-1}(g)$ consists only of vectors of the form $(0, u)$. Thus $\text{dist}[(\dot{s}(0), \dot{v}_1^\epsilon(0)), T_{(s,v)}W^{-1}(g)] \geq \delta s$ independently of ϵ , which would be impossible if $W^{-1}(g)$ were a submanifold, according to the following lemma:

Lemma 2.1 *Suppose that X is a C^1 manifold modeled on a Banach space E , $U \subset X$ a coordinate neighborhood, $TX|_U \cong U \times E$ a coordinate trivialization, and $M \subset U$ a C^1 -submanifold. Let $x_\epsilon \rightarrow x$ in M , $\xi_\epsilon \rightarrow \xi \in E$ so that $(x_\epsilon, \xi_\epsilon) \in T_{x_\epsilon}M \subset T_{x_\epsilon}X$. Then there exist $\eta_\epsilon \in E$ so that*

$$(i) \quad (x, \eta_\epsilon) \in T_x M$$

$$(ii) \quad \|\xi_\epsilon - \eta_\epsilon\|_E \rightarrow 0, \text{ i.e. } \text{dist}[(x_\epsilon, \xi_\epsilon), T_x M] \rightarrow 0.$$

Proof. $M \subset U$ is a C^1 submanifold if and only if there exist closed subspaces $E_1, E_2 \subset E$, $E = E_1 \oplus E_2$ and C^1 diffeomorphism $\Phi : U \rightarrow E$ so that $\Phi(U)$ is a neighborhood of the origin, and

$$M = \Phi^{-1}(\{0\} \oplus E_2) \cap U.$$

Thus $T_x M = D\Phi(x)^{-1}\{\{0\} \oplus E_2\}$, and there exist $\zeta_\epsilon \in E_2$ so that $\xi_\epsilon = D\Phi(x_\epsilon)^{-1}(0, \zeta_\epsilon)$.

Set $\eta_\epsilon = D\Phi(x)^{-1}(0, \zeta_\epsilon)$. Since Φ is C^1 , $D\Phi(x_\epsilon)^{-1} \rightarrow D\Phi(x)^{-1}$ uniformly, and $D\Phi(x)$ is an isomorphism, so

$$\begin{aligned} \eta_\epsilon &= D\Phi^{-1}(x)D\Phi(x_\epsilon)\xi_\epsilon \\ &= D\Phi(x)^{-1}D\Phi(x_\epsilon)(\xi_\epsilon - \xi) \\ &\quad + (D\Phi(x)^{-1}D\Phi(x_\epsilon) - I)\xi + \xi \\ &\rightarrow \xi \qquad \qquad \qquad \text{q.e.d.} \end{aligned}$$

Since $\dot{v}_1^\epsilon(0) = 0$, $\dot{s}(0) = \delta s$ for all ϵ , we can replace $\xi_\epsilon = \xi \rightarrow (\delta s, 0)$, $x_\epsilon \rightarrow (s, v^\epsilon)$, $x \rightarrow (s, v)$. No lower bound like the one stated before the lemma would be possible if $W^{-1}(g)$ were a submanifold of \tilde{E} — so it isn't. q.e.d.

Remark 2.2. The feasible sets of smooth constrained optimization become submanifolds because of the so-called *constraint qualification*. See e.g. Luenberger [1973]. This condition, involving the Lagrange multipliers, has as its sole function in the theory to ensure that the feasible set and its perturbations are a submanifold — though this point is generally unmentioned in the literature on such things. We have shown elsewhere that the constraint qual-

ification fails even for fixed s , in the detection problem (Symes [1988]). It is amusing, and a little deeper, that $W^{-1}(g)$ actually fails to be a submanifold.

