

The Plane-Wave Detection Problem

William W. Symes

August 1992  
(revised April 1993)

TR92-22



# The Plane-Wave Detection Problem

William W. Symes

Department of Computational and Applied Mathematics

Rice University

Houston, TX 77005

U.S.A

## Abstract

The plane-wave detection problem is: to estimate the incidence angle and waveform of a transient plane traveling wave, from samples recorded at a linear array of receivers. This simple problem shares several important mathematical features with other inverse problems of wave propagation, and is of interest in its own right as a model problem in ocean acoustic signal analysis. Straightforward formulation as a nonlinear least squares problem yields a nonconvex objective for which the minima are not stably dependent on the data. In contrast, an infeasible point formulation, in which the signal at each receiver is explained to some extent independently, proves to yield a smooth convex optimization problem with stable optima. Numerical experiments illustrate the theoretical results about the infeasible point approach, differential semblance optimization.



**Acknowledgement:** I would like to express my appreciation to Hua Song, who contributed to the work reported here in several ways. In particular some of the examples at the end of Section 1 and the “two step” intermediate approach at the beginning of Section 2 are due to him. This work was partially supported by the National Science Foundation (DMS8905878), The Office of Naval Research (N00014-89-J-1115), the Defense Advanced Research Projects Agency and the Air Force Office of Scientific Research (AFOSR 90-0334), the Texas Geophysical Parallel Computation Project, and the Rice Inversion Project. TRIP Sponsors for 1992 are Amoco Research, Conoco, Inc., Cray Research, Earth Modeling Systems, Exxon Production Research Co., and Mobil Research and Development Corp.



## 0 Introduction

This paper presents a detailed study of a simple but nontrivial nonlinear least-squares problem in infinite dimensions. 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 ~~first~~ author (Symes [1991a], [1991c], Symes and Carazzone [1991], Symes [1993]). Suppose that a scalar field in three-dimensional space-time is sampled at every point in the interval  $[-X, X]$  on the  $x$ -axis:

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

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  $\theta$  (or equivalently its sine  $s = \sin \theta$ ), given the *data*  $z(x, t)$ .

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

$$\phi[s, u] = u(t + sx) .$$

In this notation, one attempts to find  $[s, u]$  to solve the *output least-squares* problem

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

where  $\|\cdot\|$  denotes the  $L^2$  norm on a suitable domain. This formulation may be attacked numerically after suitable discretization. Note that  $\phi$  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, applied to some (constrained or unconstrained) optimization problem.

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 *differential semblance optimization 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 Tarantola [1987]). 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.

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 signal is close to that which would be produced by a plane-wave, and ask that the direction sine and wave form be estimated with comparable accuracy.

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 in terms of subjective probabilities, or Bayesian statistics (see especially Tarantola [1987], Menke [1984]). We shall not discuss this “standard inverse theory” here, or otherwise offer much serious argument about the epistemology of inverse problems. Our goal here is modest and mathematical: we want merely to show that some formulations of our simple model problem have properties (i)–(iii) above, whereas others do not.

A part of condition (iii) is the presumption that  $S$  is defined by the solution of an optimization problem. The rest of (iii) demands that this optimization problem, which may incorporate model constraints, be solved by local mathematical programming, i.e., methods that find local minima. Therefore we must regard the collection of all local solutions of the defining optimization problem as the “value” of  $S$ : that is,  $S$  is multi-valued when the defining optimization problem possesses multiple local minima. In this

case the stability property (ii) must be reinterpreted. The largest possible error in the estimated model is the radius of the smallest ball containing all values of  $S$ , i.e., all local optima. So the stability property (ii) must mean that the radius of this ball is proportional to the data error. In particular, for noise-free (consistent) data, there must exist exactly one local (hence global) minimizer (as is also implied by (i) and (iii) together).

Condition (iii) is motivated by the role of the detection 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  $\phi$  (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.

Our first main result (Section 1) is that, for a large class of obvious choices of metric in the model space  $\{m = [s, u]\}$ , the output least squares problem stated above does not define an estimator  $S$  for the detection problem with the properties (i)–(iii) above. We choose the  $L^2$  topology for the range of  $\phi$ , as is implied in the setting of the least squares problem, and a subspace of  $\mathbb{R} \times H^k[0, T]$ ,  $k \geq 0$ , as the domain. We show that, depending on domain metric (i.e. on  $s$ ), either the objective function fails to be differentiable, or its Hessian fails to be coercive. Thus the standard sufficient conditions for superlinear convergence of Newton’s method are violated, i.e. condition (iii) appears to be violated (Dennis and Schnabel [1983], Ch. 5). Moreover, we show by explicit example that the least squares functional can have many local minima even for consistent data, so that condition (ii) cannot possibly hold.

Our second main result (Section 2) is that a reformulation of the output least squares, involving enlargement of the model space and introduction of new constraints, defines an estimator with properties (i)–(iii) above. To motivate this reformulation, note that the output least-squares formulation demands that the signals  $z(x, \cdot)$  be fit simultaneously, for many values of  $x$ ,

by the predicted signal  $u(t + sx)$  based on a single input waveform  $u(t)$ . If  $z$  is at all oscillatory, a good fit is only possible (if at all) when the slope  $s$  is very nearly correct. On the other hand, fitting the data for a single value of  $x$  is easy — in fact, any  $s$  and  $u(t) = z(x, t - sx)$  will do the job! Equally well, it is easy to fit  $z(x, t)$  with the  $x$ -dependent waveform  $u(x, t)$  regardless of the slope  $s$ . In allowing the waveform  $u$  to depend on  $x$  we have lost the “physics” of the problem, of course: that is, the assumption that the field sampled to produce  $\phi[s, u]$  is a *plane wave*. We reintroduce this assumption in a measured way by means of the objective function (of  $s$  and the  $x$ -dependent waveform  $u(x, t)$ ):

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

where  $\| \cdot \|$  is the  $L^2$  norm and  $u$  ranges over a suitable subspace of  $H^1([-X, X] \times [0, T])$  (precise choices are identified in Section 2.) We claim that this objective function yields an estimator satisfying the criteria (i)–(iii), for proper choice of  $\sigma > 0$  and minimization strategy.

We draw the reader’s attention to the role of the parameter  $\sigma$ . It is a penalty parameter, enforcing a problem constraint in a soft way. It is *not* a regularization parameter, and does not exhibit the “ $L$ -curve” influence on the solution typical of regularization weights. Some information on the dependence of the solution on  $\sigma$  is given in Section 2, but much remains to be explained.

Note that we have implicitly changed variables by introducing the  $t$ -derivative in the data misfit term. Specifically, if  $(s, u)$  is a minimizer of the above functional, then

$$s, t \mapsto \frac{1}{2X} \int_{-X}^X dx \frac{\partial u}{\partial t}(x, t)$$

is an estimator for the plane-wave detection problem. Note that if  $z(x, t) = z_*(x, t) := u_*(x + s_*t)$  is consistent data for  $u_* \in L^2_{\text{loc}}$ , then

$$s = s_*, u(x, t) = \int_0^t dt' u_*(t')$$

minimizes the above functional — in fact, its value is  $\equiv 0$ . Moreover, the estimator just defined gives  $(s_*, u_*)$  as the estimated solution of the detection

problem. Thus the new formulation is consistent — that is, it has property (i) above.

The demonstration that this estimator possesses properties (ii) and (iii) is more involved. For a variety of subspaces  $H \subset H^1([-X, X] \times [0, T])$ , the problem

$$\min_{u \in H} \left( \left\| \frac{\partial}{\partial t} \phi[s, u] - z \right\|^2 + \sigma^2 \left\| \frac{\partial u}{\partial x} \right\|^2 \right)$$

has a unique solution  $u[s]$  depending continuously on the data  $z$ . Set

$$J_\sigma[s; z] = \frac{1}{2} \left( \left\| \frac{\partial}{\partial t} \phi[s, u[s]] - z \right\|^2 + \sigma^2 \left\| \frac{\partial u[s]}{\partial x} \right\|^2 \right).$$

In Section 2 we show that  $J_\sigma$  is smooth over  $s \in (-1, 1)$  so that its local minima may be found efficiently by variants of Newton's method. Next we investigate the set of critical points of  $J_\sigma$ . Let  $e \geq 0$  be the *error level* of the datum  $z$  relative to a model  $[s^*, u^*]$ , i.e.

$$e = \left\{ \frac{\|z - \phi[s^*, u^*]\|}{\|z\|} \right\}.$$

We show that the diameter of the set of critical points of  $J_\sigma$  is proportional to  $e/\sigma$ , provided that  $\sigma$  is sufficiently small and the “target signal”  $u^*$  is *sufficiently oscillatory*. These observations (Theorems 2.9, 2.10) establish properties (ii) and (iii) for the estimator constructed by local minimization of  $J_\sigma$ , applied to oscillatory data. In particular, the solution map

$$z \mapsto \left( s[z], t \mapsto \frac{1}{2X} \int_{-X}^X dx \frac{\partial u[s[z]]}{\partial t}(x, t) \right)$$

can be computed reliably by smooth optimization methods for low-noise data sets, and is Lipschitz continuous (in the sense of set-valued maps).  $J_\sigma$  has a unique critical point, i.e. the global minimizer, if  $e = 0$ : for consistent (noise-free) data,  $J_\sigma$  is *quasiconvex*. Note that quasiconvexity is not necessary in general for an estimator to have properties (i)–(iii).

The analysis of the gradient in Section 2 yields a fairly precise description of the *shape* of  $J_\sigma$  as a function of  $\sigma$ . For noise-free oscillatory data, small

$\sigma$  implies quasiconvexity. For large  $\sigma$  on the other hand we show that  $J_\sigma$  becomes quite flat away from the global minimizer (Theorem 2.9(a)), and in fact then closely resembles the output least squares objective in its general form, as one would suspect. Therefore to retain properties (i)–(iii)  $\sigma$  *must* be kept bounded away from both 0 and  $\infty$ . The estimates of Section 2 are qualitative in nature. Precision in the constants appearing in these estimates would have little value. Instead, an adaptive algorithm for adjustment of  $\sigma$  is needed. We do not give such an algorithm here.

The results of Section 2 are illustrated by a number of numerical experiments discussed at the end of the section.

The term  $\sigma^2 \|\partial u / \partial x\|^2$  measures the extent to which the waveforms  $u(\cdot, x)$  resemble each other for nearby values of  $x$ . We call this term a *differential semblance* measure, and  $\sigma$  the differential semblance weight.

As mentioned earlier, the planewave detection problem is studied here mainly as a simple model for various inverse problems in hyperbolic partial differential equations. Versions of differential semblance optimization, analogous to minimization of  $J_\sigma$ , have been applied to a variety of such problems by the author and his associates; see the references cited at the beginning of this section.

Norms in the  $L^2$ -based Sobolov spaces  $H^k$  will be denoted by  $\|\cdot\|_k$ . Thus  $\|\cdot\|_0$ , or simply  $\|\cdot\|$ , is the norm in  $L^2$ . Domains are either evident from context or specified explicitly. Boundary conditions will be introduced as needed and specified by subscripts.

# 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 order to fulfill the standard conditions for convergence of Newton's method, we must severely restrict the smoothness of the signal  $u(t)$ . Having made this restriction, we then find that, even so, the least squares cost functional has multiple local minima, which can be as close together as we like — even for noise-free data! As pointed out in the introduction, this means that the least squares formulation cannot be regarded as well-posed, or alternatively that the amount of regularization necessary to render the problem well-posed is bounded away from zero, even when the data error vanishes.

We shall assume that

(i) the wave form and signal are square-integrable;

(ii) either

a) the waveform  $u$  is periodic with period  $T$ , and  $\int_0^T dt u(t) = 0$

or

b) the support in  $t$  of  $u(t + sx)$  is known *a priori* to be contained in the interval  $[0, T]$  for any  $s \in [-1, 1]$  and  $x \in [-X, X]$

where  $[-X, X]$  is the interval of recording positions, and  $[0, T]$  is the time interval of the recording. Note that the direction parameter  $s$  is necessarily in the interval  $[-1, 1]$ . Cases (a) and (b) in assumption (ii) correspond to the two most common plane wave signals: periodic and impulsive. Under our assumption, we can view (b) as a special case of (a) by extending  $u$   $T$ -periodically, but we prefer to single it out to emphasize the more common impulsive case. Assumption (ii) ensures that we have sufficient amount of information of the plane wave at each recording position. The second part of assumption (iia) implies that  $u$  is oscillatory. We will not use this assumption until the next section. The signal resulting from the waveform  $u$  at direction  $s$  is

$$\phi[s, u](x, t) = u(t + sx) .$$

We may regard  $\phi$  as a mapping

$$\phi : [-1, 1] \times L^2[0, T] \rightarrow L^2([-X, X] \times [0, T]) .$$

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

given data

$$z \in L^2([-X, X] \times [0, T])$$

find

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

to minimize

$$J[s, u; z] = \|\phi(s, u) - z\|_0^2 .$$

Suppose that we regard the minimization of  $J$  as equivalent to finding a stationary point of its gradient, i.e. we ignore the possible presence of stationary points other than minima. Formally,

$$\text{grad } J[s, u; z] = 2D\phi[s, u]^T(\phi[s, u] - z)$$

where

$$D\phi[s, u][\delta s, \delta u](x, t) = \delta u(t + sx) + \delta s x \frac{du}{dt}(t + sx) .$$

The presence of a derivative on the right-hand side of the last equation suggests that  $D\phi$  is not well-defined at general  $[s, u] \in [-1, 1] \times L^2[0, T]$ . In fact it is easy to show that, while  $\phi$  as defined above is continuous, it is not locally uniformly continuous, hence *a fortiori* not differentiable.

If we are to employ smooth optimization methods at all, we must change the definition of  $\phi$ , either by strengthening the topology on the domain or weakening the topology on the range. We choose to maintain the  $L^2$  measure of error on the range. Then the signal  $u(t)$  must have a derivative in the sense of  $L^2$  in order that the derivative  $D\phi$  of  $\phi$  exist. We shall state the following result in a fashion appropriate to the “impulsive” case ((iib) above). Similar statements can be made concerning the periodic case ((iia) above); we leave the formulation of these to the reader. The proof of the following result is straightforward:

**Theorem 1.1** For  $k = 0, 1, 2, \dots$ , regard  $\phi$  as a map

$$\phi : [-1, 1] \times H_0^k[0, T] \rightarrow L^2([-X, X] \times [0, T]),$$

then:

(a)  $\phi$  is of class  $C^{j-1,1}$  if and only if  $k \geq j$ .

(b) the gradient of  $J$  is given by

$$\text{grad } J[s, u; z] = [g_s, g_u]$$

where

$$g_s = 2 \int_{-X}^X dx \int_0^T dt x \frac{\partial}{\partial t} u(t + sx) (u(t + sx) - z(x, t))$$

and  $g_u$  is the solution of the boundary-value problem (in weak sense)

$$\sum_{i=0}^k (-1)^i \frac{\partial^{2i}}{\partial t^{2i}} g_u(t) = 2Xu(t) - \int_{-X}^X dx z(x, t - sx)$$

with boundary conditions

$$g_u(t) = \frac{d}{dt} g_u(t) = \dots = \frac{d^{k-1}}{dt^{k-1}} g_u(t) = 0 \quad \text{at } t = 0, T.$$

We digress to recall the standard conditions for local quadratic convergence of Newton's method for the equation  $F(x) = 0$  to a solution  $x_* \in U$ , where  $F : U \rightarrow \bar{Y}$ ,  $U \subset X$  open,  $X$  and  $Y$  are Banach spaces:

- (1)  $F$  should be of class  $C^{1,1}(U, Y)$ , i.e. be continuously differentiable with Lipschitz-continuous derivative;
- (2) the Jacobian map  $F'(x_*)$  should be coercive, i.e.,

$$|F'(x_*)\delta x|_Y \geq C|\delta x|_X$$

(see e.g. Dennis and Schnabel [1983], Theorem 5.2.1). We wish to take  $\text{grad} J$  for  $F$ , one of the spaces  $[-1, 1] \times H^k[0, T]$  for  $X$ , and  $L^2([-X, X] \times [0, T])$



for  $Y$ . For  $\text{grad } J$  to be of class  $C^{1,1}$ , it is sufficient that  $\phi$  is of class  $C^{2,1}$  which requires  $k \geq 3$ , according to Theorem 1.1. In fact it is not hard to prove that  $k \geq 3$  is also necessary (at least so long as  $k$  is restricted to be an integer). Accordingly we shall regard  $\phi$  as a map

$$\phi : [-1, 1] \times H_0^3[0, T] \longrightarrow L^2([-X, X] \times [0, T]) .$$

The Jacobian of the gradient is the *Hessian map*. Its action on a vector of the form  $[0, \delta u]$  is given by

$$H[s, u][0, \delta u] = h_{uu} ,$$

where  $h_{uu}$  is the solution of the boundary-value problem:

$$\sum_{i=0}^3 (-1)^i \frac{\partial^{2i}}{\partial t^{2i}} h_{uu}(t) = 2X \delta u(t)$$

$$h_{uu}(t) = \frac{d}{dt} h_{uu}(t) = \frac{d^2}{dt^2} h_{uu}(t) = 0 \quad \text{at } t = 0, T .$$

Evidently  $H$  is a smoothing operator  $H^{-3}[0, T] \rightarrow H_0^3[0, T]$ . Thus the spectrum of  $H$  accumulates at zero, and the second sufficient condition (coercivity) fails. This implies that the cost functional might not be globally convex, thus multiple minima might be present. This conjecture is proved in Symes [1991b], and illustrated by the following very simple example. In this example we view  $\phi$  once again as a map

$$\phi : [-1, 1] \times L^2[0, T] \rightarrow L^2([-X, X] \times [0, T])$$

as the choice of norm in the domain does not affect the properties illustrated here.

