

Velocity Inversion: A Case Study in
Infinite-Dimensional Optimization

William W. Symes

July, 1988

TR88-14

Department of Mathematical Sciences
Rice University
P. O. Box 1892
Houston, Texas 77251

Velocity Inversion: a Case Study in Infinite-Dimensional Optimization

William W. Symes
Rice University

July 1988

Abstract.

The goal of seismic velocity inversion is the estimation of seismic wave velocities inside the earth by attempting to predict, in a least-error sense, seismic waveforms measured at its surface. We present velocity inversion as a case study in the various "infinite-dimensional" pathologies which may afflict practically important problems of distributed parameter identification, treated as optimization problems in function spaces. These features differentiate various problem formulations far beyond the degree one would expect for finite- (small-) dimensional problems.

1 Introduction

The estimation of seismic wave velocities within the earth from data measured on its surface is a fundamental preoccupation of seismology. The version of this problem natural to reflection seismology has recently come under study by many academic and industrial scientists around the world as a nonlinear least-error problem, in which a velocity model is sought to minimize some measure of the misfit between a measured seismogram and one predicted from the model (by means of linear elastodynamics, for instance). This direct, “model-based” approach to velocity estimation goes under the name “velocity inversion”, to distinguish it from older, less comprehensive techniques embedded in conventional seismic data-processing technology.

This problem, when set within the linear-elastodynamics model of seismic wave propagation, is infinite-dimensional: both the model and the data are fields, i.e. functions of one or several variables. Of course, sample rates of actual measurements are finite, but are rapid enough that the study of the continuum problem provides an accurate guide to the behaviour of reasonable discretizations.

In this paper we discuss three distinct but related approaches to the velocity inversion problem. All are based on closely related variational principles of least-squares type, but these principles have very different analytical properties. These analytical differences are essentially consequences of the infinite-dimensional nature of the velocity inversion problem, and have a profound bearing on the efficiency of computational algorithms. We show that the first of these approaches (“output-least-squares”), the subject of most work in this area, is intrinsically ill-suited to numerical optimization, whereas the third approach (“coherency optimization”), a reformulation of output-least-squares, amounts to a smooth, convex minimization problem. We give both analytical arguments and numerical evidence to support these assertions.

Thus the central goal of this paper is to illustrate, by means of a nontrivial example, that both the functional-analytic structure and the application-derived phenomenology of an infinite-dimensional optimization problem may have profound consequences for computation, and may even force a complete reformulation of the problem in the interests of computational efficiency, or even feasibility.

The investigation of the velocity inversion problem, at a level of detail sufficient to justify these conclusions, unavoidably requires the introduction of a mathematical description of the map relating the model (velocity distribution in the earth, in this case) to the data (seismogram), to which we devote section 2. For the sake of conceptual simplicity, we work in the context of layered acoustics, with plane-wave sources and signals, and we use the linearized approximation to the high-frequency components of the waveform, derived from geometric optics.

This approximation simplifies the mathematics considerably, but not essentially: generalization to less restrictive classes of propagating media and experimental models may be had at the price of more complex arguments. As we wish to emphasize the optimization-theoretic aspects of the velocity inversion problem, we will remain with its simplest setting.

The first approach to velocity inversion, discussed in section 3, is the “output-least-squares” formulation, which is the subject of a rapidly growing literature. In this approach, a model is sought which predicts a data set (seismogram) minimizing the mean-square error with a given (“measured”) data set. It turns out that, in the ideal case of infinite signal bandwidth, the objective function (mean-square error of fit) is either nowhere differentiable or has a Hessian with spectrum accumulating at zero, depending on the choice of norms. In the more realistic case of finite signal bandwidth, the objective inherits these features in the form of extreme non-convexity, as well as severe ill-conditioning of local quadratic models. The deleterious consequences for computation have been noted in the literature, and have been a major factor in preventing this approach to velocity estimation from gaining a place in the seismic data processing stream.

As we show in section 4, the objective function can be rendered better-behaved by enlarging the space of models, at the price of adding a constraint. Norms can be chosen so that this second “constrained least-squares” version of the velocity inversion problem involves only smooth functions, and satisfies the “second-order sufficiency condition” at zero-residual optima: i. e. the objective Hessian is positive-definite on the null space of the linearized constraints. Unfortunately, the so-called “constraint qualification” fails, so that the constraint set may fail to be a submanifold — at every point! We illustrate this pathology with a simple, closely related quadratic variational problem with linear constraints.

The illustrative quadratic problem becomes a convex unconstrained problem with well-conditioned Hessian when the constraint is incorporated as a “soft” L^2 -penalty term instead of as a hard constraint. This observation motivates the introduction in section 5 of the third approach to the velocity inversion problem, based on the enlarged model space with an L^2 -penalty treatment of the constraint mentioned above. This is a “nice” problem: we briefly review some theoretical results, which establish the convexity of the penalized objective in a “large” region, for near-consistent data, and for a constructively estimated range of penalty parameters. We have solved this problem numerically using a model-trust-region Gauss-Newton algorithm, in which the linear subproblem is solved by conjugate residual iteration, truncated in the fashion suggested by Steihaug (1981) and Toint (1981). Iterative construction of the Newton step is mandatory, as the dimension of the model space exceeds 10^4 in our simplified examples, and will lie in the range 10^5 – 10^6 in applications to field data sets. The numerical tests are quite successful, conform closely to the predictions of the

theory, and suggest several as-yet unresolved theoretical and practical questions.