Remark 2.3. Since $W^{-1}(g)$ fails to be a submanifold on such a fat, dense subspace of L^2 , it is natural to ask whether $W^{-1}(g)$ is *ever* a submanifold. It is rather easy to construct $g \in \text{Range } W$ for which only vectors of the form $(0, *)$ are tangent, but our examples so far always have accumulation points along the s -direction, so are not submanifolds. Conceivably for suitable g , there exists an *isolated* s for which

$$\frac{\partial v}{\partial x} + s \frac{\partial v}{\partial t} = g$$

has H^1 solutions. Then $W^{-1}(g)$ would simply be the affine space of such solutions, parameterized by boundary values, hence a local submanifold. Whether this can occur seems to be rather delicate, and we have been unable to resolve the point. It concerns the *stability of the range* for first order hyperbolic p.d.e.s acting on Sobolev spaces. A definite “no” would provide a nice counterpart to Sard’s theorem, i.e. a nontrivial example of a smooth map with no regular points.

Remark 2.4. It is easy to produce examples in the spirit of Section 1, of a dense set of points in $W^{-1}(0)$ and sequences of directions in which the curvature of $W^{-1}(0)$ increase without bound. Thus $W^{-1}(0)$ has a “quasi-cusp” at a dense set of points — showing once again that the pathological features of $W^{-1}(0)$ are truly “infinite dimensional”!

3 The Penalized Form of the Coherency Problem

Note: In this section we will work with the impulsive case ($f = \delta$) only.

Since the coherency reformulation of the detection problem cannot be treated as a constrained problem, one might try a least-squares approach instead; that is, introduce a norm in the augmented data space \tilde{F} , and minimize

$$\tilde{J}[s, \tilde{u}; \tilde{z}] = \left\| \tilde{\phi}_\delta[s, \tilde{u}] - \tilde{z} \right\|_{\tilde{F}}^2 .$$

At the very least, for consistent data $\tilde{z} = (z, 0)^T$, and $z \in \text{Range}(\phi_\delta)$, then $\tilde{J} = 0$ is a global minimum value, which is achieved at the model indicated in Theorem 2.1. That is, for consistent data, the coherency least-squares problem has the "same" solution as the original problem.

It remains to verify that, for some choice of norm in \tilde{F} , neighborhood of a model corresponding to consistent data, and data noise level, the reformulated problem is well-posed. Our principal tool will be the following simple consequence of Chavent's theory.

Theorem 3.1 *Suppose that E is a Banach Space, F a Hilbert space, and $C \subset E$ a closed convex set. Let $\phi : C \rightarrow F$ be a C^2 map, and set*

$$\begin{aligned} L(x) &= \inf_{\delta x \in E \setminus \{0\}} \left(\frac{\|\phi'(x)\delta x\|_F}{\|\delta x\|_E} \right) \\ M &= \sup_{x \in C} \|\phi'(x)\|_{E,F} \\ S &= \sup_{x \in C} \|\phi''(x)\|_{E,E,F} \end{aligned}$$

$$\kappa(x) = \frac{2\sqrt{2}M}{L(x)}$$

$$\alpha(x) = \frac{L(x)}{S}$$

Suppose that $x_0 \in C$, $z \in F$ and

$$\|\phi(x_0) - z\|_F < \epsilon(x_0) := L(x_0)\alpha(x_0)p^*(1-p^*)$$

where $p^* = p^*(\kappa(x_0))$ is the solution of

$$(1-p)^2 - \frac{\kappa^2}{8} p^2(1-p)^{-2} = p(1-p)$$

with $0 \leq p \leq 1/2$. Then

(i) for $R = p^*\alpha(x_0)$, the least squares problem

$$\min_{x \in B_R(x_0) \cap C} \|\phi(x) - z\|_F^2$$

has a unique global minimizer $x^*(z)$, which lies interior to $B_R(x_0)$.

All other local minimizers have residuals $\geq \epsilon(x_0)$.

(ii) for $x_1, x_2 \in B_{R(x_0)}(x_0)$,

$$\|x_1 - x_2\|_E \leq \frac{2}{L(x_0)} \|\phi(x_1) - \phi(x_2)\|$$

(iii) for $z_1, z_2 \in \phi(B_{R(x_0)}(x_0)) + B_{\epsilon(x_0)}(0)$, set $d = \epsilon(x_0) - \max_{i=1,2} \text{dist}(z_i, \phi(B_{R(x_0)}))$.