Let  $X = \frac{1}{2}$ ,  $T = \pi$ , and  $z(x, t) = \sin \omega t$  be consistent and noise-free data. It is easy to show via normal equation that, given  $s$ , the unique minimizer

$$u[s] = \operatorname{argmin}_u J[s, u; z]$$

is

$$u[s](t) = \int_{-X}^X dx \sin \omega(t - sx)$$

and straightforward calculations lead to

$$J_{\text{opt}}[s] := J[s, u[s]] = \frac{\pi}{4} \left[ 1 - \frac{\sin^2\left(\frac{\omega s}{2}\right)}{\left(\frac{\omega s}{2}\right)^2} \right].$$

Figure 1.1 and Figure 1.2 are graphs of  $J_0$  with  $\omega = 50$  and  $\omega = 500$  respectively. A quick examination of  $J[s]$  shows that this function has as many local minima as one likes, and they can be as close as one likes by choosing  $\omega$  sufficiently large. For any local minimizer  $s^*$  of  $J[s]$ ,  $(s^*, u[s^*])$  is a local solution of

$$\min_{s,u} \|\phi[s, u] - z\|_0^2.$$

Hence, this output least squares problem can have as many solutions as one likes, they spread as wide as one likes, and they can be as close as one likes too, by just choosing  $\omega$  large enough! Keep in mind that  $z$  is smooth and of class  $L^2$  for any  $\omega$ , and it is consistent and noise-free! Therefore the solution of the output least-squares problem cannot be stable, in the sense of the discussion in the introduction.

This example also reveals a typical feature of the output least squares detection problem for bandlimited data: if the frequencies of the data concentrate near  $\omega$ , then the size of the neighborhood of the global minimizer of  $J[s]$  in which  $J[s]$  is convex is of  $O(\frac{1}{\omega})$ . Outside of this neighborhood,  $J[s]$  tends to be flat (“saturated”), with many local minima.

Another example of this nature is the output least squares detection of a plane wave produced from the Ricker wavelet (i.e. the second derivative of a Gaussian pulse — see Figure 1.3). Figure 1.4 is the noise-free sampling of the plane wave on spatial range  $[-0.5, 0.5]$  in a time interval  $[0, 1.0]$ , and Figure 1.5 is the sampling with 160% noise. There are 41 receiver signals (“traces”) in both instances, each with 512 samples. The time shifts of the sampled input wave form are performed *via* the Fourier transform. All calculations were done in MATLAB.

Figure 1.6 and Figure 1.7 are the corresponding numerically computed mean square errors  $J[s]$  vs. direction  $s$ . Both curves are nonconvex and very sharp at their global minima  $s = 0.1$ . Success in minimizing such functions relies heavily on a good initial guess or on a global search. Note also that

Figure 1.7 shows that with noise-corrupted data,  $J[s]$  tends to have multiple minima.

## 2 Differential Semblance Optimization

In this section we introduce an alternative to the output least-squares functional, which we have named *differential semblance optimization* (DSO). We develop the basic regularity and convexity properties of the new objective function, then present some numerical examples. We begin with some remarks intended to motivate the new approach.

As the plots of residuals in the last section suggest, one of the difficulties faced by output least-squares is the necessity of fitting all of the data traces simultaneously with a single waveform  $u$  and slope  $s$ . On the other hand the data-fitting problem for a single trace:

$$\min_{u,s} \int [u(t + sx) - z(x, t)]^2 dt$$

while nonconvex for the same reasons as before, is nonetheless trivial to solve. In fact, zero residual can always be achieved simply by setting

$$u(t) = z(x, t - sx)$$

and this residual is independent of  $s$ , in contrast to the multi-trace case considered in the last section. Alternatively, suppose that the models for each trace are regarded as independent, i.e.  $u$  is allowed to depend on  $x$ : then, once again zero residual is achieved, independently of  $s$ , through solution of the quadratic problem

$$\min_{u,s} \int \int dx dt |u(x, t + sx) - z(x, t)|^2$$

the solution of which is

$$u(x, t) = z(x, t - sx)$$

with no restriction on  $s$ .

Of course the price paid for this complete relaxation of the data-fitting problem is the complete loss of the plane-wave model, which is the “physics” of the plane-wave detection problem.

The plane-wave model is retrieved by imposition of any constraint forcing  $u$  to be independent of  $x$ ; for example, minimizing the above function under

the constraint

$$\frac{\partial u}{\partial x} = 0$$

is apparently equivalent to the output least-squares problem of Section 1 and therefore will not be amenable to solution through versions of Newton's method or its relatives adapted to equality-constrained problems, for instance sequential quadratic programming.

The following reasoning suggests our resolution of this difficulty. We view  $\phi$  as a mapping

$$\begin{aligned} \phi : [-1, 1] \times L^2([-X, X] \times [0, T]) &\longrightarrow L^2([-X, X] \times [0, T]) \\ \phi[s, u](x, t) &= u(x, t + sx) \end{aligned}$$

and solve the detection problem in two steps:

Step 1. Solve

$$\min_u \|\phi[s, u] - z\|^2$$

to get  $u[s]$ ;

Step 2. Solve

$$\min_s \left\| \frac{\partial}{\partial x} u[s] \right\|^2 =: J[s]$$

to get  $s_*$ , hence  $u[s_*]$ . It is easy to see that

$$\begin{aligned} u[s](t, x) &= z(x, t - sx) \\ J[s] &= \left\| \frac{\partial z}{\partial x} - s \frac{\partial z}{\partial t} \right\|^2 \end{aligned}$$

and

$$s_* = \frac{\langle \frac{\partial z}{\partial x}, \frac{\partial z}{\partial t} \rangle}{\langle \frac{\partial z}{\partial t}, \frac{\partial z}{\partial t} \rangle}$$

where  $\langle \cdot, \cdot \rangle$  is the inner product on  $L^2([-X, X] \times [0, T])$ , and we used the fact that  $\phi[s, \cdot]$  is an isometry on this space under the assumption (iib) in Section 1. Note that Step 1 is easily solved via the normal equation, which is a linear equation, and in Step 2  $J[s]$  is perfectly smooth and convex. The obvious problem is that one needs to compute the derivatives of  $z$ , which are generally

not square integrable. In fact, evidently  $u$  (which must be differentiated) is only as smooth as  $z$ , which doesn't have a derivative. We resolve this conflict by parametrizing the waveform as  $\partial u / \partial t$ , rather than as  $u$ , minimizing the penalized functional

$$\frac{1}{\sigma^2} \left\| \frac{\partial}{\partial t} \phi[s, u] - z \right\|^2 + \left\| \frac{\partial}{\partial x} u \right\|^2$$

in two steps as above. The change of unknown ensures that the presumption that  $z \in L^2$  will not be violated. We will obtain a linear well-posed problem for  $u$  (Step 1). Moreover the minimum (over  $u$ ) is a smooth, globally convex function of  $s$  if  $z$  is close to a consistent datum and  $\sigma$  is sufficiently small. In this case, at the stage of Step 2, the first term in the functional will be small and not destroy the convexity of the second term. Note also that in the limit  $\sigma \rightarrow 0$  this problem becomes constrained by the first term, therefore reduces to the problem discussed above.

That is, for small  $\sigma$ , this problem resembles the minimization of  $\|\partial u / \partial x\|_0^2$  subject to the constraint that  $u$  predicts the data correctly. The  $x$ -derivative of  $u$  measures the extent to which the waveforms, chosen to explain signals at *neighboring* receivers, resemble each other. This idea motivates the name *differential semblance*.

On the other hand, as  $\sigma \rightarrow \infty$  the problem tends to resemble the minimization of the mean-square error subject to the constraint  $\partial u / \partial x \equiv 0$ , which is of course exactly the output least-squares problem of the previous section. Thus the differential semblance functional *interpolates* between the "pure semblance" problem described above ( $\sigma \rightarrow 0$ ) and output least squares ( $\sigma \rightarrow \infty$ ). This feature will be illustrated numerically at the end of the section.

We begin the development of these results by defining some useful function spaces.

Define the extended model space  $\tilde{E}$  by

$$H_b^1 = \left\{ u \in H^1([-X, X] \times [0, T]) : u(x, 0) = u(x, T) \right. \\ \left. |x| \leq X ; \int_0^T dt \int_{-X}^X dx u \equiv 0 . \right\}$$

$$\tilde{E} = \mathbb{R} \times H_b^1 .$$

We define a family of norms on  $\tilde{E}$  by

$$\|u\|_{\tilde{E},\sigma}^2 = (|s|^2 + \|u\|_{1,\sigma}^2)$$

where

$$\|u\|_{1,\sigma}^2 = \int_{-X}^X dx \int_0^T dt \left( \left| \frac{\partial u}{\partial t} \right|^2 + \sigma^2 \left| \frac{\partial u}{\partial x} \right|^2 \right)$$

for  $x = [s, u] \in \tilde{E}$ .

The norm is definite for  $\sigma > 0$  because of Poincaré's inequality (Gilbarg and Trudinger (1983), p. 164). Define the extended data space  $\tilde{F}$  by

$$\tilde{F} = F \times F$$

where as before  $F = L^2([-X, X] \times [0, T])$ , viewed as a subspace of T-periodic functions in  $t$ , with the obvious norm. Define finally for  $\sigma \in \mathbb{R}_+$ :

$$\tilde{\phi}_\sigma : \tilde{E} \longrightarrow \tilde{F}$$

by

$$\tilde{\phi}_\sigma[s, u](x, t) = \begin{bmatrix} \frac{\partial u}{\partial t}(x, t + sx) \\ \sigma \frac{\partial u}{\partial x}(x, t) \end{bmatrix} .$$

For  $z \in F$ , let  $\tilde{z} = [z, 0]^T \in \tilde{F}$ .

*Remark.* The reason for the definition of  $\tilde{\phi}$  using  $\partial u / \partial t$ , rather than  $u$  as in Section 1, was hinted above and will be explained further below.

Evidently  $\tilde{\phi}_\sigma$  is once again quite irregular (continuous, but not locally uniformly). Motivated by the remarks at the beginning of the section, define nonetheless

$$\tilde{J}_\sigma[s] = \inf_{u \in H_b^1} \left\| \tilde{\phi}_\sigma[s, u] - \tilde{z} \right\|_{\tilde{F}}^2 .$$

**Theorem 2.1**  $\tilde{J}_\sigma[s] = \|\tilde{\phi}_\sigma[s, u[s]] - \tilde{z}\|_{\tilde{F}}^2$  where  $u[s]$  is the (weak) solution of the boundary value problem

$$-\left(\frac{\partial^2}{\partial t^2} + \sigma^2 \frac{\partial^2}{\partial x^2}\right) u(x, t) = -\frac{\partial z}{\partial t}(x, t - sx)$$

$$u(x, 0) = u(x, T), -X \leq x \leq X$$

$$\frac{\partial u}{\partial x}(-X, t) = \frac{\partial u}{\partial x}(X, t) \equiv 0, \quad 0 \leq t \leq T$$

$$\int_{-X}^X dx \int_0^T dt u = 0.$$

Also, the estimate

$$\|u[s]\|_{1,\sigma} \leq \|z\|_0$$

holds.

*Proof.* Denote by  $H_b^{-1}$  the dual space of  $H_b^1$ . For each  $s \in \mathbb{R}$ , define the operator  $T_s$  on  $F$  by

$$T_s f(x, t) = f(x, t + sx)$$

(interpreted as 1-periodic in  $t$ ).  $T_s$  is bounded and continuous for each  $s \in \mathbb{R}$ , and the following relations are easily verified:

$$T_s^T f(x, t) = f(x, t - sx)$$

$$T_s^T T_s = I.$$

Since the derivatives  $\partial/\partial t$  and  $\partial/\partial x$  map  $H_b^1 \rightarrow F$  continuously, we can view their transposes (formal adjoints) as maps  $F \rightarrow H_b^{-1}$ . Finally, define for each  $s \in \mathbb{R}$  the map

$$\Phi_{s,\sigma} : H_b^1 \rightarrow \tilde{F}$$

by

$$\Phi_{s,\sigma} u = \tilde{\phi}_\sigma[s, u].$$

Evidently  $\Phi_{s,\sigma}$  is a bounded linear map, and

$$\Phi_{s,\sigma} u = \begin{bmatrix} T_s \frac{\partial u}{\partial t} \\ \sigma \frac{\partial u}{\partial x} \end{bmatrix}, \quad u \in H_b^1.$$



Moreover,

$$\begin{aligned}
\Phi_{s,\sigma}^T \Phi_{s,\sigma} &= \left( \frac{\partial}{\partial t} \right)^T T_s^T T_s \frac{\partial}{\partial t} + \sigma^2 \left( \frac{\partial}{\partial x} \right)^T \frac{\partial}{\partial x} \\
&= - \left( \frac{\partial^2}{\partial t^2} + \sigma^2 \frac{\partial}{\partial x^2} \right) \\
\Phi_{s,\sigma}^T \tilde{z} &= - \frac{\partial}{\partial t} T_{-s} z
\end{aligned}$$

where we have identified  $(\partial/\partial t)^T$  with  $-\partial/\partial t$  as usual and thus regarded the Laplace operator as a map:  $H_b^1 \rightarrow H_b^{-1}$ .

The objective function of the variational problem posed in the statement of the theorem can be written as

$$\begin{aligned}
&\|\Phi_{s,\sigma} u - \tilde{z}\|_{\tilde{F}}^2 \\
&= \|\Phi_{s,\sigma} u\|_{\tilde{F}}^2 - 2\langle \Phi_{s,\sigma} u, \tilde{z} \rangle_{\tilde{F}} + \|\tilde{z}\|_{\tilde{F}}^2 \\
&= \|\Phi_{s,\sigma} u\|_{\tilde{F}}^2 - 2\langle u, \Phi_{s,\sigma}^T \tilde{z} \rangle_{H_b^1, H_b^{-1}} + \|\tilde{z}\|_{\tilde{F}}^2 .
\end{aligned}$$

The first term defines a coercive quadratic form on  $H_b^1$ , thanks to Poincaré's inequality. The second term defines a continuous linear form on  $H_b^1$ . According to a basic result in the calculus of variations (the Lax-Milgram Theorem, see e.g. Gilbarg and Trudinger (1983), p. 83), the objective function attains a (global) minimum at  $u \in H_b^1$  solving the *normal equations*

$$\Phi_{s,\sigma}^T \Phi_{s,\sigma} u = \Phi_{s,\sigma}^T \tilde{z} .$$

The translation into the boundary value problem stated in the theorem follows the standard pattern of the direct method (Gilbarg and Trudinger (1983), Ch. 8).

To prove the bound, suppose first that  $z$  is smooth, T-periodic in  $t$ , and satisfies

$$\frac{\partial}{\partial x} (z(x, t - sx)) \equiv 0 \quad \text{if } x = \pm X .$$

Then

$$z(x, t - sx) = \sum_{\substack{n, m \in \mathbb{Z}; \\ |n| + |m| > 0}} \hat{z}_{mn} \cos \left( \frac{\pi m x}{2X} + \frac{m\pi}{2} \right) e^{\frac{2\pi i n t}{T}}$$

the series being absolutely convergent together with all derivatives. Accordingly

$$u(x, t) = \sum_{\substack{n, m \in \mathbb{Z}; \\ |n| + |m| > 0}} \hat{u}_{mn} \cos\left(\frac{\pi m x}{2X} + \frac{m\pi}{2}\right) e^{2\pi \frac{in t}{T}}$$

with

$$\hat{u}_{mn} = \frac{2\pi i n / T}{\sigma^2 \left(\frac{\pi m}{2X}\right)^2 + \left(\frac{2\pi n}{T}\right)^2} \hat{z}_{mn}$$

and in particular

$$\left|2\frac{\pi n}{T} u_{mn}\right| \leq |z_{mn}|, \left|\sigma \frac{\pi m}{2X} u_{mn}\right| \leq |z_{mn}|.$$

The indicated bound follows immediately from the fact that  $T_{-s}$  is an isometry of  $F$ . Note that more generally for  $\gamma > 0$

$$\|u[s]\|_{1, \gamma} \leq \frac{\gamma}{\sigma} \|z\|.$$

q.e.d.