A good general reference for inverse problems in geophysics, treated as optimization problems, is the book by Tarantola (1987). Other references are scattered throughout the following sections.

2 The Convolutional Model for Plane-Wave Reflection Seismograms

This approximation to the linear elastodynamics description of the reflection seismogram has been of fundamental importance in the development of contemporary seismic data processing technology; see for example Robinson and Treitel (1980). A development of the multiple-plane-wave version used here is given in Santosa and Symes (1988).

It is assumed that the earth is adequately modeled as a horizontally stratified fluid with constant density. Thus the local mechanics of the model are described entirely by the *sound velocity*, which is a function of the depth z only. Further, it is assumed that the sound velocity is the *sum* of a slowly varying (smooth) *background velocity* $c(z)$ and a (possibly) rapidly varying, or oscillatory, *reflection series* $\Delta c(z)$. This hypothesis is consistent with direct measurements (*sonic logs*) of sound velocity in the upper crust — see e.g. Waters (1981), Figure 4.3A.

Recall that a *plane wave* with velocity c and wave-vector \vec{p} in space-time \mathfrak{R}^4 is a disturbance of the form

$$u(\vec{x}, t) = f(ct - \vec{p} \cdot \vec{x})$$

When c is constant, such a function solves the wave equation

$$\left(\frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \nabla^2 \right) u(\vec{x}, t) = 0$$

provided that $c^2 |\vec{p}|^2 = 1$, which is assumed. It is further assumed that the earth model occupies all of \mathfrak{R}^3 , and that c is constant for $z (= x_3) < 0$, while the seismogram, i.e. the field u , is measured at the surface $z = 0$ (“seismology at the bottom of an infinitely deep ocean”). Then it can be shown that solutions of the wave equation exist which for $z, t \ll 0$ are plane-waves, as above.

While such an incident-plane-wave experiment cannot be performed in reality, of course, it is generally believed that field data from actual experiments can be massaged to correspond to the data which would be obtained from such an experiment, at least to some extent — see Gutowski, Treitel, and Wagner (1982), Chapman (1978).

The plane-wave seismogram is the record of surface ($z = 0$) motion resulting from an incident plane-wave disturbance moving downward into the layered model. As this incident wave encounters velocity changes, part of its energy is reflected back toward the surface, also as a plane-wave (“echo”). Roughly speaking, the time-of-arrival of the reflected wave at the surface is determined by the depth of the causative velocity inhomogeneity and the velocity distribution between it and the surface, and the amplitude by its strength: these notions are the conceptual basis of reflection seismology.

More precisely: the plane waves are naturally indexed by the angle θ made by the equal-phase plane of the wave with the (horizontal) surface $z = 0$, or more usually by the *slowness* or ray parameter

$$p = \frac{\sin \theta}{c(0)}, \quad \vec{p} = \left(p, 0, \sqrt{\frac{1}{c^2} - p^2} \right)$$

The vertical velocity of the plane wave of slowness p is given by

$$v(z, p) = \frac{c(z)}{\sqrt{1 - c^2(z)p^2}}$$

The *travel time* to depth z and back to the surface is

$$\tau(z, p) = 2 \int_0^z \frac{dz'}{v(z', p)}$$

and its inverse function $\zeta(t, p)$ is defined by

$$t = 2 \int_0^{\zeta(t, p)} \frac{dz'}{v(z', p)}$$

The *plane wave reflectivity* is given by

$$r(t, p) = \frac{\Delta c(z, p)}{c(z)(1 - c^2(z)p^2)} \Big|_{z=\zeta(t, p)}$$

With these notations (a normalized version of) the plane-wave seismogram is given (approximately) by the convolutional expression

$$S(t, p) = \int dt' f(t - t') \frac{\partial r}{\partial t}(t', p)$$

where f is the functional form of the incident plane wave. Usually f is called the *source wavelet*, since it is determined by a model of the energy source, presumed to be acting at such a distance ($z, t \ll 0$) that the wave produced has essentially the plane-wave form.

A very important property of source wavelets is their bandlimited nature. A typical model source wavelet, which gives reasonable-looking seismograms, is depicted in Figure 1, with its power spectrum in Figure 2. This is a so-called Ricker wavelet, peaked at 20 Hertz. Most of the energy content of this wavelet lies at wavelengths which are small compared to the length scale of significant change in the background velocity model. This circumstance underlies the justification of the convolutional model *via* geometric optics. See Santosa and Symes (1988). Note that the above expressions are sensible only when $c(z)|p| < 1$. This is the so-called *precritical* assumption; we refer the reader to the above-cited

references for discussion of its profound physical meaning. For simplicity we will assume *a priori* that the velocity satisfies, for all z ,

$$0 < c_{\min} \leq c(z) \leq c_{\max}$$

and consider only p for which

$$c_{\max}|p| < 1$$

For an attempt to maximize the range of p , see Santosa and Symes (1988).

The above equations complete the description of the convolutional plane-wave model. The extent to which this model reproduces actual seismograms is a subject of much debate. Many qualitative features are certainly predicted rather well, while various others are not. In any case, it captures some part of the information content of reflection seismograms, and is likely a special case of any more complete model.