If $|z_1 - z_2| < d$, then

$$\|x^*(z_1) - x^*(z_2)\|_E \leq \frac{2\epsilon(x_0)}{dL(x_0)} \|z_1 - z_2\|_F$$

Remark 3.1. That is, coercivity at a point of the derivative of a C^2 map implies local well-posedness of the corresponding nonlinear least-squares problem for near-consistent data. We will derive this conclusion from Chavent's quasiconvexity estimates. While the basic qualitative conclusion follows immediately from the implicit function theorem, Chavent's estimates seem to yield quantitatively correct results as well, which will be useful in the sequel.

The various constants depend on the lower bound for the derivative, the ratio of this lower bound to an upper bound for the second derivative (measuring nonlinearity) and the condition number of the derivative (essentially measuring linear well-posedness). As $S \mapsto 0$ and ϕ becomes linear, the stability estimates (ii) and (iii) reduce to well-known absolute stability results for linear least-squares (Golub and Van Loan, 1983).

Proof. For $R > 0$, set $\Pi_R = \{\text{linear segments in } B_R(x_0) \cap C\}$, $\mathbb{P}_R = \phi(\Pi_R)$. For $x \in B_R(x_0) \cap C$, $\delta x \in E$,

$$\begin{aligned} \|\phi'(x)\delta x\|_F &= \left\| \phi'(x_0)\delta x + \int_0^1 d\nu \phi''((1-\nu)x_0 + \nu x)(\delta x, x - x_0) \right\|_F \\ &\geq L \|\delta x\|_E - S \|\delta x\|_E \|x - x_0\|_E \\ &\geq (L - SR) \|\delta x\|_E, \end{aligned}$$

i.e. for $R \leq \frac{L}{S}$, $\phi'(x)$ is uniformly coercive over $B_R(x_0)$ (here $L = L(x_0)$).

Accordingly, for $\pi \in \Pi_R$,

$$\pi(\nu) = x + \nu \delta x, \quad \|\delta x\|_E \leq 2R$$

we get ($P = \phi \circ \pi$)

$$\begin{aligned} \|\dot{P}(\nu)\|_F &= \|\phi'(x + \nu \delta x) \cdot \delta x\|_F \\ &\geq (L - SR) \|\delta x\|_E \end{aligned}$$

while

$$\begin{aligned}\delta(P) &= \int_0^1 d\nu \|\dot{P}(\nu)\|_E \leq M \|\delta x\|_E \\ &\leq 2FR.\end{aligned}$$

For the acceleration vector $a(\nu)$,

$$\begin{aligned}\|a(\nu)\|_F &= \left\| \ddot{P}(\nu) - \frac{\dot{P}(\nu)}{\|\dot{P}(\nu)\|_F} \left\langle \ddot{P}(\nu), \frac{\dot{P}(\nu)}{\|\dot{P}(\nu)\|_F} \right\rangle \right\|_F \\ &\leq \|\ddot{P}(\nu)\|_F \leq S \|\delta x\|_E^2.\end{aligned}$$

For the radius of curvature $\bar{\rho}(\nu)$ we obtain

$$\frac{1}{\bar{\rho}(\nu)} = \frac{\|a(\nu)\|_F}{\|\dot{P}(\nu)\|_F^2} \leq \frac{S}{(L - SR)^2}$$

whence a lower estimate for the size-curvature ratio $\gamma(P)$ is

$$\begin{aligned}\gamma(P) &\geq \inf_{\nu \in [0,1]} \bar{\rho}(\nu) - \frac{\delta(P)^2}{8 \inf_{\nu \in [0,1]} \bar{\rho}(\nu)} \\ &\geq \frac{(L - SR)^2}{S} - \frac{SM^2R^2}{2(L - SR)^2} =: \gamma_0(R).\end{aligned}$$