If we denote the solution operator of the problem in Theorem 2.1 by  $N_\sigma$ , then similar arguments show that  $\frac{\partial}{\partial t} N_\sigma \frac{\partial}{\partial t}$  is contractive on  $F$ , i.e. we have

**Proposition 2.2**  $\left\|\frac{\partial}{\partial t} N_\sigma \frac{\partial}{\partial t} u\right\| \leq \|u\|$  for  $u \in F$ .

**Theorem 2.3**  $J_\sigma \in C^\infty(\mathbb{R})$

*Proof.* Write the objective function in the statement of Theorem 2.1 in the form

$$\left\|T_s \frac{\partial u}{\partial t} - z\right\|_F^2 + \sigma^2 \left\|\frac{\partial u}{\partial x}\right\|_F^2.$$

Note that  $T_s$  and  $\partial/\partial t$  commute. Also  $H_b^1 \subset F \subset H_b^{-1}$  and  $T_s$  and its formal adjoint  $T_s^T$  both restrict to isomorphisms of  $H_b^1$  interpreted as consisting of 1-periodic functions. Thus

$$\left\|T_s \frac{\partial u}{\partial t} - z\right\|_F^2 + \sigma^2 \left\|\frac{\partial u}{\partial x}\right\|_F^2$$

$$\begin{aligned}
&= \left\| \frac{\partial}{\partial t} T_s u - z \right\|_F^2 + \sigma^2 \left\| \frac{\partial u}{\partial x} \right\|_F^2 \\
&= \left\| \frac{\partial v}{\partial t} - z \right\|_F^2 + \sigma^2 \left\| \frac{\partial}{\partial x} T_s^T v \right\|_F^2
\end{aligned}$$

where  $v = T_s u$  (so that  $u = T_s^T v$ ). It follows that

$$\tilde{J}_\sigma[s] = \inf_{v \in H_b^1} \left\{ \left\| \frac{\partial v}{\partial t} - z \right\|_F^2 + \sigma^2 \left\| \frac{\partial}{\partial x} T_s^T v \right\|_F^2 \right\}.$$

Since  $\frac{\partial}{\partial x} T_s^T = T_s^T \left( \frac{\partial}{\partial x} - s \frac{\partial}{\partial t} \right)$  and  $T_s^T$  is an isometry of  $F$ , it follows that also

$$J_\sigma[s] = \inf_{v \in H_b^1} \left\{ \left\| \frac{\partial v}{\partial t} - z \right\|_F^2 + \sigma^2 \left\| \frac{\partial v}{\partial x} - s \frac{\partial v}{\partial t} \right\|_F^2 \right\}.$$

The problem just defined is also a convex variational problem, as its quadratic part is also coercive:

$$\begin{aligned}
&\left\| \frac{\partial v}{\partial t} \right\|_F^2 + \sigma^2 \left\| \frac{\partial v}{\partial x} - s \frac{\partial v}{\partial t} \right\|_F^2 = \\
&= (1 + s^2 \sigma^2) \left\| \frac{\partial v}{\partial t} \right\|_F^2 + \sigma^2 \left\| \frac{\partial v}{\partial x} \right\|_F^2 - 2\sigma^2 s \left\langle \frac{\partial v}{\partial x}, \frac{\partial v}{\partial t} \right\rangle_F \\
&\geq \frac{1}{2} \left\| \frac{\partial v}{\partial t} \right\|_F^2 + \frac{\sigma^2}{1 + 2s^2 \sigma^2} \left\| \frac{\partial v}{\partial x} \right\|_F^2
\end{aligned}$$

So in exactly the same way as before it attains its minimum at  $v[s] \in H_b^1$ , given by  $v[s] = \tilde{N}_{s,\sigma} z$ . Here  $\tilde{N}_{s,\sigma}$  is the inverse operator of

$$v \mapsto \left\{ \left( \frac{\partial}{\partial t} \right)^T \left( \frac{\partial}{\partial t} \right) + \sigma^2 \left( \frac{\partial}{\partial x} - s \frac{\partial}{\partial t} \right)^T \left( \frac{\partial}{\partial x} - s \frac{\partial}{\partial t} \right) \right\} v =: \tilde{M}_{s,\sigma} v \in H_b^{-1}$$

where the transposes are taken in the sense of  $F \rightarrow H_b^{-1}$ . Thus

$$J_\sigma[s] = \left\| \frac{\partial v[s]}{\partial t} - z \right\|_F^2 + \sigma^2 \left\| \frac{\partial v[s]}{\partial x} - s \frac{\partial v[s]}{\partial t} \right\|_F^2.$$

The operator-valued function

$$s \mapsto \frac{\partial}{\partial x} - s \frac{\partial}{\partial t} : H_b^1 \longrightarrow F$$

is  $C^\infty$  (linear!) in the operator norm, therefore so is its adjoint  $F \rightarrow H_b^{-1}$ , whence

$$s \mapsto \tilde{M}_{s,\sigma} : H_b^1 \longrightarrow H_b^{-1}$$

is a  $C^\infty$  operator-valued function. Since its values are invertible, with (locally) uniformly bounded inverse  $\tilde{N}_{s,\sigma}$  it follows from the calculus that

$$s \mapsto \tilde{N}_{s,\sigma} : H_b^{-1} \longrightarrow H_b^1$$

is  $C^\infty$ . Therefore  $s \rightarrow v[s] = \tilde{N}_{s,\sigma} z \in H_b^1$  is  $C^\infty$ , and the statement of the theorem follows immediately. q.e.d.

*Remark.* In a prior incarnation of this work (Symes [1990]), the entire treatment was based on the map

$$\tilde{\Phi}_{s,\sigma} v \mapsto \begin{bmatrix} \frac{\partial v}{\partial t} \\ \sigma \left( \frac{\partial v}{\partial x} - s \frac{\partial v}{\partial t} \right) \end{bmatrix}$$

which plays a major role in the proof of Theorem 2.3. There is a rough equivalence between variational problems based on  $\tilde{\Phi}_{s,\sigma}$  and those based on  $\Phi_{s,\sigma}$  though only in the case of  $t$ -periodic models does it seem possible to establish an exact relationship.

*Remark.* We will sketch here another proof, which works when the definition of the map  $\Phi$  incorporates a cutoff:

$$\Phi_{s,\sigma} u = \begin{bmatrix} g T_s \frac{\partial u}{\partial t} \\ \sigma \frac{\partial u}{\partial x} \end{bmatrix}$$

with  $g \in C_0^\infty(\mathbb{R}^2)$ . In fact, this problem setting is more prototypical of the seismic inverse problems which motivate this work, which require a similar approach. It is natural to take for the model space  $H_0^1(\Omega)$ , where  $\Omega$  is a

domain large enough that data in the support of  $g$  can be reproduced with models supported in  $\Omega$ , for the range of slopes  $s$  considered. Evidently the normal operator of  $\Phi$  is no longer elliptic, and this is also a characteristic feature of the reflection inversion problems. It is necessary to regularize the normal operator; in the present case, one could for example redefine

$$\Phi_{s,\sigma,\lambda} = \left[ gT_s \frac{\partial}{\partial t}, \sigma \frac{\partial}{\partial x}, \lambda \frac{\partial}{\partial t} \right]^T$$

$$H_0^1(\Omega) \longrightarrow F \times F \times F .$$

Now zero residual is no longer possible, as usual for regularized problems. The normal equations now read

$$- \left[ \frac{\partial}{\partial t} T_s^T g^2 T_s \frac{\partial}{\partial t} + \sigma^2 \frac{\partial^2}{\partial x^2} + \lambda^2 \frac{\partial}{\partial t^2} \right] u = - \frac{\partial}{\partial t} T_s^T g z$$

$$u \equiv 0 \quad \text{on} \quad \partial\Omega .$$

If it is arranged — as it always can be — that the support of the right-hand side is interior to  $\Omega$ , then the solution operator  $N_{s,\sigma,\lambda}$  of this Dirichlet problem differs from a certain pseudo-differential operator by an infinitely smoothing error. The symbol of the pseudodifferential operator  $N_{s,\sigma,\lambda}$  can be calculated explicitly, and when  $u = -N_{s,\sigma,\lambda} \frac{\partial}{\partial t} T_s^T g z$

$$J_\sigma[s] = \left\| \Phi_{s,\sigma,\lambda} u - (z, 0, 0)^T \right\|_{F \times F \times F}^2 = \left\| \left( gT_s \frac{\partial}{\partial t} N_{s,\sigma,\lambda} \frac{\partial}{\partial t} T_s^T g - I \right) z \right\|^2$$

$$- \sigma^2 \left\langle z, gT_s \frac{\partial}{\partial t} N_{s,\sigma,\lambda} \frac{\partial^2}{\partial x^2} N_{s,\sigma,\lambda} \frac{\partial}{\partial t} T_s^T g z \right\rangle$$

$$- \lambda^2 \left\langle z, gT_s \frac{\partial}{\partial t} N_{s,\sigma,\lambda} \frac{\partial^2}{\partial t^2} N_{s,\sigma,\lambda} \frac{\partial}{\partial t} T_s^T g z \right\rangle .$$

Each of the terms in this sum involves operators of the form  $T_s$  (pseudo D.O.)  $T_s^T$ . It turns out that such operators are also (essentially) pseudodifferential (this observation is related to Egorov's theorem (Taylor (1980), pp. 147 ff.); their symbols can be computed explicitly and are smooth functions of  $s$ . It follows immediately that  $J_\sigma[s]$  as defined above is smooth.

For more comment on this reasoning see Symes (1991c).

Next we compute the gradient of  $J$ . We assume temporarily that the data  $z$  is smooth. Then  $u[s]$  is smooth and periodic, and

$$\frac{d}{ds} (\Phi_{s,\sigma} u[s]) = \left[ x \frac{\partial}{\partial t} T_s \frac{\partial}{\partial t} u[s] \right] + \Phi_{s,\sigma} \frac{d}{ds} u[s].$$

So

$$\begin{aligned} \frac{d}{ds} J[s] &= \frac{d}{ds} \|\Phi_{s,\sigma} u[s] - \tilde{z}\|_F^2 \\ &= 2 \left\langle \frac{d}{ds} (\Phi_{s,\sigma} u[s]), \Phi_{s,\sigma} u[s] - \tilde{z} \right\rangle \\ &= 2 \left\langle x \frac{\partial u}{\partial t} T_s \frac{\partial u}{\partial t} [s], T_s \frac{\partial}{\partial t} u[s] - z \right\rangle \\ &\quad + 2 \left\langle \frac{d}{ds} u[s], \Phi_{s,\sigma}^T (\Phi_{s,\sigma} u[s] - \tilde{z}) \right\rangle \end{aligned}$$

and the second term in the sum vanishes by virtue of the normal equations.

As noted before,  $\frac{\partial}{\partial t}$  and  $T_s$  commute. Note also that (multiplication by)  $x$  and  $T_s$  commute, so we obtain

$$\frac{d}{ds} J[s] = 2 \left\langle x \frac{\partial}{\partial t} u[s], \left( T_s \frac{\partial}{\partial t} \right)^T \left( T_s \frac{d}{dt} u[s] - z \right) \right\rangle.$$

We will need the following technical result.

**Theorem 2.4** *Suppose that  $z$  is a smooth ( $C^\infty$ ) function, 1-periodic in  $t$ . Then  $u[s]$  and all of its  $s$ -derivatives are smooth functions, 1-periodic in  $t$ .*

*Proof.* Recall that the r.h.s. of the normal equation for  $u[s]$  is

$$-T_s^T \frac{\partial z}{\partial t}.$$

Evidently if  $z$  is smooth then this quantity is a smooth function all  $s$ -derivatives of which are also smooth. The first derivative with respect to  $s$  is

$$\begin{aligned} \frac{d}{ds} \left( -T_s^T \frac{\partial z}{\partial t} \right) (x, t) &= \frac{d}{ds} \left( -\frac{\partial z}{\partial t} (x, t - sx) \right) \\ &= x \frac{\partial^2 z}{\partial t^2} (x, t - sx) \\ &= x \frac{\partial}{\partial t} T_s^T \frac{\partial z}{\partial t} \end{aligned}$$

and so on. Since the r.h.s. of the normal equations are smooth in  $s$ , so is the solution. The same reasoning applies to all higher derivatives. q.e.d.

Now suppose that  $z \in C^\infty$ . The normal equations can also be written

$$\left( T_s \frac{\partial}{\partial t} \right)^T \left( T_s \frac{\partial}{\partial t} u - z \right) - \sigma^2 \frac{\partial^2 u}{\partial x^2} = 0$$

so

$$\begin{aligned} \frac{d}{ds} J[s] &= 2\sigma^2 \left\langle x \frac{\partial}{\partial t} u[s], \frac{\partial^2}{\partial x^2} u[s] \right\rangle \\ &= -2\sigma^2 \left\langle \frac{\partial}{\partial x} \left( x \frac{\partial}{\partial t} u[s] \right), \frac{\partial}{\partial x} u[s] \right\rangle \\ &= -2\sigma^2 \left[ \left\langle \frac{\partial}{\partial t} u[s], \frac{\partial}{\partial x} u[s] \right\rangle \right. \\ &\quad \left. + \left\langle x \frac{\partial}{\partial t} \left( \frac{\partial}{\partial x} u[s] \right), \frac{\partial}{\partial x} u[s] \right\rangle \right] \end{aligned}$$

where we have used the assumed smoothness of  $u[s]$  to take advantage of the Neumann boundary conditions in integrating by parts.

It is easy to see that the operator  $x \frac{\partial}{\partial t}$  is antisymmetric on smooth functions periodic in  $t$ , so the second term above vanishes and we are left with the result

**Theorem 2.5**

$$\frac{d}{ds} J[s] = -2\sigma^2 \left\langle \frac{\partial}{\partial t} u[s], \frac{\partial}{\partial x} u[s] \right\rangle.$$

*Proof.* We have established this formula under the assumption that  $z \in F$  was actually smooth. However, note that  $u[s]$  is a continuous  $H_b^1$ -valued function of  $z \in F$ ; it follows that the formula is correct in general. q.e.d.

*Remark.* At this point it is possible to justify the inclusion of  $\partial/\partial t$  in the definition of  $\tilde{\phi}_\sigma$ . Indeed, suppose we were to define instead

$$\phi_\sigma[s, u] = \begin{pmatrix} T_s u \\ \sigma \frac{\partial u}{\partial x} \end{pmatrix}$$

and regard  $\phi_\sigma$  as mapping  $H^1$ -functions of  $x$ , with values in  $L^2$  1-periodic functions of  $t$ , into  $\tilde{F}$ . The normal equations may then be written

$$\left( 1 - \sigma^2 \frac{\partial^2}{\partial x^2} \right) u = T_s^T z.$$

*Exactly the same calculation* leads to the formula stated in Theorem 2.5, under the assumption of smooth  $z$ . In particular, the  $t$ -derivative of  $u$  appears, because the  $s$ -derivative of  $u(x, t + sx)$  produces it! Now, however, the quadratic form in  $z$ , defined in Theorem 2.5, does not extend to  $z \in F$ : the gradient of  $T$  is only defined at a dense set of  $z \in F$ . In fact, when  $\partial/\partial t$  is not included in the definition of  $\tilde{\phi}_\sigma$ , the normal operator is not elliptic, and does not smooth  $u$  sufficiently to make  $J[s]$  smooth in  $s$ , for arbitrary data  $z \in F$ . The inclusion of  $\partial/\partial t$  in the definition of  $\tilde{\phi}_\sigma$ , on the other hand, induces exactly the necessary amount of regularity in  $u[s]$  to make  $J[s]$  smooth.