3 The Output Least-Squares Problem

First we collect the components of the definition of the plane-wave seismogram:

$f(t)$: source wavelet
 $c(z)$: background velocity (smooth)
 $\Delta c(z)$: reflection series (velocity
 perturbation, oscillatory)
 $\tau(z, p)$: travel time

$$\tau(z, p) = 2 \int_0^z dz' \sqrt{\frac{1}{c^2(z')} - p^2}$$

$\zeta(t, p)$: inverse travel time

$$t = 2 \int_0^{\zeta(t, p)} dz' \sqrt{\frac{1}{c^2(z')} - p^2}$$

$r(t, p)$: plane-wave reflectivity

$$r(t, p) = \frac{\Delta c(z)}{c(z)(1 - c^2(z)p^2)} \Big|_{z=\zeta(t, p)}$$

$S[c, \Delta c](t, p)$: plane-wave seismogram

$$S[c, \Delta c](t, p) = \int dt' f(t - t') r \frac{\partial r}{\partial t}(t', p)$$

We have emphasized the functional dependence of S on $c, \Delta c$ by including the latter as arguments of the former.

The *output least-squares* version of the velocity inversion problem is simply the task of finding $(c, \Delta c)$ to best-fit a given plane-wave seismogram $D(t, p)$ (for “data”) in the least-squares sense:

$$\min_{c, \Delta c} J_{LS}[c, \Delta c; D]$$

where

$$\begin{aligned}
 J_{LS}[c, \Delta c; D] &= \|S[c, \Delta c] - D\|^2 \\
 &= \int_0^{t^{\max}} dt \int_0^{p^{\max}} dp |S[c, \Delta c](t, p) - D(t, p)|^2
 \end{aligned}$$

See Tarantola (1987) for a profound attempt to justify this problem formulation on statistical grounds, as well as a discussion of the rationale for *weighted* least-squares. We ignore the issue of weights, as we feel that it adds nothing to the discussion to follow, of the basic mathematical properties of least-squares.

In order to proceed in examining this problem, *norms* must be introduced in the domain $(c, \Delta c)$ and range spaces.

Since c is bound to remain quite smooth, as is required to justify the convolutional model, we confine c to the exponential of a space of piecewise polynomials. Specifically,

$$c(z) = c_0 \exp \int_0^z \ell$$

where ℓ is a spline on a bounded mesh of fixed nodes. Perturbations δc in c then have the form

$$\delta c(z) = c(z) \int_0^z \delta \ell$$

We norm the space of spline perturbations $\delta \ell$ with a weighted version of L^2 ; note that this space is finite-dimensional. To keep all of the objects of our discussion well-defined, it is also necessary to maintain the *a-priori* upper and lower bounds on $c(z)$ mentioned in the previous section. For the most part, we shall ignore this matter by presuming that the solution of any of the several optimization problems to follow lies in the *interior* of the set defined by these bounds. Ultimately, these problems should be reformulated to include the upper and lower velocity bounds as inequality constraints.

For the seismogram S , the choice of norms is implicit in the definition of the output least-squares problem: we choose the L^2 -norm on the data rectangle $[0, t_{\max}] \times [0, p_{\max}]$.

This choice essentially implies a choice of norms for Δc . In fact, since S is bilinear in f and Δc , properties of one of these two need to be fixed, to avoid ambiguities. We will assume that the operator

$$w(t) \longmapsto \int dt' f(t-t')w(t')$$

defines a bounded operator on L^2 . For instance $f \in L^1(\mathfrak{R})$, or, more generally, f is a finite measure. We explicitly want to include as the “ideal” case

$$f(t) = \delta(t)$$

i.e. the Dirac delta-function, as this choice yields a seismogram containing the most information about Δc , and any other choice, e.g. Figure 1, may be viewed as an approximation.

Thus we must demand that $\partial r / \partial t$ is square-integrable. It is clear from the definition of r that this will occur only if $\partial \Delta c / \partial z$ is square-integrable. The least-restrictive class with this property is the Sobolev space

$$H_{\text{loc}}^1(\mathfrak{R}) = \{\Delta c \in L_{\text{loc}}^2(\mathfrak{R}) : \partial \Delta c / \partial z \in L_{\text{loc}}^2(\mathfrak{R})\}$$

A convenient way of making this a normed space is to demand that $\Delta c \equiv 0$ for $z < 0, z > z_{\max}$. We can identify the resulting class with

$$H_0^1[0, z_{\max}] = \{\Delta c \in L^2[0, z_{\max}] : \\ \partial \Delta c / \partial z \in L^2[0, z_{\max}], \Delta c(0) = \Delta c(z_{\max}) = 0\}$$

The norm in $H_0^1[0, z_{\max}]$ is

$$\|\Delta c\|_1^2 = \int_0^{z_{\max}} dz [|c(z)|^2 + |\partial c / \partial z(z)|^2]$$

The conventional reference for information on the Sobolev spaces is Adams (1975).

Having defined norms, we can attempt to check the conditions which guarantee existence, uniqueness, and continuous dependence of the minimizer of J_{LS} on the data, *via* the implicit function theorem. As it happens, these are exactly the "standard conditions" needed to ensure local convergence of Newton's method.