On the other hand,

$$\begin{aligned}\phi(x) - \phi(x_0) &= \int_0^1 d\nu \phi'((1 - \nu)x_0 + \nu x)(x - x_0) \\ &= \phi'(x_0)(x - x_0) + \int_0^1 d\mu \int_0^1 d\nu \phi''((1 - \mu)x_0 \\ &\quad + \mu((1 - \nu)x_0 + \nu x))(x - x_0, x - x_0)\end{aligned}$$

so

$$\|\phi(x) - \phi(x_0)\|_F \geq L \|x - x_0\|_E - S \|x - x_0\|_E^2.$$

If $\|x - x_0\|_E = R$, then

$$\|\phi(x) - \phi(x_0)\|_F \geq (L - SR)R.$$

The function of R on the r.h.s. attains its maximum at $R = L/2S = 1/2 \alpha$.
If $0 < p < 1/2$ and

$$\|x - x_0\| = p\alpha \quad \text{then} \quad \|\phi(x) - \phi(x_0)\| \geq p(1-p)L\alpha.$$

Accordingly, if $R \leq p\alpha$ and

$$\|\phi(x_0) - z\| < p(1-p)L\alpha$$

then necessarily

$$x^*(z) = \operatorname{argmin}_{x \in B_R(x_0)} \|\phi(x) - z\|,$$

if it exists, lies in the interior of $B_R(x_0)$.

Insert the expression for R into the estimate for γ to get

$$\gamma(P) \geq \left[(1-p)^2 - \frac{\kappa^2 p^2}{8} (1-p)^{-2} \right] L\alpha$$

for any $P \in \mathbb{P}$. It is elementary to see that the equation

$$(1-p^2) - \frac{\kappa^2 p^2}{8} (1-p)^{-2} = p(1-p)$$

has a unique root $p^*(\kappa)$ in $[0, 1/2]$ for $H \geq 1$, with $p^* \rightarrow 0$ monotonically as $\kappa \rightarrow \infty$; in fact $p^* = 0(1/\kappa)$.

The conclusions now follow immediately from Chavent, Corollary 4.19.

q.e.d.

We will apply this result to the family of least-squares problems

$$\min \|\tilde{\phi}_{\delta, \sigma}[s, \tilde{u}] - z\|_{\tilde{F}}^2$$

where

$$\tilde{\phi}_{\delta,\sigma}[s, \tilde{u}] = \begin{pmatrix} \frac{\partial \tilde{u}}{\partial t} \\ \sigma W[s, \tilde{u}] \end{pmatrix}$$

and $\tilde{F} = F \oplus F$. So the objective function is

$$\left\| \frac{\partial \tilde{u}}{\partial t} - z \right\|_F^2 + \sigma^2 \|W[s, \tilde{u}]\|_F^2$$

and our problem has the appearance of a penalized least-squares problem. We shall show how a proper choice of the penalty parameter σ emerges from Chavent's theory.

The maps $\tilde{\phi}_{\delta,\sigma}$ are defined on the closed, convex subset \tilde{C} of the closed subspace

$$\tilde{E} \subset \mathbb{R} \oplus H_b^1([-1, 1] \times [-1, 3])$$

as in the preceding section. We shall define a σ -dependent family of norms on \tilde{E} . Before doing so, we present the crux of the proof that the derivative $\tilde{\phi}'_{\delta,\sigma}$ is coercive, at least sometimes.

Theorem 3.2 *Suppose $(s_0, \tilde{u}_0) \in \tilde{C}$ and $W[s_0, \tilde{u}_0] = 0$. For a constant K_0 independent of (s_0, \tilde{u}_0) , any $(\delta s, \delta \tilde{u}) \in \tilde{E}$*

$$\left\| \frac{\partial \tilde{u}_0}{\partial t} \right\|_0^2 |\delta s|^2 \leq K_0 \left(\left\| \frac{\partial \delta \tilde{u}}{\partial t} \right\|_0^2 + \|DW[s_0, \tilde{u}_0][\delta s, \delta \tilde{u}]\|_0^2 \right).$$

Proof. Set $D = DW[s_0, \tilde{u}_0][\delta s, \delta \tilde{u}]$. Thus

$$\frac{\partial \delta \tilde{u}}{\partial x} + s_0 \frac{\partial \delta \tilde{u}}{\partial t} = D - \delta s \frac{\partial \tilde{u}_0}{\partial t}.$$