Next we shall compute the Hessian, i.e. second derivative, of  $J[s]$ . To accomplish this, we need to compute the derivative of  $u[s]$ , which satisfies the normal equations, written in yet another way:

$$-\left( \frac{\partial^2}{\partial t^2} + \sigma^2 \frac{\partial^2}{\partial x^2} \right) u[s] = -\frac{\partial}{\partial t} T_s^T z.$$



Therefore

$$-\left(\frac{\partial^2}{\partial t^2} + \sigma^2 \frac{\partial^2}{\partial x^2}\right) \frac{d}{ds} u[s] = -x \frac{\partial}{\partial t} \left(-\frac{\partial}{\partial t} T_s^T z\right).$$

On the other hand,

$$\begin{aligned} x \frac{\partial}{\partial t} \left[ -\left(\frac{\partial^2}{\partial t^2} + \sigma^2 \frac{\partial^2}{\partial x^2}\right) u[s] \right] &= \\ &= -\left(\frac{\partial^2}{\partial t^2} + \sigma^2 \frac{\partial^2}{\partial x^2}\right) \left(x \frac{\partial}{\partial t} u[s]\right) + \sigma^2 \left[x, \frac{\partial^2}{\partial x^2}\right] \frac{\partial}{\partial t} u[s] \\ &= -\left(\frac{\partial^2}{\partial t^2} + \sigma^2 \frac{\partial^2}{\partial x^2}\right) \left(x \frac{\partial}{\partial t} u[s]\right) - 2\sigma^2 \frac{\partial^2}{\partial x \partial t} u[s] \\ &= x \frac{\partial}{\partial t} \left(-\frac{\partial}{\partial t} T_s^T z\right) \end{aligned}$$

Comparing equations, we see that

$$-\left(\frac{\partial^2}{\partial t^2} + \sigma^2 \frac{\partial^2}{\partial x^2}\right) \left(\frac{d}{ds} u[s] + x \frac{\partial}{\partial t} u[s]\right) = -2\sigma^2 \frac{\partial^2}{\partial x \partial t} u[s]$$

Let  $N_\sigma$  be the inverse of  $-(\partial^2/\partial t^2 + \sigma^2 \partial^2/\partial x^2) : H_b^1 \rightarrow H_b^{-1}$ . Then we have

$$\frac{d}{ds} u[s] = -x \frac{\partial}{\partial t} u[s] - 2\sigma^2 N_\sigma \frac{\partial^2}{\partial x \partial t} u[s].$$

Differentiating the result of Theorem 2.5, we have

$$\begin{aligned} \frac{d^2}{ds^2} J[s] &= -2\sigma^2 \left\{ \left\langle \frac{\partial}{\partial t} \frac{d}{ds} u[s], \frac{\partial}{\partial x} u[s] \right\rangle + \left\langle \frac{\partial}{\partial t} u[s], \frac{\partial}{\partial x} \frac{d}{ds} u[s] \right\rangle \right\} \\ &= -2\sigma^2 \left\{ \left\langle \frac{\partial}{\partial t} \left(-x \frac{\partial}{\partial t} - 2\sigma^2 N_\sigma \frac{\partial^2}{\partial x \partial t}\right) u[s], \frac{\partial}{\partial x} u[s] \right\rangle \right. \\ &\quad \left. + \left\langle \frac{\partial}{\partial t} u[s], \frac{\partial}{\partial x} \left(-x \frac{\partial}{\partial t} - 2\sigma^2 N_\sigma \frac{\partial^2}{\partial x \partial t}\right) u[s] \right\rangle \right\} \\ &= -2\sigma^2 \left\{ \left\langle -x \frac{\partial}{\partial t} \left(\frac{\partial}{\partial t} u[s]\right), \frac{\partial}{\partial x} u[s] \right\rangle \right\} \end{aligned}$$

$$\begin{aligned}
& + \left\langle \frac{\partial}{\partial t} u[s], -x \frac{\partial}{\partial t} \left( \frac{\partial}{\partial x} u[s] \right) \right\rangle + \left\langle -\frac{\partial}{\partial t} u[s], \frac{\partial}{\partial t} u[s] \right\rangle \Big\} \\
& + 4\sigma^4 \left\{ \left\langle \frac{\partial}{\partial t} N_\sigma \frac{\partial^2}{\partial x \partial t} u[s], \frac{\partial}{\partial x} u[s] \right\rangle \right. \\
& \left. + \left\langle \frac{\partial}{\partial t} u[s], \frac{\partial}{\partial x} N_\sigma \frac{\partial^2}{\partial x \partial t} u[s] \right\rangle \right\} \\
& = 2\sigma^2 \left\| \frac{\partial}{\partial t} u[s] \right\|^2 - 8\sigma^4 \left\langle \frac{\partial^2}{\partial x \partial t} u[s], N_\sigma \frac{\partial^2}{\partial x \partial t} u[s] \right\rangle \\
& = 2\sigma^2 \left\| \frac{\partial}{\partial t} u[s] \right\|^2 + 8\sigma^4 \left\langle \frac{\partial}{\partial x} u[s], \frac{\partial}{\partial t} N_\sigma \frac{\partial}{\partial t} \frac{\partial}{\partial x} u[s] \right\rangle
\end{aligned}$$

where we have used the antisymmetry of  $x \frac{\partial}{\partial t}$  as before. Since both terms define continuous quadratic forms in  $z \in F$ , we have proved

**Theorem 2.6**

$$\begin{aligned}
\frac{d^2}{ds^2} J_\sigma[s] & = 2\sigma^2 \left\| \frac{\partial}{\partial t} u[s] \right\|^2 \\
& + 8\sigma^4 \left\langle \frac{\partial}{\partial x} u[s], \frac{\partial}{\partial t} N_\sigma \frac{\partial}{\partial t} \frac{\partial}{\partial x} u[s] \right\rangle .
\end{aligned}$$

The next results give together a fairly precise characterization of the shape of  $J_\sigma$ .

**Corollary 2.7** *Suppose that  $u[s] \not\equiv 0$ ,  $\frac{\partial}{\partial x} u[s] \equiv 0$ . Then  $s$  is a strict local minimizer of  $J_\sigma[s]$  for any  $\sigma > 0$ .*

*Proof.* From Theorem 2.5,  $\frac{d}{ds} J[s] = 0$ , while from Theorem 2.6,

$$\frac{d^2}{ds^2} J[s] = 2\sigma^2 \left\| \frac{\partial u[s]}{\partial t} \right\|^2 > 0 ,$$

since  $u[s] \not\equiv 0$  and  $u[s]$  has mean zero.

q.e.d.

The periodic 1-variable Sobolev spaces  $H_{per}^s$  are defined to be the completions of  $C_{per}^\infty[0, T]$  in the norms

$$\|u\|_s^2 = \sum_n \left(1 + \frac{4n^2\pi^2}{T^2}\right)^s |\hat{u}_n|^2.$$

**Corollary 2.8** *Suppose that the datum  $z \in F$  is consistent, i.e. for some  $s^* \in \mathbb{R}$ ,  $u^* \in H_{per}^1$ ,  $z = \frac{\partial}{\partial t} T_{s^*} u^*$ . Then  $s^*$  is a strict local minimizer of  $J_\sigma$  for all  $\sigma > 0$ .*

*Proof.* Indeed if we set  $u(x, t) = u^*(t)$  then evidently  $u = u[s^*]$ ,  $\partial u[s^*]/\partial x \equiv 0$ , and  $J[s^*] = 0$ , so the conclusion follows from the last corollary. q.e.d.

It is convenient to introduce notation for norm ratios: for  $u \in \bigcup_s H_{per}^s$  define

$$R_{s,t}[u] = \frac{\|u\|_s}{\|u\|_t}$$

whenever the latter makes sense.

**Theorem 2.9** *Suppose that the datum  $z$  is consistent, i.e.  $z = \frac{\partial}{\partial t} T_{s^*} u^*$  for  $s^* \in \mathbb{R}$ ,  $u^* \in H_{per}^1$ . Then with  $R_{s,t} := R_{s,t}[u^*]$ ,*

(a) *If additionally  $u^* \in H_p^{\frac{3}{2}}[0, T]$ , then*

$$\begin{aligned} & \left( \|z\| - \text{const.} \left\{ \frac{R_{0,1}^2}{|s-s^*|} + \frac{R_{\frac{3}{2},1}^2}{\sigma} \right\}^{\frac{1}{2}} \|u^*\|_1 \right)^2 \\ & \leq J_\sigma[s] \leq \left( \|z\| + \text{const.} \left\{ \frac{R_{0,1}^2}{|s-s^*|} + \frac{R_{\frac{3}{2},1}^2}{\sigma} \right\}^{\frac{1}{2}} \|u^*\|_1 \right)^2 \end{aligned}$$

(b)

$$\left| \frac{dJ_\sigma}{ds}[s] \right| \geq \sigma^2 |s - s^*| \|u^*\|_1^2 \left\{ \frac{8}{\pi^2} - \text{const.} R_{0,1}^2 - \text{const.} \sqrt{\sigma} R_{\frac{1}{2},1} \right\}$$

(c) If  $\sqrt{\sigma} R_{\frac{1}{2},1}$  and  $R_{0,1}$  are sufficiently small, then  $J_\sigma$  is quasiconvex over  $[-1,1]$ .

*Remark.* Before giving the proof of Theorem 2.9, we discuss its meaning. The norm ratios  $R_{s,t}$  measure *oscillation*. For example

$$R_{s,t} \left[ \sin \frac{n\pi t}{T} \right] = \left| \frac{n\pi}{T} \right|^{s-t}.$$

This gives us a way to interpret the results in terms of the shortest and longest significant wavelengths in an essentially bandlimited target signal  $u^*$ .

The bound (a) means:  $|J_\sigma[s] - \|z\|^2|$  is small, provided that  $\sigma$  is much larger than the largest frequency (upper bandlimit), and that  $|s - s^*|$  is larger than the longest wavelength (corresponding to the lower bandlimit). That is, as  $\sigma \rightarrow \infty$ ,  $J_\sigma$  becomes flat if the slope  $s$  is “more than a wavelength in error”. This is exactly the behaviour observed in Section 1 for the mean square error function, and justifies the intuitive notion that  $J_\sigma$  must approximate the mean square error for large  $\sigma$ .

The bound (b) states that  $J_\sigma$  has only one critical point, so long as the longest wavelength in  $u^*$  is sufficiently short. Moreover the  $\sqrt{\sigma} R_{\frac{1}{2},1}$  term suggests that larger  $\sigma$  can be allowed for shorter wavelengths. Otherwise put:  $J_\sigma$  is quasiconvex so long as  $\sigma \approx 1$  and  $u^*$  is *sufficiently oscillatory*.

*Proof.* Assume for the time being that  $u^*$ , hence  $u[s]$ , is smooth. The result as stated will follow from continuity.

As in the proof of Theorem 2.1,

$$\frac{\partial}{\partial t} u[s](x, t) = \sum_{\substack{\mathbf{k} \geq \mathbf{0} \\ n = -\infty}}^{\infty} \left[ \frac{\partial}{\partial t} u[s] \right]_{\mathbf{k},n}^{\wedge} \cos \left( \frac{k\pi x}{2x} + \frac{k\pi}{2} \right) e^{\frac{2\pi i n t}{T}}$$

so

$$\frac{\partial}{\partial t} u[s](\pm X, t) = \sum_{\substack{\mathbf{k} \geq \mathbf{0} \\ n = -\infty}}^{\infty} (\pm 1)^k \left[ \frac{\partial}{\partial t} u[s] \right]_{\mathbf{k},n}^{\wedge} e^{\frac{2\pi i n t}{T}}.$$

Here for  $n \neq 0$ ,

$$\left[ \frac{\partial}{\partial t} u[s] \right]_{\mathbf{k},n}^{\wedge} = \frac{1}{XT(1 + \delta_{0k})} \int_{-X}^X dx \int_0^T dt \cos \left( \frac{k\pi x}{2X} + \frac{k\pi}{2} \right) e^{-\frac{2\pi i n t}{T}} \frac{\partial}{\partial t} u[s](x, t)$$

$$\begin{aligned}
&= \left[ \frac{\partial}{\partial t} N_\sigma \left( T_{-s} \frac{\partial^2}{\partial t^2} T_{s^*} u^* \right) \right]_{kn}^\wedge \\
&= \frac{2\pi i n}{T} \left( \frac{n^2 \pi^2}{T^2} + \sigma^2 \frac{k^2 \pi^2}{4X^2} \right)^{-1} \frac{4n^2 \pi^2}{T^2} A_{kn} \hat{u}_n^* \\
&= \frac{2\pi i n}{T} \left( 1 + \sigma^2 \frac{T^2}{16X^2} \frac{k^2}{n^2} \right)^{-1} A_{kn} \hat{u}_n^*
\end{aligned}$$

Of course  $\left[ \frac{\partial u[s]}{\partial t} \right]_{k0} \equiv 0$ , so the sum on  $n$  implicitly avoids  $u = 0$ .

$$\begin{aligned}
A_{kn} &= \frac{1}{2X} \int_{-X}^X dx \left\{ \exp \left[ i\pi \left( \frac{2n(s^* - s)}{T} + \frac{k}{2X} \right) x + \frac{k\pi}{2} \right] \right. \\
&\quad \left. + \exp \left[ i\pi \left( \frac{2n(s^* - s)}{T} - \frac{k}{2X} \right) x - \frac{k\pi}{2} \right] \right\} \\
&= \frac{i^k}{2} \operatorname{sinc} \left( \pi \left( \frac{2n(s^* - s)X}{T} + \frac{k}{2} \right) \right) \\
&\quad + \frac{i^{-k}}{2} \operatorname{sinc} \left( \pi \left( \frac{2n(s^* - s)X}{T} + \frac{k}{2} \right) \right)
\end{aligned}$$

Evidently,  $|A_{kn}| \leq 1$  uniformly. Thus

$$\begin{aligned}
\left\| \frac{\partial}{\partial t} u[s] \right\|^2 &= \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} \sum_{k=0}^{\infty} \left| \left[ \frac{\partial}{\partial t} u[s] \right]_{kn}^\wedge \right|^2 \\
&\leq \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} \frac{4n^2 \pi^2}{T^2} |\hat{u}_n^*|^2 \sum_{k=0}^{\infty} \left( 1 + \sigma^2 \frac{T^2}{16X^2} \frac{k^2}{n^2} \right)^{-2} |A_{kn}|^2 \\
&\leq \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} \frac{4n^2 \pi^2}{T^2} |\hat{u}_n^*|^2 |A_{0n}|^2
\end{aligned}$$

$$+ \frac{\text{const.}}{\sigma} \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} \frac{n^3 \pi^3}{T^3} |\hat{u}_n^*|^2$$

by the simple domination of a sum over  $k$  by an integral. On the other hand

$$|A_{0n}| \leq \frac{1}{n|s - s^*|}, \quad n \neq 0$$

so we conclude that for  $s \neq s^*$

$$\begin{aligned} \left\| \frac{\partial}{\partial t} u[s] \right\|^2 &\leq \text{const.} \left\{ \frac{\|u^*\|_0}{|s - s^*|^2} + \frac{\|u^*\|_{\frac{3}{2}}^2}{\sigma} \right\} \\ &\leq \text{const.} \|u^*\|_1^2 \left\{ \frac{R_{0,1}^2}{|s - s^*|} + \frac{R_{\frac{3}{2},1}^2}{\sigma} \right\}. \end{aligned}$$

Similarly

$$\frac{\partial u[s]}{\partial x}(x, t) = \sum_{n=-\infty}^{\infty} \hat{u}_n^* e^{\frac{2xint}{T}} \sum_{k=0}^{\infty} \frac{k\pi}{2X} \left(1 + \sigma \frac{T^2}{4X^2} \frac{k^2}{n^2}\right)^{-1} A_{kn} \sin\left(\frac{k\pi x}{2X} + \frac{k\pi}{2}\right).$$

The family  $\left\{ \sin\left(\frac{k\pi x}{2X} + \frac{k\pi}{2}\right) \right\}$  is also orthogonal on  $[-X, X]$  so

$$\begin{aligned} \left\| \frac{\partial u[s]}{\partial x} \right\|^2 &= \sum_{n=-\infty}^{\infty} |\hat{u}_n^*|^2 \sum_{k=0}^{\infty} \frac{k^2 \pi^2}{16X^2} \left(1 + \sigma^2 \frac{T^2}{4X^2} \frac{k^2}{n^2}\right)^{-2} |A_{kn}|^2 \\ &\leq \frac{\text{const.}}{\sigma^3} \|u^*\|_{\frac{3}{2}}^2 = \frac{\text{const.}}{\sigma^3} R_{\frac{3}{2},1}^2 \|u^*\|_1^2 \end{aligned}$$

by another simple integral majorization. The first part of the theorem follows immediately *via* the triangle inequality.