Specifically, let $c_0, \Delta c_0, D_0$ be such that

$$J_{LS}[c_0, \Delta c_0; D_0] = 0$$

Suppose D is near D_0 . Then we would like to conclude that there exists a local minimizer $(c, \Delta c)$ of $J_{LS}[c, \Delta c; D]$ near $(c_0, \Delta c_0)$. The implicit function theorem will provide such a result, provided that J_{LS} is twice differentiable, with Lipschitz-continuous second derivative. This regularity condition alone is insufficient, of course, but we are already in trouble: J_{LS} is not even differentiable, in the generality given above. Indeed, suppose $f = \delta$, so that

$$S[c, \Delta c](t, p) = \frac{\partial r}{\partial t}(t, p) \\ = \left[\frac{\partial}{\partial z} \left(\frac{\Delta c(z)}{c(z)(1 - c^2(z)p^2)} \right) \right]_{\zeta(t, p)} \cdot \frac{\partial \zeta}{\partial t}(t, p)$$

Then formally a simple calculation using the chain rule gives

$$DS[c, \Delta c] \cdot [\delta c, 0] \\ + \frac{\partial^2}{\partial z^2} \left(\frac{\Delta c}{c(1 - c^2 p^2)} \right) \Big|_{z=\zeta[c]} \cdot D\zeta[c] \delta c \frac{\partial \zeta[c]}{\partial t} \\ + \frac{\partial}{\partial z} \left(\frac{\Delta c}{c(1 - c^2 p^2)} \right) \Big|_{z=\zeta[c]} \frac{\partial}{\partial t} D\zeta[c] \delta c$$

in obvious notation, taking into account the functional dependence of ζ on c .

This expression actually gives the correct (Gateaux) derivative of S in the direction $(\delta c, 0)$ when Δc has *two* square integrable derivatives — i.e. when it makes sense. Since DS involves the second derivative of Δc , however, it cannot be continuous in the sense of $H_0^1[0, z_{\max}]$. Thus there is no possibility of meeting the requirements of the implicit function theorem.

In reality, arbitrarily high frequencies are not measured: that is, reasonable f should be more regular than the delta function. A minimal degree of smoothness, met in all practical situations, is that $|\partial f/\partial t|$ be integrable. Then a simple integration-by-parts argument shows that S is differentiable. If $|\partial^2 f/\partial t^2|$ is integrable, then S is twice differentiable, as is J_{LS} , and the regularity requirements of the implicit function theorem are met.

Unfortunately, these assumptions (of regularity of f) create new difficulties. It is easily seen that, as soon as $|\partial f/\partial t|$ is integrable, the operator

$$\delta\Delta c \mapsto DS[c, \Delta c] \cdot [0, \delta\Delta c] = S[c, \delta\Delta c]$$

(note that S is *linear* in Δc) is *compact*, which means that the Hessian of J_{LS} cannot be positive-definite. This is the other major requirement of the implicit function theorem of course, and also the “second order sufficiency condition” for convergence of Newton’s method and relatives. This difficulty can be alleviated to some extent by *a priori* bounds, e.g. in the form of a penalty term, on higher derivatives of $\delta\Delta c$ (Tikhonov and Arsenin, 1977).

Even worse is in store, because of the qualities which must be assumed of Δc and f in order to model real-world seismograms. While $|\partial f/\partial t|$ is indeed integrable, for realistic choices of scales, frequencies, etc., it is *much* larger than $|f|$. Indeed, for the example given in Figure 1,

$$\begin{aligned} \int |f| &\cong 0.32 \\ \int |\partial f/\partial t| &\cong 44.4 \end{aligned}$$

(The integrals are approximated using the trapezoidal rule). Examining the expressions above for the partial derivatives of S with respect to c , respectively, Δc , we see that these involve convolution of $\partial f/\partial t$, respectively f , with expressions involving $\partial\Delta c/\partial z$ (in the first case, integration by parts is required to achieve this). Thus, if $\partial\Delta c/\partial z$ is “large”, then the partial derivative of S with respect to c is *much larger* than the partial derivative with respect to Δc . In fact, realistic Δc ’s are quite oscillatory nearly everywhere, so we can expect DS , hence $\text{Hess } J_{LS}$ to be *very ill-conditioned* at a near-consistent minimum. Also, S itself involves only convolution of f with an expression involving $\partial\Delta c/\partial z$, so the rate of change of S in “ c -directions” near the solution is very large compared to the size of S itself. Thus J cannot be convex over a large region: as c is changed, it must *saturate*. We conclude that the output-least-squares problem is most likely rather *non-convex*.

To illustrate that this behavior in fact occurs, we show a $(c, \Delta c)$ pair in Figure 3 derived (ultimately!) from a well-log data, supplied to us by Dr. Pierre Kolb of SNEA(P), whose group has used it extensively in their studies of the output least-squares method. Shown are the total velocity profile, the exponential integral of projection of its logarithmic derivative into a 10-dimensional space of cubic splines covering the interval of definition (i.e. the background velocity $c(z)$), and the difference (reflection series) $\Delta c(z)$. The spline space is illustrated in Figure 4. Figure 5 shows the convolutional model seismogram computed using the source wavelet of Figure 1, which we take to be D in the definition of J . Figure 6 shows a background velocity perturbation $\delta c(z)$. Figure 7 shows the function of a parameter h , $-1 \leq h \leq 1$, defined by

$$h \longmapsto J_{LS}[c + h\delta c, \Delta c, D]$$

Clearly for these choices J_{LS} is highly non-convex and very ill-conditioned near the global minimum, as predicted.