For $(x_1, t_1) \in [-1, 1] \times [-1, 3]$, and x_2 satisfying

$$-1 - t_1 + s_0 x_1 \leq s_0 x_2 \leq 3 - t_1 + s_0 x_1$$

integrate both sides along the line segment

$$x \in [x_1, x_2] \mapsto (x, t_1 + s_0(x - x_1))$$

to get

$$\begin{aligned} & \delta \tilde{u}(x_2, t_1 + s_0(x_2 - x_1)) - \delta \tilde{u}(x_1, t_1) = \\ & \int_{x_1}^{x_2} dx D[x, t_1 + s_0(x - x_1)] + \delta s(x_2 - x_1) \frac{\partial \tilde{u}_0}{\partial t}(x_1, t_1) \end{aligned}$$

where we have made use of $W[s_0, \tilde{u}_0] = 0$. Thus

$$\begin{aligned} & |\delta s|^2 (x_2 - x_1)^2 \left| \frac{\partial \tilde{u}_0}{\partial t}(x_1, t_1) \right|^2 \leq \\ & 4 \left[\int_{x_1}^{x_2} dx (D(x_1 t_1 + s_0(x - x_1)))^2 + \int_{-1}^{t_1} dt \left[\frac{\partial \delta \tilde{u}}{\partial t} \right]^2(x_1, t) \right. \\ & \left. + \int_1^{t_1 + s_0(x_2 - x_1)} dt \left[\frac{\partial \delta \tilde{u}}{\partial t} \right]^2(x_2, t) \right] \end{aligned}$$

where we have used the boundary condition $\delta \tilde{u}(x, -1) \equiv 0, -1, \leq x \leq 1$.

Now integrate both sides in x_1, x_2 , and t_1 and do a little algebra to get

$$|\delta s|^2 \left\| \frac{\partial \tilde{u}_0}{\partial t} \right\|_0^2 \leq K_0 \left(\left\| \frac{\partial \delta \tilde{u}}{\partial t} \right\|_0^2 + \|D\|_0^2 \right)$$

for a suitable $K_0 > 0$.

q.e.d.

Theorem 3.3 Define the norm $\| \cdot \|_{\tilde{E}, \sigma}$ on \tilde{E} by

$$\|[s, \tilde{u}]\|_{\tilde{E}, \sigma}^2 = \left(\left\| \frac{\partial \tilde{u}}{\partial t} \right\|^2 + \frac{\sigma}{\sqrt{1 + \sigma^2}} \left(|s|^2 + \left\| \frac{\partial \tilde{u}}{\partial x} \right\|^2 \right) \right)$$

Appendix. Proof of Theorem 1.2

We will establish:

There exist consistent data $z = \phi[s_0, u_0]$ with $\|[s_0, u_0]\|_{E_2} \leq 1$ and $\delta > 0$ for which the problem

$$\min_{[s, u] \in E_2} \|\phi(x) - z\|_F$$

has (local) solutions satisfying $|s| < 1$, $\|u\|_{H^2[0,1]} < 1$, $|s - s_0| + \|u - u_0\| \geq \delta$

Thus restricting the H^2 -size of the solution does not restore well-posedness to the best-fit version of the detection problem, even for noise-free data!

Set $z(\xi, t) = u_0(t) = a\chi(t) \sin \omega t$ with ω and a to be determined, and $\chi \in C_0^\infty(0, 1)$ fixed. Then

$$z = \phi[0, u_0].$$

For $u \in L^2[0, 1]$, $s \in [-1, 1]$,

$$\begin{aligned} \langle z, \phi[s, u] \rangle &= a \int_{-1}^1 d\xi \int dt u(t - s\xi) \chi(t) \sin \omega t \\ &= a \int_{-1}^1 d\xi \int dt u(t) \chi(t + s\xi) (\sin \omega t \cos \omega s\xi + \cos \omega t \sin \omega s\xi) \\ &= 2a \frac{\sin \omega s}{\omega s} \int dt u(t) \chi(t) \sin \omega t \\ &\quad + \int_{-1}^1 d\xi \int dt u(t) (\chi(t + s\xi) - \chi(t)) [\sin \omega t \cos \omega s\xi + \omega \cos \omega t \sin \omega s\xi]. \end{aligned}$$