Since  $Tz^*$  is a function of  $t + (s^* - s)x$  it follows that

$$O = - \left( \frac{\partial^2}{\partial t^2} + \sigma^2 \frac{\partial}{\partial x^2} \right) \left( \frac{\partial}{\partial x} - (s^* - s) \frac{\partial}{\partial t} \right) u[s].$$

Since  $u[s]$  obeys the homogeneous Neumann condition at  $x = \pm X$ , we can view

$$v := \left( \frac{\partial}{\partial x} - (s^* - s) \frac{\partial}{\partial t} \right) u[s]$$

as the solution of the boundary value problem

$$-\left( \frac{\partial^2}{\partial t^2} + \sigma^2 \frac{\partial^2}{\partial x^2} \right) v \equiv 0 \quad \text{in } [-X, X] \times [0, T]$$

$$v(\pm X, \cdot) = (s^* - s) \frac{\partial u[s]}{\partial t} (\pm X, \cdot)$$

$v$  is  $T$ -periodic in  $t$ .

Straightforward estimation shows that

$$\|v\|^2 \leq C \sigma (s^* - s)^2 \left\| \frac{\partial u[s]}{\partial t} \Big|_{x=\pm X} \right\|_{-\frac{1}{2}}^2$$

Using the Fourier expansion developed previously, we can write

$$\begin{aligned} \frac{\partial u[s]}{\partial t} (\pm X, t) &= \sum_{n=-\infty}^{\infty} \sum_{k \geq 0} \left[ \frac{\partial u[s]}{\partial t} \right]_{kn}^{\wedge} e^{\frac{2\pi i n t}{T}} \cos \left( \frac{k\pi x}{2X} + \frac{k\pi}{2} \right) \Big|_{x=\pm X} \\ &= \sum_{n=-\infty}^{\infty} e^{\frac{2\pi i n t}{T}} \sum_{k \geq 0} (\mp 1)^k \left[ \frac{\partial u[s]}{\partial t} \right]_{kn}^{\wedge} \end{aligned}$$

so

$$\begin{aligned} \left\| \frac{\partial u[s]}{\partial t} (\pm X, \cdot) \right\|_{-\frac{1}{2}}^2 &= \\ &= \sum_{n=-\infty}^{\infty} \frac{4n^2\pi^2}{T^2} |\hat{u}_n^*|^2 \left| \sum_{k \geq 0} (\mp 1)^k \left( 1 + \sigma \frac{T^2}{16X^2} \frac{k^2}{n^2} \right)^{-1} A_{kn} \right|^2 \\ &= \sum_{n=-\infty}^{\infty} \frac{n^2\pi^2}{T^2} |\hat{u}_n^*|^2 \left| \sum_{k=-\infty}^{\infty} (\mp i)^k \left( 1 + \sigma^2 \frac{T^2}{16X^2} \frac{k^2}{n^2} \right)^{-1} \right. \\ &\quad \left. \operatorname{sinc} \left( \pi \left( \frac{2n(s^* - s)X}{T} + \frac{k}{2} \right) \right) \right|^2. \end{aligned}$$

For each  $n \in Z$ , denote by  $k[n]$  the unique  $k \in Z$  satisfying

$$\left| k + \frac{4n(s - s^*)X}{T} \right| < \frac{1}{2}.$$

Split the sum over  $k$  above into three pieces, according to whether  $k - k[n]$  is zero, odd, or even:

$$\begin{aligned} & \sum_{k=-\infty}^{\infty} (\mp i)^k \left( 1 + \sigma^2 \frac{T^2}{16X^2} \frac{k^2}{n^2} \right)^{-1} \operatorname{sinc} \left( \pi \left( \frac{2n(s^* - s)X}{T} + \frac{k}{2} \right) \right) \\ &= (\mp i)^{k[n]} \left\{ \left( 1 + \sigma^2 \frac{T^2}{16X^2} \frac{k[n]^2}{n^2} \right)^{-1} \operatorname{sinc} \pi \lambda[n] \right. \\ & \quad - \frac{\sin \pi \lambda[n]}{\pi} \sum_{\substack{k=-\infty \\ k \neq 0}}^{\infty} \left[ \left( 1 + \sigma^2 \frac{T^2}{16X^2} \frac{(k[n] + 4k)^2}{n^2} \right) (\lambda[n] + 2k + 1)^{-1} \right. \\ & \quad \quad \left. - \left( 1 + \sigma^2 \frac{T^2}{16X^2} \frac{(k[n] + 4k + 2)^2}{n^2} \right) (\lambda[n] + 2k + 1)^{-1} \right] \\ & \quad \left. \pm i \frac{\sin \pi (\lambda[n] + \frac{1}{2})}{\pi} \sum_{k=-\infty}^{\infty} \left[ \left( 1 + \sigma^2 \frac{T^2}{16X^2} \frac{(k[n] + 4k + 1)^2}{n^2} \right)^{-1} (\lambda[n] + 2k + \frac{1}{2})^{-1} \right. \right. \\ & \quad \quad \left. \left. - \left( 1 + \sigma^2 \frac{T^2}{16X^2} \frac{(k[n] + 4k + 3)^2}{n^2} \right) (\lambda[n] + 2k + \frac{3}{2})^{-1} \right] \right\}. \end{aligned}$$

Here

$$\lambda[n] = \frac{2(s - s^*)X}{T} + \frac{k[n]}{2}$$

(so  $|\lambda[n]| < \frac{1}{4}$ , all  $n$ ), and the sums have been “unrolled” a bit take advantage of the periodicity of  $\sin$ .

One easily sees that

$$\left| \frac{k[n]T}{4Xn} + (s - s^*) \right| < \frac{T}{8Xn}$$

whence follows a uniform bound on the first term.



Set

$$f(x) = \left(1 + \frac{\sigma^2 T^2}{16X^2} \frac{(k[n] + 2x)^2}{n^2}\right)^{-1} (\lambda[n] + x)^{-1}.$$

Then the second sum can be written

$$\sum_{\substack{k=-\infty \\ k \neq 0}}^{\infty} f(2k) - f(2k + 1)$$

and the third as

$$\sum_{k=-\infty}^{\infty} f\left(2k + \frac{1}{2}\right) - f\left(2k + \frac{3}{2}\right).$$

The arguments of  $f$  in the sums avoid the interval  $(-\frac{1}{2}, \frac{1}{2})$ . Outside of this interval

$$|f(x)| < 4.$$

In addition to the pole at  $x = -\lambda[n]$ ,  $f$  has at most two critical points, so the two half-axes  $(-\infty, \frac{1}{2}]$  and  $[\frac{1}{2}, \infty)$  are divided into at most four subintervals on each of which  $f(x)$  is monotone. Thus each sum can be grouped into at most four subsums, each of which is an alternating sum of a monotone sequence. Each of these is estimated by the following simple principle: Suppose  $\{a_n\}$  is a monotonic sequence with  $a_- \leq a_n \leq a_+$ , all  $n$ . Then

$$\left| \sum_{n=0}^{\infty} a_{2n+1} - a_{2n+2} \right| \leq a_+ - a_-.$$

Consequently

$$\left| \sum_{\substack{k=-\infty \\ k \neq 0}}^{\infty} f(2k) - f(2k + 1) \right| \leq 32$$

and a similar bound holds for the third term. In view of the previous estimates of  $k[n]$ ,  $\lambda[n]$  this gives a uniform bound on the sums. We obtain

$$\left\| \frac{\partial u[s]}{\partial t}(\pm X, \cdot) \right\|^2 \leq \text{const.} \|u^*\|_1^2.$$

Similarly,

$$\left\| \frac{\partial u[s]}{\partial t}(\pm X, \cdot) \right\|_{-\frac{1}{2}}^2 \leq \text{const.} \|u^*\|_{\frac{1}{2}}$$

whence

$$\begin{aligned} \frac{dJ_\sigma}{ds}[s] &= -2\sigma^2 \left\langle \frac{\partial u[s]}{\partial t}, \frac{\partial u[s]}{\partial x} \right\rangle \\ &= -2\sigma^2 (s - s^*) \left\| \frac{\partial u[s]}{\partial t} \right\|^2 \\ &\quad - 2\sigma^2 \left\langle \frac{\partial u[s]}{\partial t}, \frac{\partial u[s]}{\partial x} - (s - s^*) \frac{\partial u[s]}{\partial t} \right\rangle \\ \left| \frac{dJ_\sigma}{ds}[s] \right| &\geq 2\sigma^2 |s - s^*| \left\| \frac{\partial u[s]}{\partial t} \right\|^2 \\ &\quad - \text{const.} \sigma^{\frac{5}{2}} |s - s^*| \|u^*\|_1 \|u^*\|_{\frac{1}{2}} \end{aligned}$$

From the Fourier expansion of  $\frac{\partial u[s]}{\partial t}$  derived before,

$$\begin{aligned} \left\| \frac{\partial u[s]}{\partial t} \right\|^2 &= \frac{1}{2} \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} \frac{4n^2 \pi^2}{T^2} |\hat{u}_n^*|^2 \\ &\quad \sum_{k=-\infty}^{\infty} \left| \text{sinc} \left( \pi \left( \lambda[n] + \frac{k}{2} \right) \right) \right|^2 \left( 1 + \sigma^2 \frac{T^2}{16X^2} \frac{(k[n] + k)^2}{n^2} \right)^{-2} \\ &\geq \frac{1}{2} \sum_{n=-\infty}^{\infty} \frac{4n^2 \pi^2}{T^2} |u_n^*|^2 |\text{sinc} \pi \lambda[n]|^2 \left( 1 + \sigma^2 \frac{T^2 k[n]}{16X^2 n^2} \right)^{-2} \end{aligned}$$

simply by dropping all terms with  $k \neq 0$ . From the definition of  $k[n]$ ,

$$|\pi \lambda[n]| = \left| \pi \left( \frac{k[n]}{2} + \frac{2n(s - s^*)X}{T} \right) \right| < \frac{\pi}{4}$$

so

$$|\text{sinc} \pi \lambda[n]| \geq \frac{2\sqrt{2}}{\pi}.$$

Also

$$\left(1 + \sigma^2 \frac{T^2 k[n]}{16X^2 n^2}\right)^{-2} \geq 1 - O\left(\frac{1}{n^2}\right)$$

where the  $O(\frac{1}{n^2})$  bound is uniform over any finite interval of  $\sigma$ . Therefore

$$\begin{aligned} \left\|\frac{\partial u[s]}{\partial t}\right\|^2 &\geq \frac{4}{\pi^2} \left\|\frac{\partial u^*}{\partial t}\right\|^2 - \text{const.} \|u\|_0^2 \\ &\geq \left(\frac{4}{\pi^2} - \text{const.} \mathcal{R}_{0,1}^2\right) \|u^*\|_1^2. \end{aligned}$$

These estimates together establish the second part of the theorem. q.e.d.

The next theorem establishes properties (ii) and (iii) of an acceptable estimator for the output of a local optimization algorithm applied to  $J_\sigma$ , as discussed in the introduction.

**Theorem 2.10** *Suppose that  $0 < \sigma$  and that  $u^* \in H_{per}^1$  is sufficiently oscillatory relative to  $\sigma$ , as in Theorem 2.9 (b), (c). Let  $z \in F$  and*

$$e = \frac{\|z - z^*\|}{\|z^*\|}$$

*be the relative error level, where  $z^* = \phi[s^*, u^*]$ . If  $s$  is a critical point of  $J_\sigma$ , then*

$$|s - s^*| \leq \text{const.} \frac{e}{\sigma}.$$

*Proof.* Recall that  $u[s]$  is the solution of

$$-\left(\frac{\partial^2}{\partial t^2} + \sigma^2 \frac{\partial^2}{\partial x^2}\right) u = -\frac{\partial}{\partial t} T_{-s} z$$

plus boundary conditions. Denote by  $u_0[s]$  the solution of the same problem with  $z$  replaced by  $z^*$ . From Theorem 2.1,

$$\begin{aligned} &\|u[s] - u_0[s]\|_{1,\sigma} \\ &= \left( \left\|\frac{\partial}{\partial t}(u[s] - u_0[s])\right\|^2 + \sigma^2 \left\|\frac{\partial}{\partial x}(u[s] - u_0[s])\right\|^2 \right)^{1/2} \\ &\leq \|z - z^*\|. \end{aligned}$$

Thus

$$\begin{aligned}
\frac{d}{ds} J_\sigma[s] &= -\sigma^2 \left\langle \frac{\partial u[s]}{\partial t}, \frac{\partial u[s]}{\partial x} \right\rangle \\
&= -2\sigma^2 \left\{ \left\langle \frac{\partial u}{\partial t} (u[s] - u_0[s]), \frac{\partial u[s]}{\partial x} \right\rangle \right. \\
&\quad + \left\langle \frac{\partial u_0[s]}{\partial t}, \frac{\partial u[s]}{\partial x} - \frac{\partial u_0[s]}{\partial x} \right\rangle \\
&\quad \left. + \left\langle \frac{\partial u_0[s]}{\partial t}, \frac{\partial u_0[s]}{\partial x} \right\rangle \right\}.
\end{aligned}$$

Denote by  $J_\sigma^*[s]$  the functional  $J_\sigma$  with data  $z^*$ . Then using Theorem 2.1 again,

$$\left| \frac{d}{ds} J_\sigma[s] - \frac{d}{ds} J_\sigma^*[s] \right| \leq 2\sigma \|z - z^*\| (\|z\| + \|z^*\|) \leq \text{const.} \sigma e \|u^*\|_1^2.$$

As noted before, if  $u^*$  is sufficiently oscillatory, the quantity in Theorem 2.9(b) obeys

$$\frac{8}{\pi^2} - \text{const.} \mathcal{R}_{0,1} - \text{const.} \sqrt{\sigma} \mathcal{R}_{\frac{1}{2},1} := \text{const.} > 0.$$

Combining Theorem 2.9(b) with the above inequalities we obtain

$$\left| \frac{d}{ds} J_\sigma[s] \right| \geq \text{const.} (\sigma^2 |s - s^*| - \sigma e) \|u^*\|_1^2,$$

so if  $s$  is a critical point of  $J_\sigma$ , then

$$|s - s^*| \leq \text{const.} \frac{e}{\sigma}.$$

q.e.d.

Up to this point we have concentrated entirely on the extent to which  $s$  can be estimated. Concerning the waveform  $u$ , we have the following result, the proof of which is straightforward:

**Corollary 2.11** *Let the hypotheses be as in Theorem 2.10 and suppose that  $s = \operatorname{argmin} J_\sigma[s]$ . Then the estimator*

$$U(t) = \frac{1}{2X} \int_{-X}^X dx \frac{\partial u[s]}{\partial t}(x, t)$$

*satisfies*

$$\int_0^T dt \left| \int_0^t dt' (u^*(t') - U(t')) \right|^2 < \operatorname{const.} \frac{\epsilon}{\sigma}.$$

*Remark.* That is, the estimate of  $u$  is Lipschitz continuous with loss of one derivative. It is easy to see that this result is sharp.

We have illustrated the foregoing results through numerical experiments. We used the data displayed in Figures 1.4 and 1.5, which led to the optimal mean-square error curves in Figures 1.6 and 1.7. Figure 2.1 shows the differential semblance objective function  $J_\sigma$  for the data of Figure 1.4, using the weight  $\sigma = 1$ . This result is directly comparable to that of Figure 1.6, and the contrast in Figure 2.1 is essentially quadratic. Figure 2.2 displays the analogous result for the noisy data of Figure 1.5, using  $\sigma = 1$ . While the values of the objective function are very much greater (by an order of magnitude), and the curve much shallower, the general shape remains quadratic with the correct minimum.

One might expect a larger error in the estimate of  $s$  on the basis of Theorem 2.10. Apparently the completely incoherent nature of the data noise in Figure 1.5 leads to anomalously small effect on the differential semblance estimator. To show that coherent noise has a much stronger effect, we use the data in Figure 2.3 (the sum of Figure 1.4 and its mirror image) to produce the plot of  $J_\sigma$  in Figure 2.4. The curve is still quite convex, but the minimizer has moved to  $s = 0$  (unsurprisingly). While these examples are suggestive, the precise influence of the nature of noise on the differential semblance estimator remains to be explained.

Our final examples illustrate the conclusion embodied in Theorem 2.9(a), that as  $\sigma \rightarrow \infty$  the differential semblance objective approximates the output least squares objective, and in particular tends to become saturated away from the global minimizer. Figures 2.5 and 2.6 show the function  $J_\sigma$  for  $\sigma = 10$ , using the data of Figures 1.4 and 1.5 respectively. The minimum is more sharply defined, as is suggested by Theorems 2.9 and 2.10, and the

curves flatten out away from the global minimizer. Even in this example, though,  $J_\sigma$  remains quasiconvex and smooth, and could easily be minimized by a properly designed quasi-Newton algorithm.