Evidence (although often not identified as such) for the disastrous effect of this structure on the convergence of quasi-Newton methods for J_{LS} may be found in Kolb *et al.* (1986), McAulay (1985, 1986), Hadjee and Collino (1988), Gauthier *et al.* (1986), Mora (1986, 1987). In fact, it seems quite clear that this structure presents a very substantial — perhaps insuperable — obstacle to the reliable solution of the output least-squares problem. We remark that Kolb and his collaborators have devised several clever devices which make the present problem more amenable to computation. The scope of these devices is far from clear, however: They are very closely tied to the “layered” hypothesis, i.e. parameters are functions of z alone, and substantial generalization seems unlikely. See Kolb *et al.* (1986) for an informative discussion.

Extensive analysis of the “real” problem (in which S is defined by solution of a boundary value problem, rather than by the convolution model), with many illustrations, may be found in Santosa and Symes (1986).

4 The Constrained Least-squares Problem

Note that the plane-wave reflectivity $r(t, p)$, and the velocity $c(z)$ contain together all information needed to construct the reflection series $\Delta c(z)$, through the relation

$$r(t, p) = \frac{\Delta c(z)}{c(z)(1 - c^2(z)p^2)} \Big|_{z=\zeta(t, p)} \quad (4.1)$$

Thus we could, in principle, view S as a function of c and r , rather than c and Δc : define

$$\tilde{S}[c, r](z, p) = \int dt' f(t - t') \frac{\partial r}{\partial t}(t', p)$$

Since \tilde{S} is linear in r and really doesn't involve c at all, it is (infinitely often) differentiable as soon as norms are selected which make it bounded. As per the discussion in the preceding section $\partial r / \partial t$ must be square-integrable, in p and t .

Since $\Delta c(z)$ is a function of one variable and $r(t, p)$ is a function of two, some constraint on $r(t, p)$ is required if it is to arise from a $\Delta c(z)$. Clearly

$$\frac{\partial}{\partial p} ((1 - c^2(z)p^2)r(\tau(z, p), p)) = 0 \quad (4.2)$$

An equivalent requirement, after some algebra (Symes 1988a, Section 2) is

$$\begin{aligned} C[c, r] &:= \frac{\partial r}{\partial p}(t, p) - p \left(\int_0^{\zeta(t, p)} dz v(z, p) \right) \frac{\partial r}{\partial t}(t, p) \\ &\quad - 2pv^2(\zeta(t, p), p)r(t, p) \\ &= 0 \end{aligned} \quad (4.3)$$

(recall that

$$v(z) = \frac{c(z)}{\sqrt{1 - c^2(z)p^2}}$$

is the vertical plane-wave velocity.)

For lack of more inspired nomenclature, we have decided to call reflectivities $r(t, p)$ satisfying (4.2) or (4.3) *coherent* (since the various "traces" $t \mapsto r(t, p)$ are related); the functional C the *incoherence*, and the condition (4.3) the *coherency condition*.

It is easy to show, conversely, that if

$$\begin{aligned} &r \in H_b^1([0, t_{\max}] \times [0, p_{\max}]) \\ &= \{r \in L^2([0, t_{\max}] \times [0, p_{\max}]) : \\ &\quad \frac{\partial r}{\partial t}, \frac{\partial r}{\partial p} \in L^2([0, t_{\max}] \times [0, p_{\max}]) \\ &\quad \text{and } r(0, p) = r(t_{\max}, p) = 0, \quad 0 \leq p \leq p_{\max}\} \end{aligned}$$

satisfies (4.3), then r is related to a suitable $\Delta c \in H_0^1([0, z_{\max}])$ for suitable z_{\max} by means of (4.1). Moreover, \tilde{S} is obviously bounded, hence C^∞ , on this space.

Thus a constrained least-squares problem equivalent (in the sense of having related minima) to the output least-squares problem of the last section, is

$$\begin{aligned} & \text{minimize}_{c,r} \quad \tilde{J}_{LS}[c, r; D] \\ & \text{subject to} \quad C[c, r] = 0 \end{aligned}$$

where

$$\tilde{J}[c, r, D] = \|S[c, r] - D\|^2$$

where the minimization is over c in the finite-dimensional space of exponential integrated splines explained in the last section, and r in the space $H_b^1([0, t_{\max}] \times [0, p_{\max}])$ defined above.

At first glance this problem looks like a great improvement. As noted above, \tilde{J}_{LS} is infinitely differentiable, being quadratic in r (and independent of c !). Inspection of (4.3) shows that perturbation with respect to c involves only differentiation of v , i.e. c , which is harmless — c is confined to a finite-dimensional space, on which differentiation is a bounded operator. Thus C is also C^∞ as a function of c and r with values in $L^2([0, t_{\max}] \times [0, p_{\max}])$, and we conclude that:

The constrained least-squares problem involves smooth objective function and constraint, for the least-restrictive choices of function classes.