There is a uniform estimate for derivatives of χ of order ≤ 2 :

$$|\chi^{(k)}(t + s\xi) - \chi^{(k)}(t)| \leq C|s|$$

for $s \in [-1, 1]$, $\xi \in [-1, 1]$, $t \in \mathbb{R}$ and $k \leq 2$. Accordingly an integration-by-parts argument shows that the second term is bounded in absolute value by

$$a\omega^{-2}C|s| \|u\|_{H^2[0,1]}.$$

Thus

$$\langle z, \phi[s, u] \rangle_F = 2 \frac{\sin \omega s}{\omega s} \langle u, u_0 \rangle_{L^2[0,1]} + O(\omega^{-2} a |s| \|u\|_{H^2[0,1]}).$$

So

$$\begin{aligned} \|\phi[s, u] - z\|_F^2 &= \|\phi[s, u]\|_F^2 + \|z\|_F^2 - 2\langle \phi[s, u], z \rangle_F \\ &\geq 2 \left[\|u_0\|_{L^2[0,1]}^2 + \|u\|_{L^2[0,1]}^2 - 2 \frac{\sin \omega s}{\omega s} \langle u, u_0 \rangle_{L^2[0,1]} \right] \\ &\quad - C\omega^{-2} a |s| \|u\|_{H^2[0,1]} \\ &\geq s \left[\left(1 - \left| \frac{\sin \omega s}{\omega s} \right| \right) (\|u\|_{L^2[0,1]}^2 + \|u_0\|_{L^2[0,1]}^2) \right] \\ &\quad - C\omega^{-2} a |s| \|u\|_{H^2[0,1]} \end{aligned}$$

Integration-by-parts shows that there exist $C_1, C_2 > 0$ so that

$$a^2 \left(C_1 + \frac{C_2}{\omega} \right) \geq \|u_0\|_{L^2[0,1]}^2 \geq a^2 \left(C_1 - \frac{C_2}{\omega} \right).$$

The hypotheses of the statement to be proved allow us to require that

$\|u\|_{H^2[0,1]} \leq 1$. Thus for $\bar{s} = \frac{\pi}{\omega}$ we get

$$\begin{aligned} \|\phi[\bar{s}, u] - z\|_F &\geq 2a^2 \left(C_1 - \frac{C_2}{\omega} \right) - Ca\pi\omega^{-3} \\ &\geq a^2 C_1 (2 - \epsilon) \end{aligned}$$

provided that

$$C_1 \epsilon \geq \frac{2C_2}{\omega} + \frac{C\pi}{a\omega^3}.$$

Now $\|u_0\|_{H^2[0,1]} \leq 1/2$, provided $|\omega| \geq 1$ and $a\omega^2 \leq K$ for suitable $K > 0$.

Thus we take $a = K\omega^{-2}$ so that the above condition becomes

$$C_1 \epsilon \geq (2C_2 + C\pi K^{-1})\omega^{-1}$$

and is satisfied for any choice of $\epsilon > 0$ as soon as ω is large enough.

Now consider the special choice $u_1 = \alpha u_0$. Then

$$\|\phi[s, u_1] - z\| \leq 2 \left[1 + \alpha^2 - 2\alpha \frac{\sin \omega s}{\omega s} \right] \|u_0\|_{L^2[0,1]}^2 + Ca\omega^{-2} |s| \alpha \|u_0\|_{H^2[0,1]}.$$

Choose $s_1 = \frac{5\pi}{2\omega}$; then the above is

$$\begin{aligned} \|\phi[s_1, u_1] - z\| &\leq 2 \left[1 + \alpha^2 - \frac{4\alpha}{5\pi} \right] \|u_0\|_{L^2[0,1]}^2 + 3Ca\alpha\pi\omega^{-3} \|u_0\|_{H^2[0,1]} \\ &\leq 2a^2 \left[1 + \alpha^2 - \frac{4\alpha}{5\pi} \right] \left(C_1 + \frac{C_2}{\omega} \right) + 3aCa\pi\omega^{-3}. \end{aligned}$$