## References

1. DENNIS, J. and SCHNABEL, R. [1983]. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice Hall, Englewood Cliffs, New Jersey.
2. GILBARG, D. and TRUDINGER, N. [1983]. *Elliptic partial differential equations of second order*, Springer-Verlag, New York.
3. MENKE, W. [1984]. *Geophysical Data Analysis: Discrete Inverse Theory*, Academic Press, New York.
4. SYMES, W. [1990]. The plane wave detection problem, Technical Report 90-1, Department of Mathematical Sciences, Rice University, Houston, TX 77251.
5. SYMES, W. [1991a]. Layered velocity inversion: a model problem from reflection seismology, *SIAM J. Math. Anal.* 22, 680-716.
6. SYMES, W. [1991b]. The plane wave detection problem: a nonlinear ill-posed problem, in: G. Roach, ed., *Inverse Problems and Imaging*, Longman Scientific and Technical, John Wiley & Sons, New York.
7. SYMES, W. [1991c]. The reflection inverse problem for acoustic waves, in: G. Cohen, L. Halpern, P. Joly, eds., *Mathematical and Numerical Aspects of Wave Propagation Phenomena*, SIAM, Philadelphia, pp. 423-433.
8. SYMES, W. and CARAZZONE, J. [1991]. Velocity inversion by differential semblance optimization, *Geophysics* 56, 654-663.
9. TAPIA, R. [1978]. Quasi-Newton methods for equality constrained optimization: equivalence of existing methods and a new implementation, in: *Nonlinear Programming*, vol. 3, Academic Press, New York.
10. TARANTOLA, A. [1987]. *Inverse Problem Theory*, Elsevier, New York.
11. TAYLOR, M. [1980]. *Pseudodifferential operators*, Princeton University Press, Princeton, New Jersey.

## Captions for figures

Figure 1.1 Plot of the reduced least squares objective

$$J_{opt}[s] = \inf_{u \in L^2[0,T]} \|\phi[s, u] - z\|^2$$

where

$$z(x, t) = \sin(\omega t)$$

and  $\omega = 50.0$ . Here  $-0.5 < x < 0.5, 0.0 < t < 1.0$  and the values of  $J_{opt}$  are computed analytically as explained in the text. Since the mean square residual has already been optimized over the waveform  $u$  here, the many local minima of  $J_{opt}$  actually represent local minima of the output least-squares objective function.

Figure 1.2 Same as 1.1 with  $\omega = 500.0$ .

Figure 1.3 A Ricker wavelet (second derivative of a Gaussian pulse) with peak frequency 20 Hz.

Figure 1.4 Plane wave data with  $u =$  Ricker waveform of Figure 1.3, slope  $s = 0.1$ .

Figure 1.5 Same as 1.4, with 160% added pseudorandom noise filtered with the waveform of Figure 1.3.

Figure 1.6  $J_{opt}$  for the noise-free data of Figure 1.4.

Figure 1.7  $J_{opt}$  for the noisy data of Figure 1.5; note the appearance of local optima.



Figure 2.1 The DSO objective function for the noise-free data of Figure 1.4,  $\sigma = 1$ .

Figure 2.2 The DSO objective function for the noisy data of Figure 1.5,  $\sigma = 1$ .

Figure 2.3 Dataset formed by summing Figure 1.4 with its mirror image; thus composed of the data for  $s = 0.1$  and  $s = -0.1$  in equal measure. Either may be regarded as coherent noise.

Figure 2.4 The DSO objective function for the “coherent noise” data of Figure 2.3,  $\sigma = 1$ .

Figure 2.5 Counterpart of Figure 2.1 with  $\sigma = 10$ .

Figure 2.6 Counterpart of Figure 2.2 with  $\sigma = 10$ .

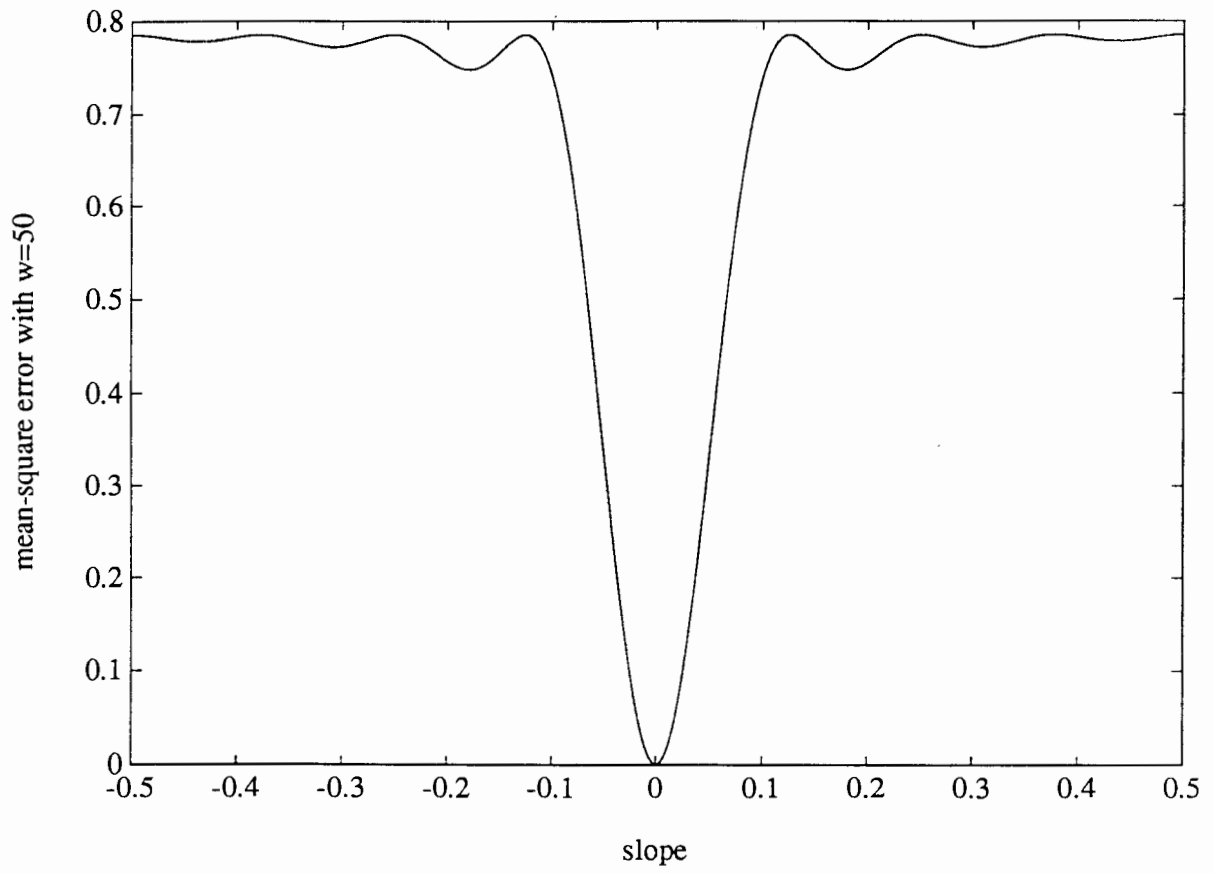


Figure 1.1

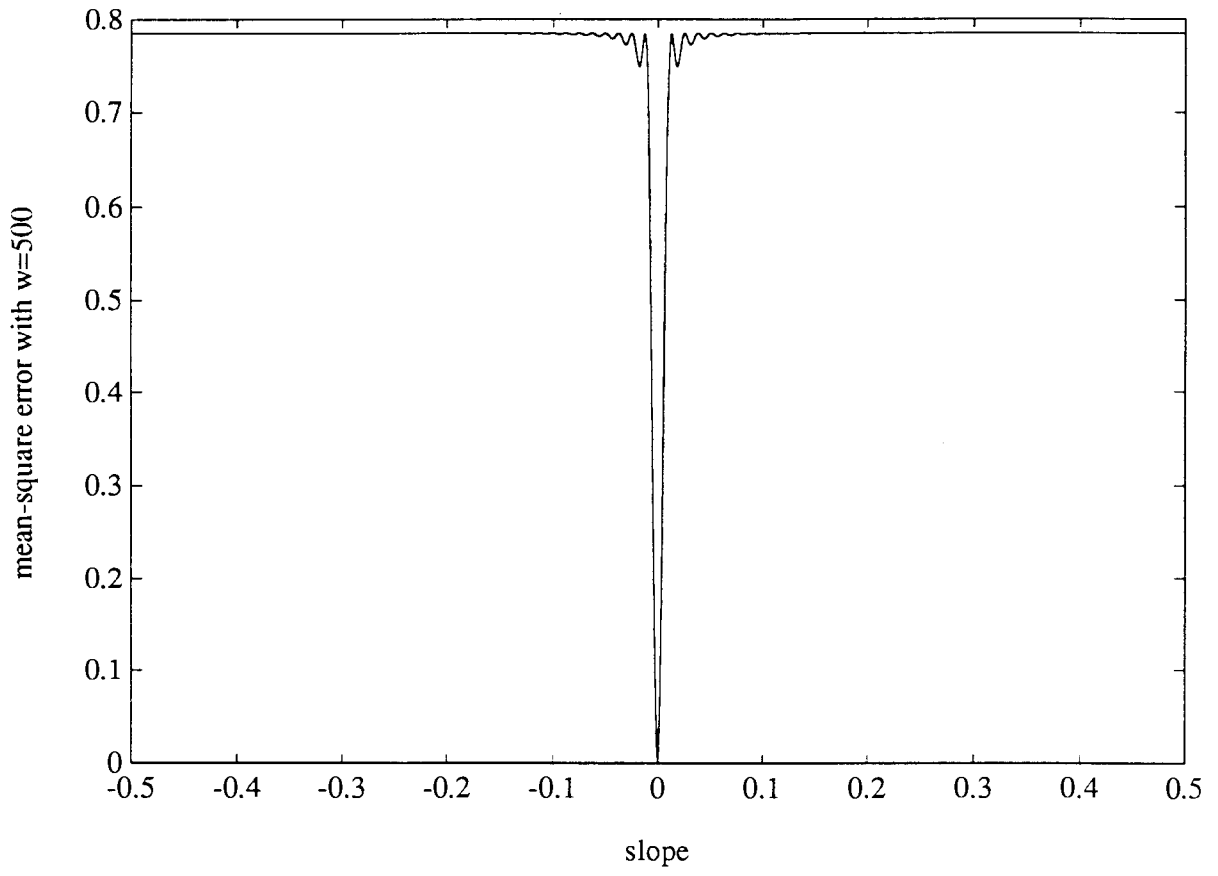


Figure 1.