The Lagrangian theory of smooth constrained problems gives sufficient conditions for existence, uniqueness, and stability of constrained minima, as well as convergence of Newton's method and relatives such as so-called sequential quadratic programming (Tapia, 1978). We recall these conditions in a general context (see e.g. Luenberger, 1973). Suppose that X, Y are Hilbert spaces, $U \subset X$ open,

$$\left. \begin{aligned} F : U &\longrightarrow \mathfrak{R} \\ G : U &\longrightarrow Y \end{aligned} \right\} C^2 \text{ functions}$$

Consider the constrained optimization problem:

$$\begin{aligned} & \text{minimize}_{u \in U} \quad F(u) \\ & \text{subject to} \quad G(u) = y_0 = G(u_0) \end{aligned}$$

Suppose that u_0 is a local constrained optimum, λ_0 the corresponding Lagrange multiplier. Then the algorithms mentioned above are superlinearly convergent to u_0 for initial estimates close enough to u_0 , and C^2 -small perturbations of F, G give problems with local optima which are small perturbations of u_0 , provided that the following conditions are satisfied:

(i) (“*Second-order sufficiency*”): for some $K > 0$

$$\begin{aligned} \langle \delta u, [\nabla \nabla F(u_0) + \langle \lambda_0, \nabla \nabla G(u_0) \rangle_Y] \delta u \rangle_X \\ \geq K \|\delta u\|_X^2 \end{aligned}$$

for all $\delta u \in X$

(ii) (“*Constraint Qualification*”): for some $K^* > 0$

$$\|DG(u_0)^* \lambda\|_X \geq K^* \|\lambda\|_Y$$

for all $\lambda \in Y$

Surprisingly, it is sometimes possible to prove that (i) holds for the constrained least-squares problem described above. The constraint qualification (ii) fails, however, in an interesting way. The demonstration of this failure for the coherency constraint C is a bit messy, so we show the same instead for a simpler, but closely related, model problem.

Define (see above)

$$\begin{aligned} U &= X = H_b^1([0, t_{\max}] \times [0, p_{\max}]) \\ Y &= L^2([0, t_{\max}] \times [0, p_{\max}]) \end{aligned}$$

Choose any function $g \in Y$, and define

$$F : U \longrightarrow \mathfrak{R}$$

by

$$F(u) = \left\| \frac{\partial u}{\partial t} - g \right\|_{L^2([0, t_{\max}] \times [0, p_{\max}])}^2$$

and

$$G : U \longrightarrow Y$$

by

$$G(u) = \frac{\partial u}{\partial p}$$

Then the adjoint operator G^* is computed as follows: for smooth $\lambda \in Y$,

$$G^* \lambda = w$$

where w solves the self-adjoint boundary value problem

$$w - \frac{\partial^2 w}{\partial t^2} - \frac{\partial^2 w}{\partial p^2} = -\frac{\partial \lambda}{\partial p}$$

$$w(0, p) = w(t_{\max}, p) = 0, \quad 0 \leq p \leq p_{\max}$$

$$\frac{\partial w}{\partial p}(t, 0) = \frac{\partial w}{\partial p}(t, p_{\max}) = 0, \quad 0 \leq t \leq t_{\max}$$

The eigenfunctions of this problem are

$$w_{n,m}(t,p) = \sin \frac{n\pi t}{t_{\max}} \cos \frac{m\pi p}{p_{\max}}$$

whence we compute that, for the choice

$$\lambda(t, p) = \sin \frac{n\pi t}{t_{\max}} \sin \frac{m\pi p}{p_{\max}}$$

$$G^* \lambda = \left(1 + \left(\frac{n\pi t}{t_{\max}} \right)^2 + \left(\frac{m\pi}{p_{\max}} \right)^2 \right)^{-1} \frac{m\pi}{p_{\max}} w_{n,m}$$

so

$$\|G^* \lambda\|_X \leq \text{const.} \left(\frac{m}{\sqrt{1 + n^2 + m^2}} \right)$$

whereas

$$\|\lambda\|_Y \geq \text{const.} > 0$$

independently of n, m . Thus no estimate of the form (ii) can possibly hold. Of course, the meaning of this failure is that we are not guaranteed that the constraint set is a *sub-manifold* of U , in general. In the present case it is, as G is linear. However, the presence of any uniform nonlinearity will produce “cusp” singularities, i.e. a sequence of independent directions in which the radii of curvature of the constraint set decreases without bound. Such a situation can be produced by an arbitrarily C^2 -small perturbation of G , and we believe that it is the case for the coherence constraint.

The resemblance between this example and the constrained least-squares problem is clear, in view of (4.2), and it is not too hard to reach a similar conclusion. The constraint qualification (ii) fails, and we do not expect to achieve reasonable computational efficiency, or stability of least-squares estimates.

5 The Coherency Optimization Problem

To begin with, we pursue the simple example, described at the end of the last section, a little further. Suppose that we weaken the constraint $G = 0$, by incorporating it as a so called L^2 -penalty term in an unconstrained minimization problem

$$\min_{u \in U} [\|F(u) - g\|_Y^2 + \sigma^2 \|G(u)\|_Y^2]$$

The Hessian quadratic form of this (quadratic) cost functional is

$$H[u] = 2 \left(\left\| \frac{\partial u}{\partial t} \right\|_Y^2 + \left\| \frac{\partial u}{\partial p} \right\|_Y^2 \right)$$

In view of the definition of X ,

$$H[u] \geq K \|u\|_X^2$$

with $K > 0$: this is a version of *Poincaré's inequality* (Nečas, 1967).

The constant K depends on σ , but is positive for any positive choice of σ : in fact $K = 0(\sigma^2)$. Thus the penalized function is convex, and this convexity is robustly dependent on the penalty parameter σ .