Choose $\alpha = \frac{2}{5\pi}$; we get

$$\begin{aligned}
&= 2a^2 \left[1 - \frac{4}{(5\pi)^2} \right] \left(C_1 + \frac{C_2}{\omega} \right) + a \frac{6C}{5\omega^3} \\
&= a^2 C_1 \left[2 \left(1 - \frac{4}{(5\pi)^2} \right) + 2 \left(1 - \frac{4}{5\pi^2} \right) \frac{C_2}{C_1 \omega} + \frac{6}{5} \frac{C}{C_1 K \omega} \right] \\
&\leq a^2 C_1 \left(2 - \frac{4}{(5\pi)^2} \right)
\end{aligned}$$

for ω large enough.

Now choose $\epsilon = (5\pi)^{-2}$. Then we have shown that, for ω sufficiently large

(i) for any u with $\|u\|_{H^2[0,1]} \leq 1$, $\bar{s} = \frac{\pi}{\omega}$:

$$\|\phi[\bar{s}, u] - z\|_F \geq a^2 C_1 (2 - \epsilon)$$

(ii) for $s_1 = \frac{5\pi}{2\omega}$, $u_1 = \frac{2}{5\pi} u_0$:

$$\|\phi[s_1, u_1] - z\|_F \leq a^2 C_1 (2 - 2\epsilon).$$

Since any continuous path from $[s_1, u_1]$ to $[0, u_0]$ must pass over the set $\{[s, u] : s = \frac{\pi}{\omega}\}$, we have shown that the set

$$\{[s, u] : \|\phi[s, u] - z\|_F \leq a^2 C_1 (2 - 3/2\epsilon), \|u\|_{H^2} \leq 1\}$$

is *not connected*; in particular the component containing $[s_1, u_1]$ is disjoint from the component containing $[0, u_0]$. The connected component of $[0, u_0]$ is contained in