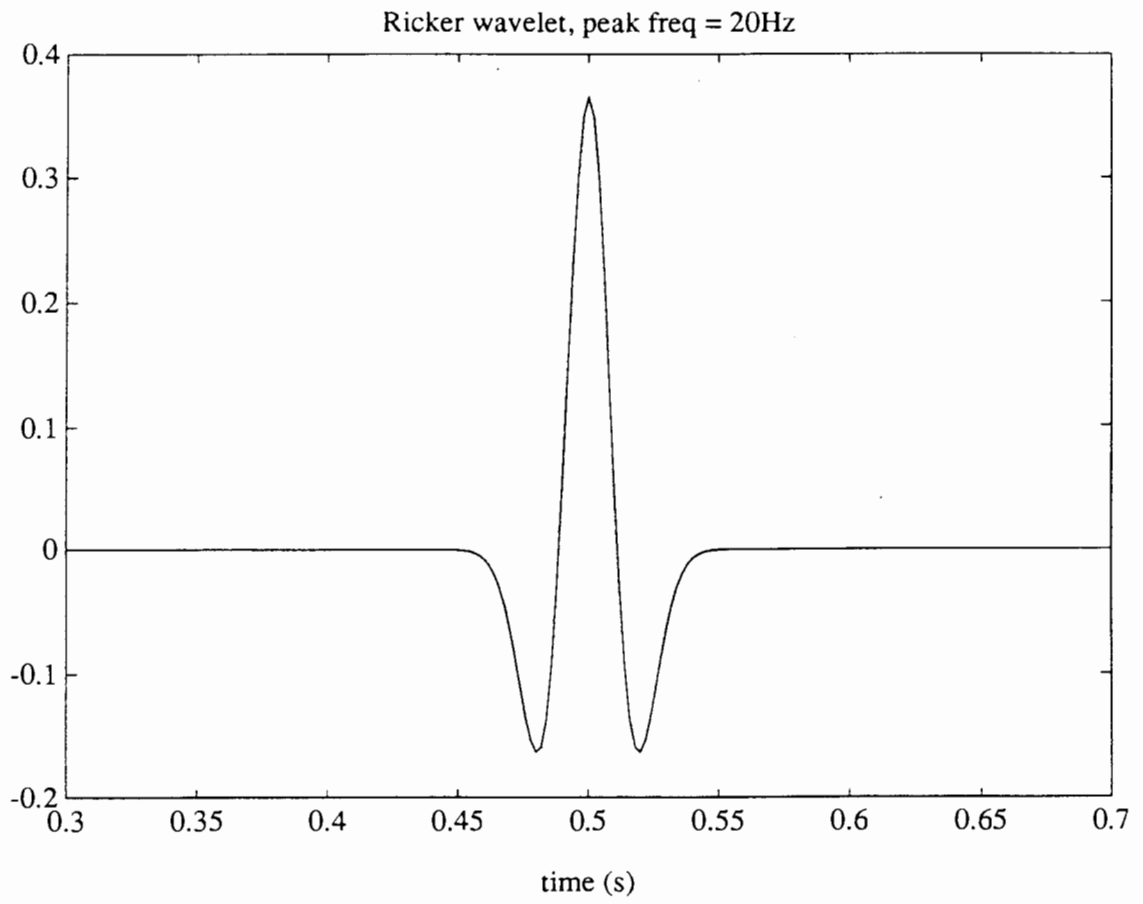


Figure 1.3

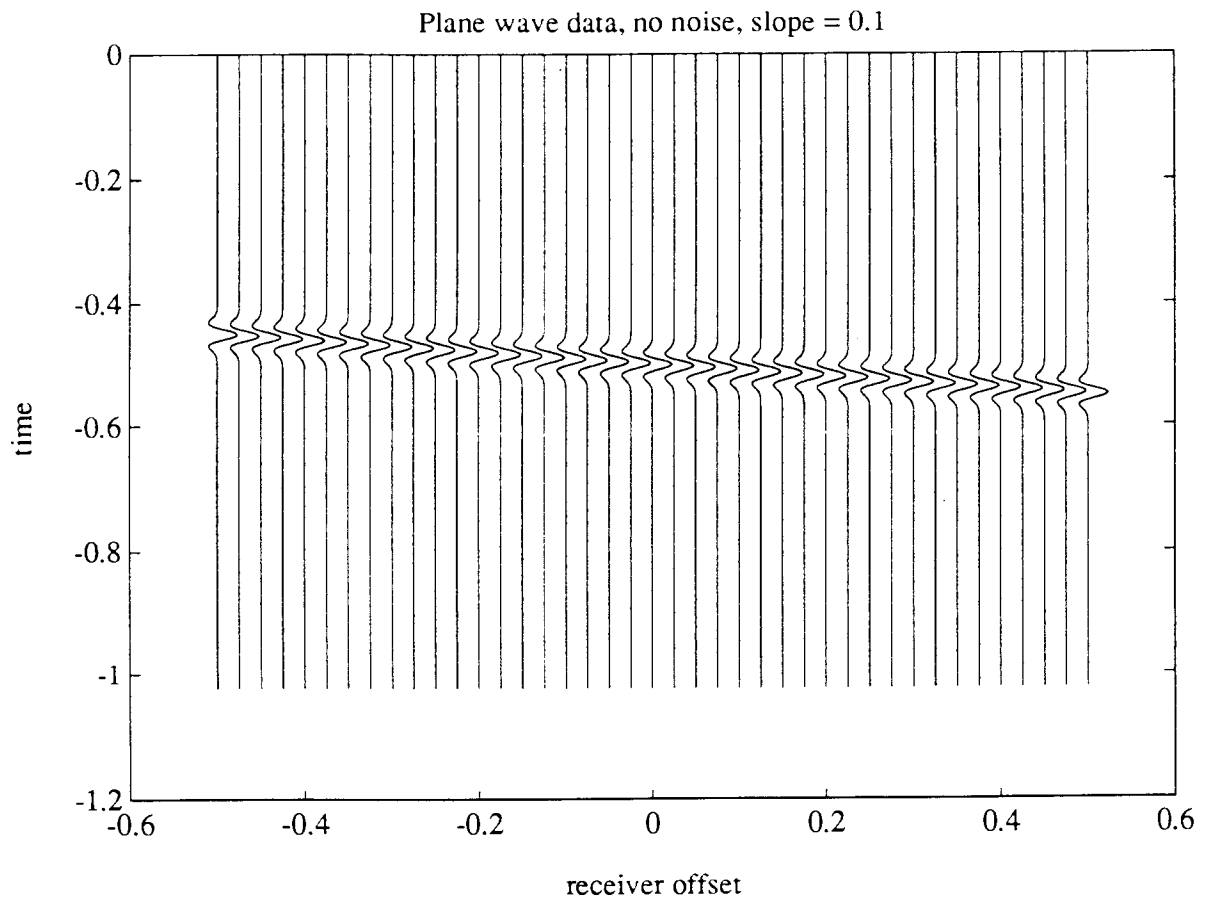


Figure 14

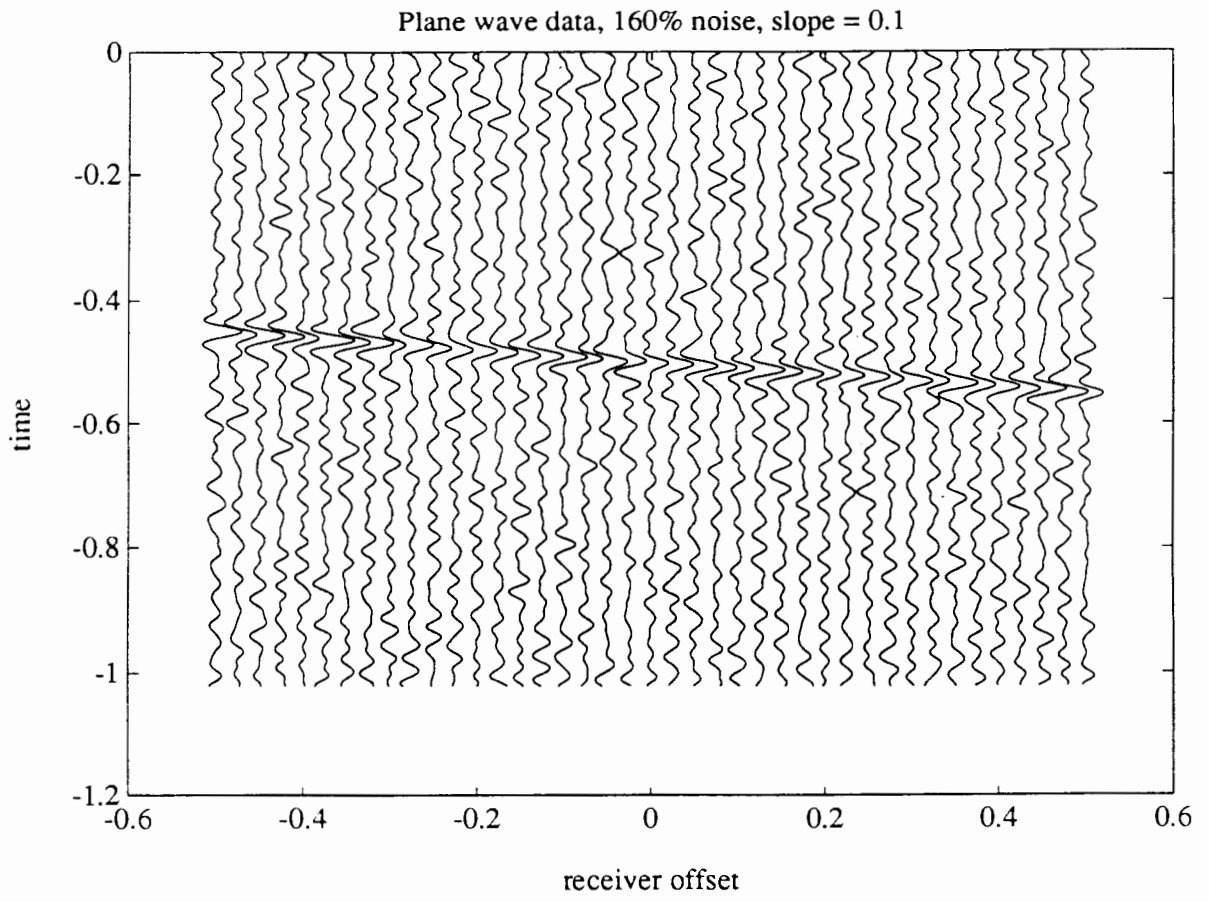


Figure 1.5

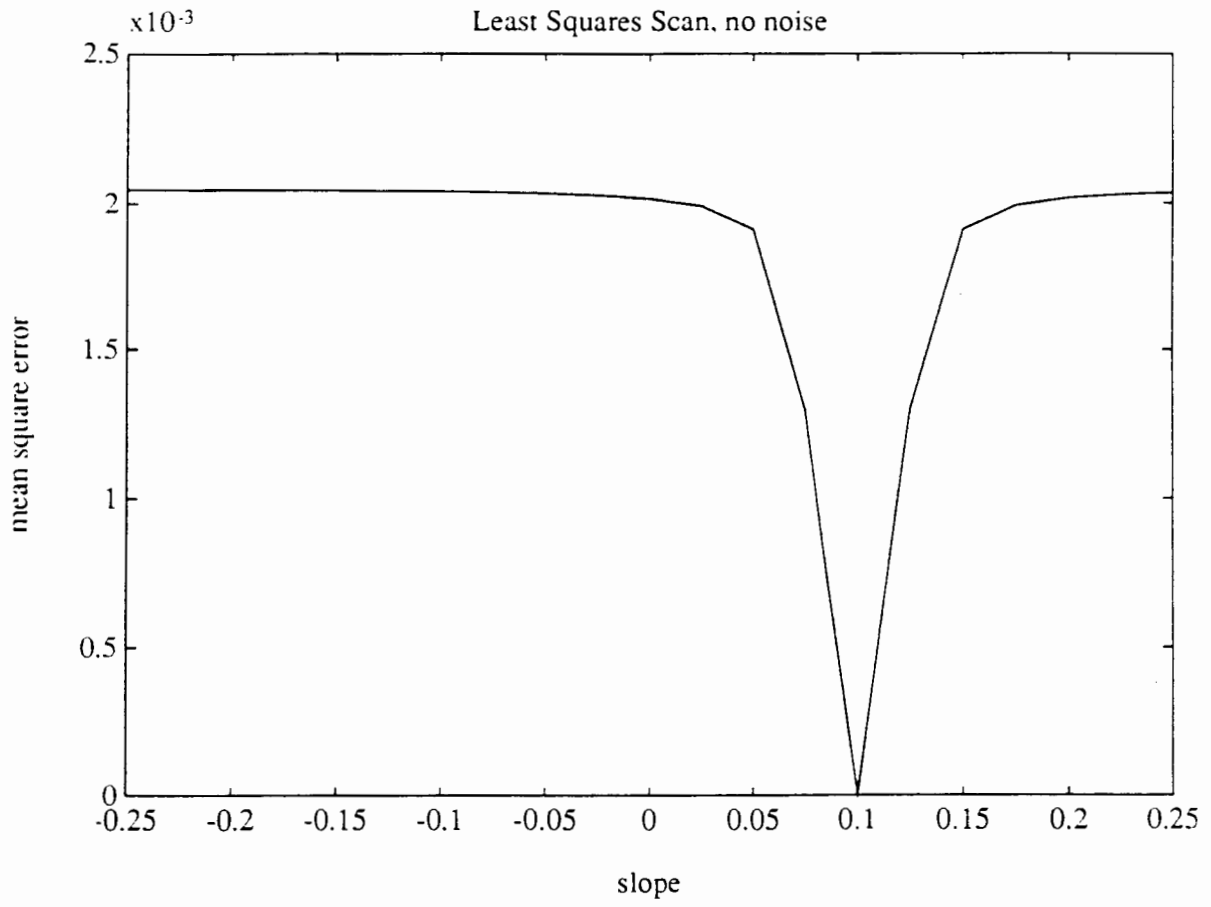


Figure 1.1

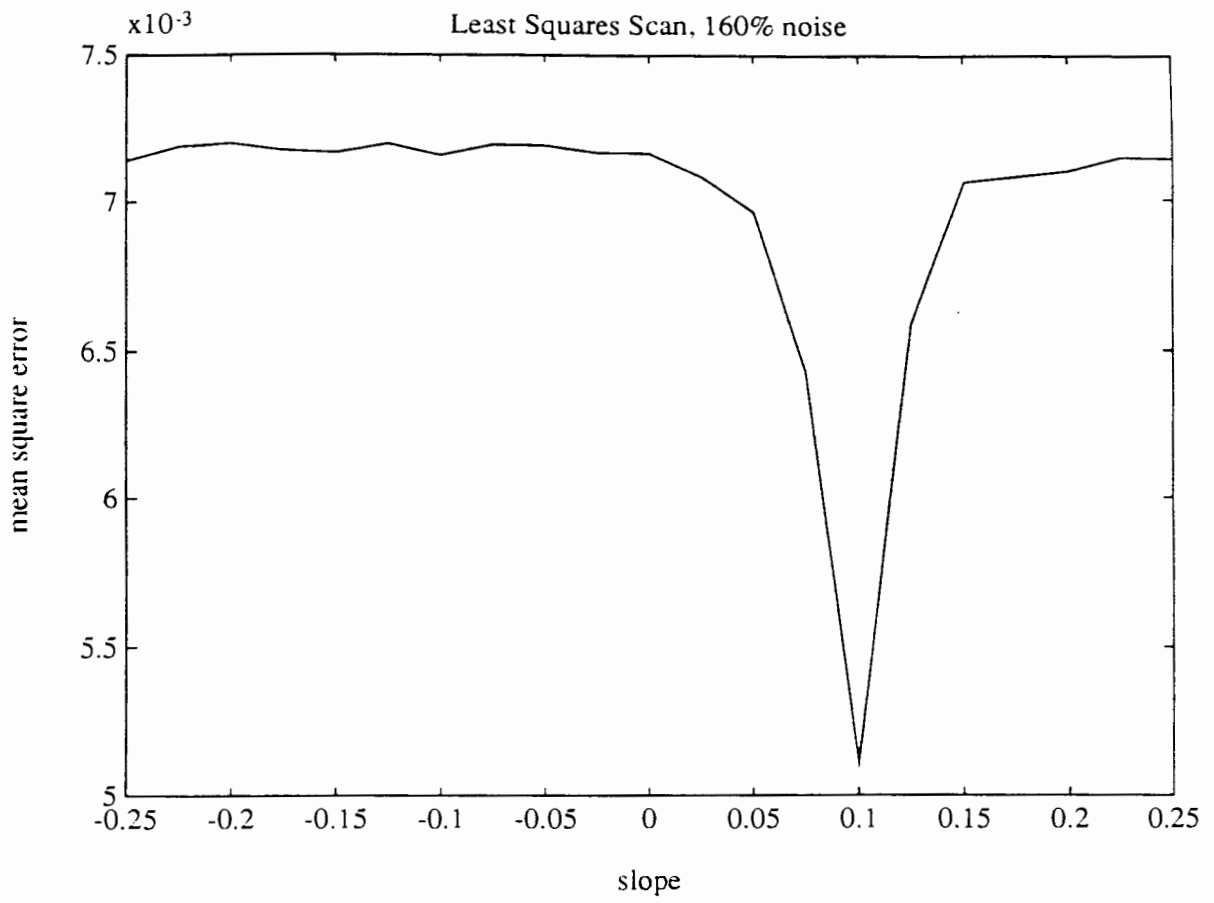


Figure 1.7



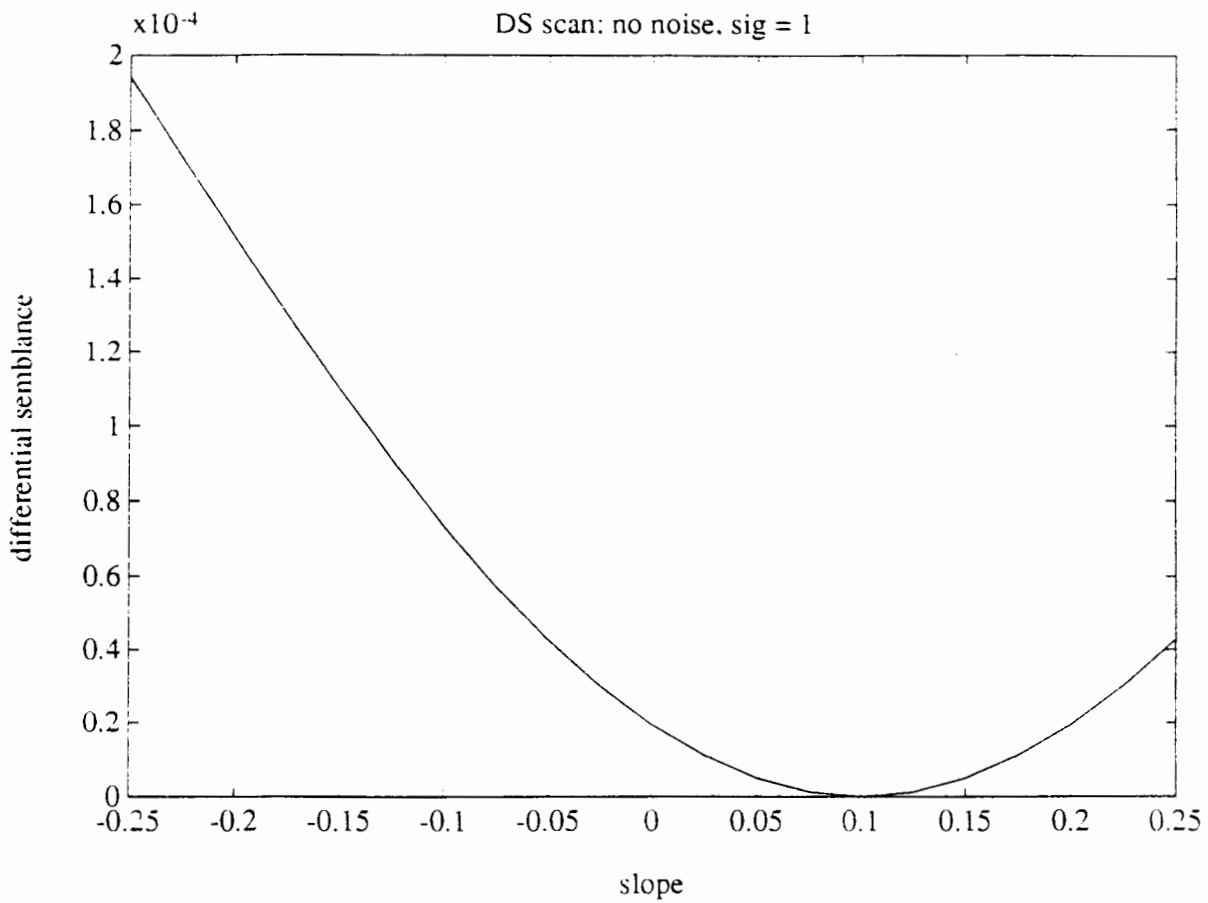