This reasoning leads to the third version of the velocity inversion problem, which we call the *coherency optimization problem* (for lack of a more inspired name):

$$\min_{c, r} J_{CM}[c, r, D]$$

where

$$J_{CM}[c, r, D] := \tilde{J}_{LS}[c, r, D] + \sigma^2 \|C[c, r]\|_{L^2([0, t_{\max}] \times [0, p_{\max}])}^2$$

The minimization is to be carried out over the set of c, r described in previous sections, on which J_{CM} is a smooth function.

To illustrate this smoothness, and hint at the convexity property to be discussed below, we scan J_{CM} over the same segment of models used to generate Figure 7 (end of Section 2). That is, in Figure 8 we plot

$$h \longmapsto J_{CM}[c + h\delta c, r, D]$$

where $c, \delta c, D$ are as indicated in Section 2, and r is derived from Δc given in Figure 3. This plot suggests that J_{CM} is both smoother and much more convex than J_{LS} , while having the same minimum.

Since this problem is unconstrained (except for the velocity bounds mentioned in Section 2, which we are assuming inactive at the solution), the key theoretical point is the positive-definiteness of the Hessian. The analysis of the

Hessian is carried out in the companion paper Symes (1988a) (see also Symes 1988b, c), Santosa and Symes (1986), Appendix E), and is somewhat involved. Here we give only an imprecise statement of the main conclusion:

Suppose that

$$D = S[c^*, \Delta c^*] + E$$

where c is sufficiently smooth and Δc is sufficiently rough, and E is a "noise" term, sufficiently small relative to the properties of c and Δc . Then the functional

$$J_{CM}[c, r, D]$$

is convex in a neighborhood of c^, r^* (where r^* is computed from $c^*, \Delta c^*$ as at the beginning of Section 2) and has a unique minimizer in this neighborhood whose distance to (c^*, r^*) is $O(\|E\|)$.*

The subtlety here is the notion of "sufficiently rough." Reflection seismograms contain information only because of the *echoes* which are caused by rapid variation in rock mechanical properties. The above-mentioned papers quantify the degree of parameter "roughness" necessary for stability in the inverse problem.

To see how this theorem would work itself out in practice, we implemented a quasi-Newton algorithm to solve the coherency optimization problem. A lengthy description of this code is given in the technical report (Symes, 1988b). Here we give a brief overview of the code, and a description of some of the early numerical results. These reveal several aspects of the coherency optimization problem at a level of detail beyond that achieved by the theory.

The tests reported below were made using a so-called truncated Newton code. This code is based on the model trust region principle (Dennis and Schnabel (1983), section 6.4) and on the extensions to it introduced by Steihaug in his Yale thesis (Steihaug (1981); see also Toint (1981)). Essentially, the Gauss-Newton linear step

$$\text{Hess } J_{CM} \cdot [\delta x, \delta r] = -\text{grad } J_{CM}$$

is solved by a conjugate residual iteration (Golub and Van Loan (1983), Ch. 10), which is terminated when the step estimate exits a ball about the current solution estimate, the radius of which is determined by a simple and robust updating strategy. This expedient avoids expensive conjugate residual steps taken outside the region in which the linearized model can be "trusted", hence the name. A more lengthy description of the code can be found in Santosa and Symes (1986), Chapter 9, where the same codes were used in solving the nonlinear output least-squares problem, using finite difference synthetic seismograms instead of the convolutional model.

An important amendment of the trust region idea is natural in this problem. All models in the iteration are supposed to obey the bound $c(z)p_{\max} < 1$ for all depths z of interest. This may cease to be true during an update step, if the velocity is increased by too much at some depth. The computation of the gradient simply flags this occurrence, and the algorithm attempts a smaller step in the same direction. Thus the trust region, for problems like the present one, may involve in a natural way constraints on the validity of the model itself.

In fact, the functional J_{CM} is sufficiently quadratic that the Gauss-Newton steps very seldom leave the ball defined by the trust-region update strategy. The intrinsic, model-validity backtrack strategy was almost always activated at some point in the iteration, however (though not near the end).

We performed a number of numerical experiments using the velocity profile c (upper curve) and perturbation Δc (lower curve) exhibited in Figure 3 and the source wavelet of Figure 1 to generate the convolutional model data of Figure 5. A target background velocity with a lower velocity zone was chosen because the structure of such a zone is impossible to determine from refraction arrival times (Aki and Richards (1980), Ch. 12), and intrinsically more difficult for output least-squares — see Santosa and Symes (1986).

The algorithm explained in this section was used to extract estimates of c and r from the data of Figure 5. In all cases, we observed the same pattern. We began with the simple estimate $c_{initial} = \text{const.} (= 1)$, $r_{initial} \equiv 0$. The first Newton step did not change the estimate of c , since the incoherence of $r \equiv 0$ vanishes for any velocity model. Otherwise put, since there are initially no reflectors, there is initially no travel-time information in the reflectivity with which to update the velocity model. The first iteration is thus devoted entirely to minimizing

$$\tilde{J}_{LS}[c, r; D]$$

to find an initial (nontrivial) estimate for r . Unless otherwise noted, each Gauss-Newton step (including the first) is approximated by five conjugate-residual iterations.