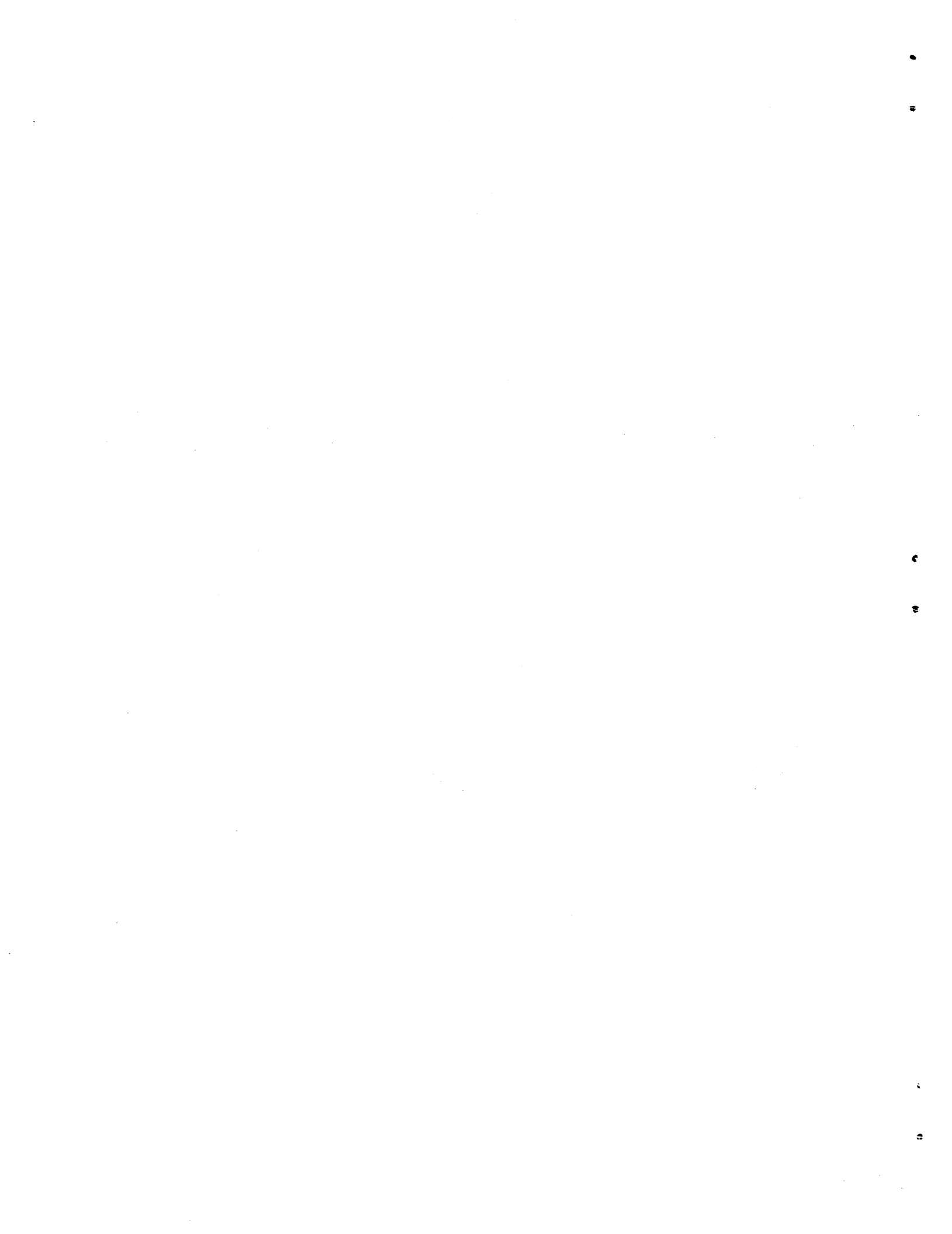
$$C_0 = \left\{ [s, u] : |s| \leq \frac{\pi}{\omega} - \delta, \|u\|_{H^2[0,1]} \leq 1 \right\}$$

for a suitable choice of $\delta > 0$, and the connected component of $[s_1, u_1]$ is contained in

$$C_1 = \left\{ [s, u] : \frac{\pi}{\omega} + \delta \leq s \leq 1, \|u\|_{H^2[0,1]} \leq 1 \right\}$$

which follows from the uniform continuity of ϕ on $[-1, 1] \times H^1[0, 1]$ and the compactness of the injection $H^2 \rightarrow H^1$. The sets C_0 and C_1 are closed, bounded, and convex in E_2 , hence weakly closed, whence follows the existence of a local minimizer in each. Clearly $[0, u_0]$ is a minimizer over all of E_2 . Since $\alpha \leq 1$, it is also the case that $[s_1, u_1]$ is interior to C_1 . Thus $[s, u_1]$ is also a local minimizer in E_2 . In particular, we have established the existence of a local minimum distinct from $[0, u_0]$, as required. q.e.d.

It is easy to extend this reasoning to generate examples with any number of local minima whatsoever. Thus even the restriction of ϕ to a ball in E_2 does not suffice to render the output least-squares problem well-posed in the nonlinear sense.



References

1. CHAVENT, G. [1988]. Quasiconvex sets and size * curvature condition, application to nonlinear inversion, Rapport *INRIA* No. 1017.
2. LUENBERGER, D.G. [1973]. Introduction to Linear and Nonlinear Programming, Addison-Wesley, Reading, Massachusetts.
3. SYMES, W. [1988]. Velocity inversion: a model problem from reflection seismology, Technical Report 88-13, Department of Mathematical Sciences, Rice University, Houston, TX 77251.
4. SYMES, W. and J. CARAZZONE [1989]. Velocity inversion by coherency optimization, Technical Report 89-9, Department of Mathematical Sciences, Rice University, Houston TX 77251.
5. TARANTOLA, A. [1987]. Inverse Problem Theory, Elsevier, New York.