Figure 2.1

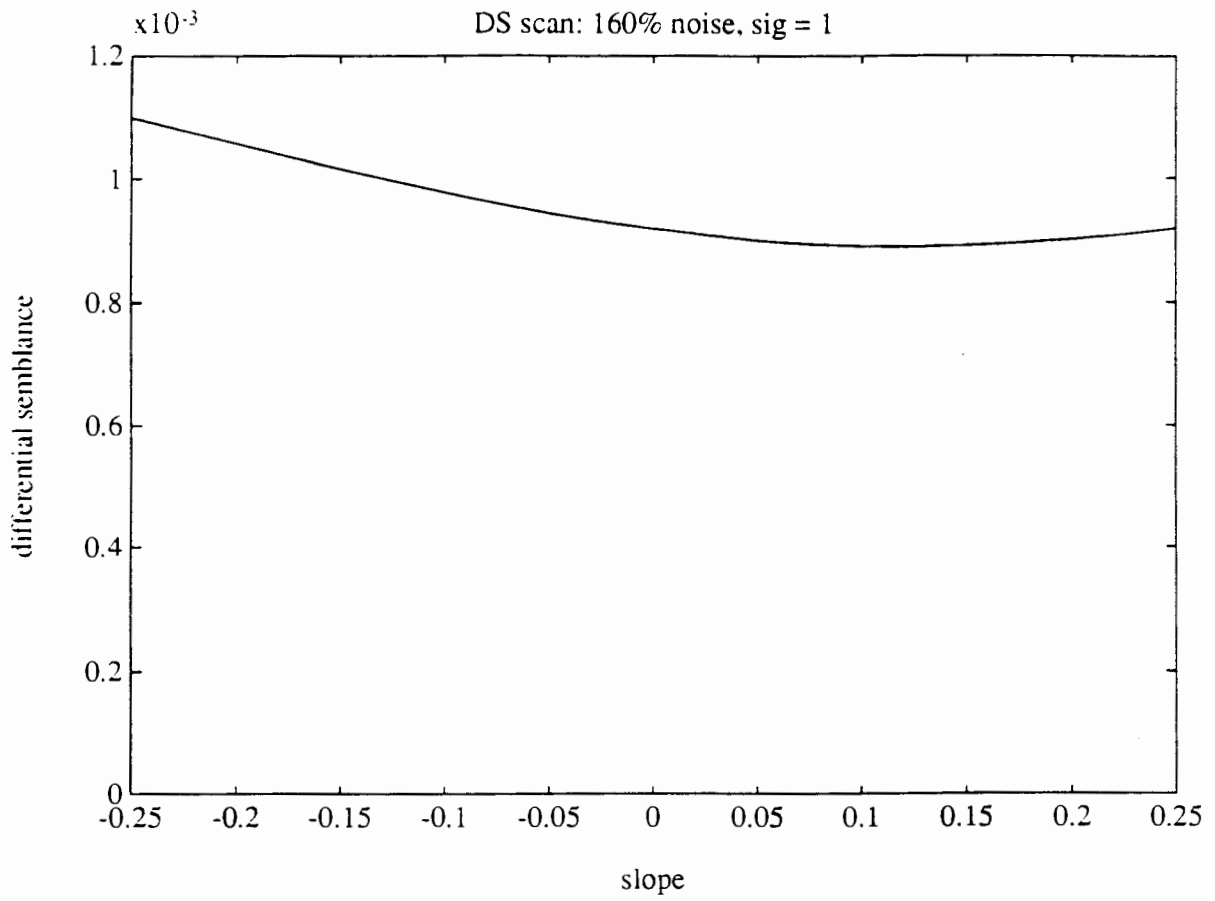


Figure 2.1

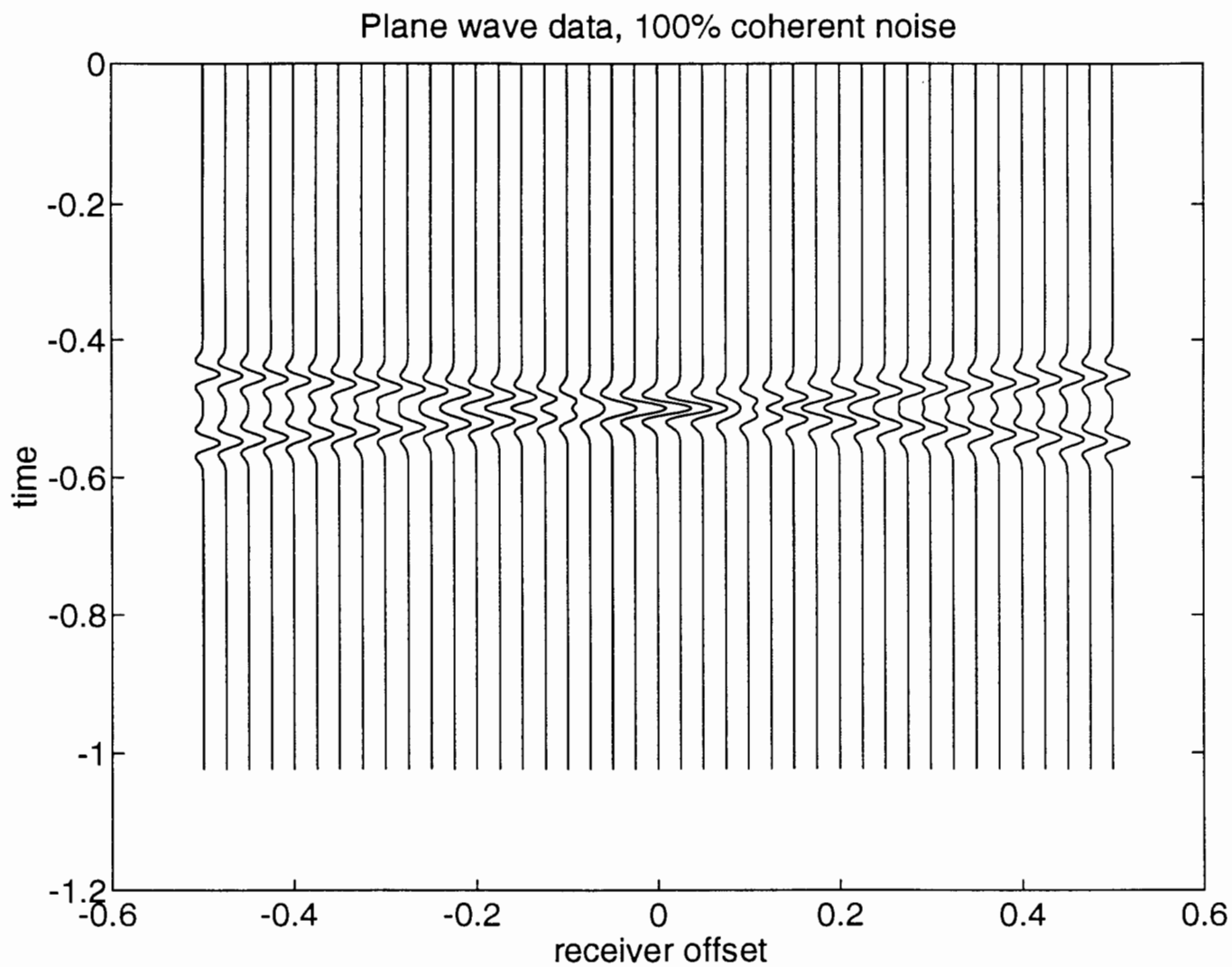


Figure 2.3

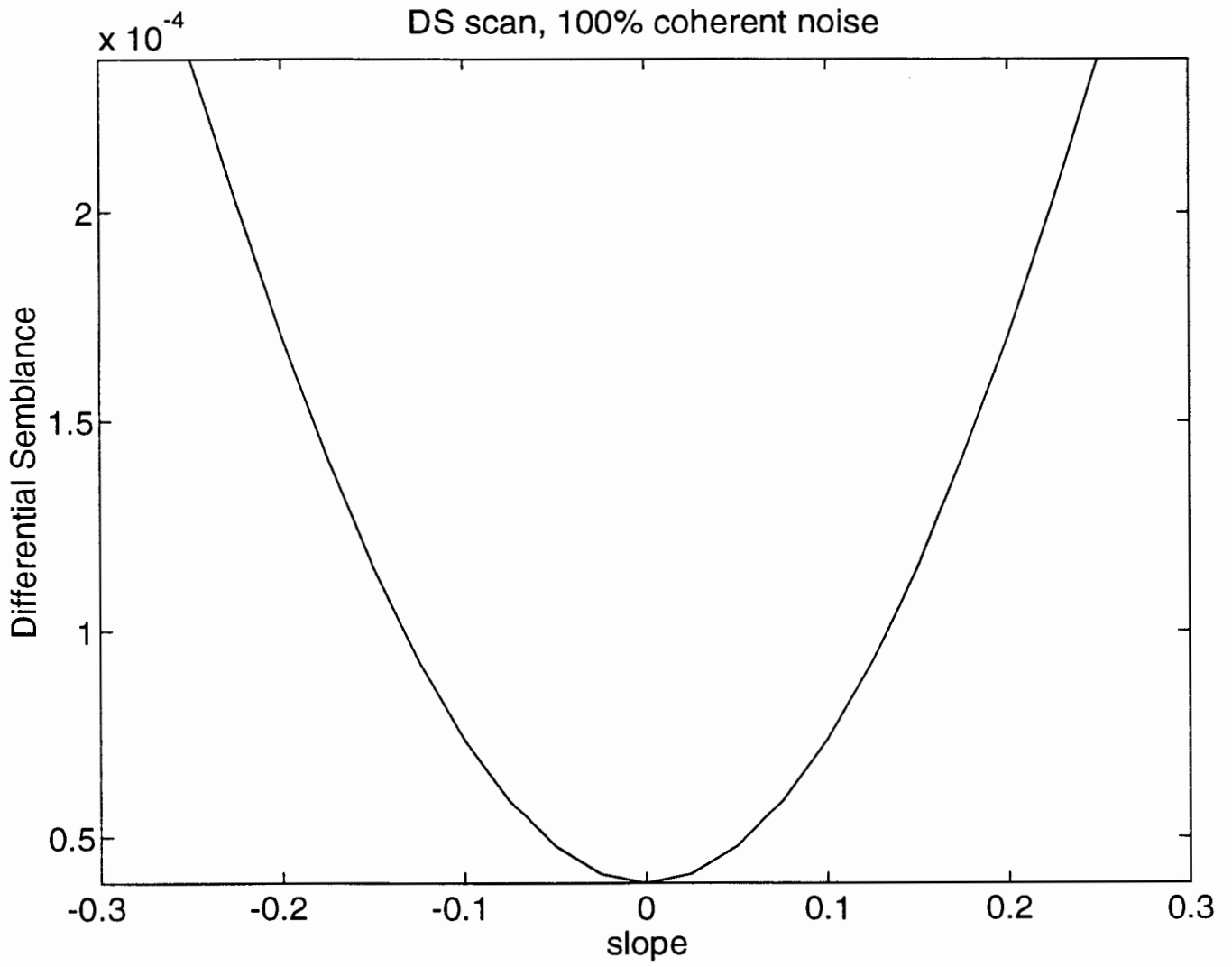


Figure 2.4

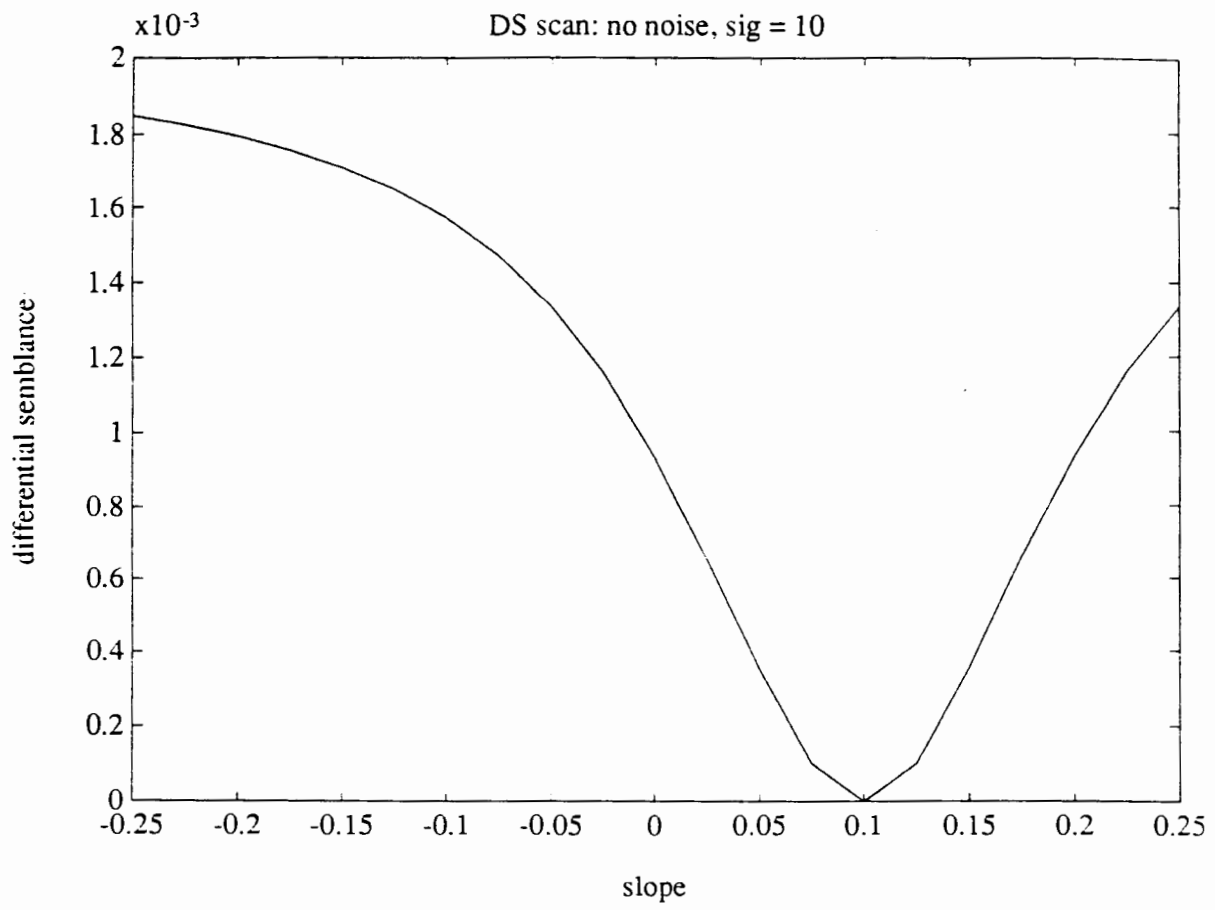


Figure 2. 5

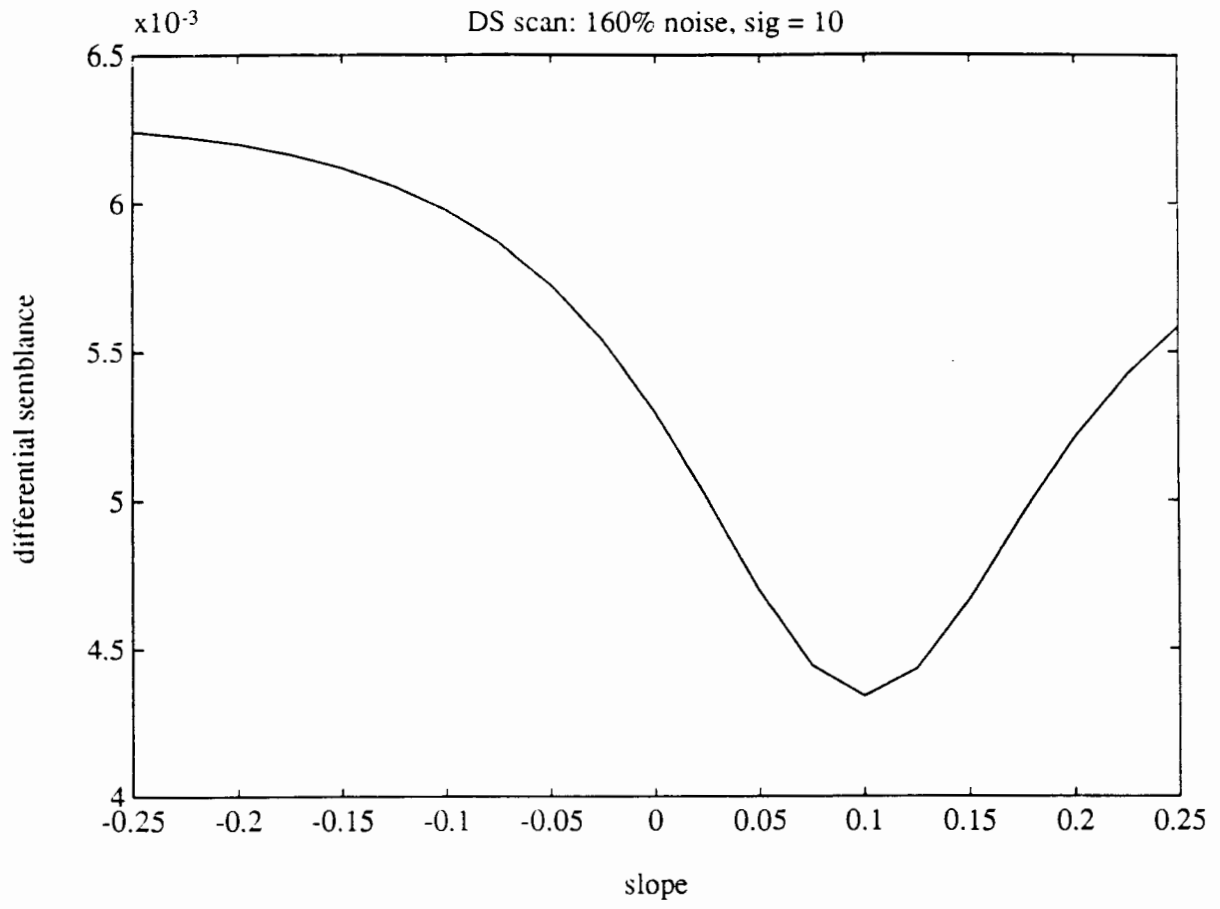


Figure 2.6