In the second and subsequent iterations, the velocity model $c(z)$ is improved, and data-noise-generated incoherency in the reflectivity is reduced. Figure 9 shows the velocity-estimate results of five and ten Gauss-Newton steps from both the constant velocity model (curve 0) and five Gauss-Newton steps from the velocity model identified as the curve 1. The results are labeled as curves 2, 3 and 4 respectively. In all cases the logarithmic derivative of the velocity was chosen from a five-dimensional space of cubic splines. This result is evidence for the independence of initial estimate of the final estimates. In particular, the results of ten iterations from a constant initial model, and five iterations from the “incorrect trend” initial model (curve 1) are virtually identical. The error is quite stable and is due to the fact that the target is not a member of the space of velocities defined by the five-node spline basis — i.e. we

get the “closest” five-node velocity estimate.

These estimates are reasonably accurate, especially considering the computational work required. Apparently the incoherence resulting from the erroneous velocity at the bottom of the model was insufficient to cause further corrections, or possibly the calculation of the incoherence is substantially inaccurate there — see comments below. Most of the reduction in J_{CM} (about an order of magnitude) came in the first iteration, in which \bar{J}_{LS} is reduced. Yet another order of magnitude is gained in the remaining iterations, in which the incoherence is reduced.

Two major points have emerged from the experimental work conducted thus far. The first concerns the number of spline nodes: the outcome is in some ways quite sensitive to this number. For example, the experiment of Figure 9 was repeated with a 13-dimensional spline space of velocity log-derivatives instead. Five, ten, and thirty iterations of the Gauss-Newton process produced the curves labeled 1, 2, and 3 in Figure 10. These look quite “wild,” and certainly the mean-square error is much greater than is the case with those in Figure 9. Recall however that the principal role of the background model is to supply travel-times, i.e. to determine $\tau(z, p)$ — see Sections 2, 4. A quite different picture emerges when the travel-times are plotted against the “true” travel time curve (from the velocity profile of Figure 6). In Figure 11 are displayed (normal-incidence, i.e. $p = 0$) travel-time curves from the “exact” velocity, curve 1 from Figure 9, and curves 1 and 2 from Figure 10. In fact, the latter two curves are closer to the “true” travel-time than is the former, despite their correspondence with velocity estimates having larger L^2 -error. This relation emerges more clearly when the (normal incidence) travel-time *errors* are plotted: see Figure 12.

Apparently, the result of increasing the number of degrees of freedom in the model is to allow a closer fit to the travel-time, at least at points corresponding to major reflectors, but at the cost of an oscillatory error which may be large in the mean-square sense. This is easy to understand: the errors oscillate on a length scale too short to affect the travel-times between major reflectors, hence correspond to *small eigenvalues* of the incoherence Hessian. While the effect on travel-times of this sort of error is *a priori* small, it does produce irritating ambiguities in the velocity estimate, and (more important) has a negative impact on the convergence of the iterative scheme. Several approaches to the removal of this ambiguity suggest themselves. Trial-and-error determination of the optimal spacing for spline nodes, as has been done here, is clearly unsatisfactory. Systematically increasing the number of nodes until a good fit is obtained requires some notion of an “acceptable” level of fit, and such information may itself only be obtainable by trial-and-error. A more satisfactory approach might be adaptive estimation of small Hessian eigenvalues, through the close relation of conjugate-residual iteration with the Lanczos algorithm

(Golub and Van Loan (1983), Ch. 10), and penalization of the corresponding eigenvector components. For the application of such “iterative deflation” to linear systems see Chan (1986) (also Meza and Symes (1987)). Since the number of small eigenvalues associated with velocity perturbations is small, and since their characterization is somewhat independent of the current velocity estimate, this deflation strategy should work rather well in combination with Gauss-Newton iteration. Finally, since the culprit is oscillatory error, penalization of a velocity derivative may regularize this problem satisfactorily. Computational trials are in progress; results will be reported elsewhere.

A second point concerned the density of p -samples. Inspections of Figure 5 clearly show that for large p and t , the “moveout”, i.e. travel-time or phase difference, in neighboring traces (constant- p sections) may be a substantial fraction of a wavelength. As we have based our difference approximations to the incoherency on centered difference approximations to the coordinate derivatives $\frac{\partial}{\partial t}$, $\frac{\partial}{\partial p}$, the possibility exists of severe undersampling in p . In fact when we increased Δp to .02 (from .01 as in Figure 8), the computation analogous to that for Figure 9 gave completely erroneous results for the deeper part of the velocity profile, apparently because the part of the incoherency due to deeper reflectors was grossly underestimated. We suspect that residual inaccuracy in the deeper parts of the curves in Figure 9 is due to a milder version of the same effect.

Besides finer sampling, methods to overcome errors in incoherency due to undersampling include higher-order difference formulas and difference formulas better adapted to the reflection phases. Indeed, low-order differences along even a crude approximation to the “moveout” curves ($t = \tau(z, p)$, z constant) should produce better results at coarser sampling than the coordinate derivatives used in our present code. These ideas are also under investigation.

It might be objected that, while the output includes an estimate of the travel-time reflectivity section $r(t, p)$, no estimate of the corresponding velocity perturbation $\Delta c(z)$ is provided. After all, the original model is $c + \Delta c$, and r has no direct physical significance. The final reflectivity is not necessarily entirely coherent, and so does not correspond to any $\Delta c(z)$, strictly speaking. Nonetheless, an estimate may be produced by averaging $r(t, p)$ on the basis of equation (5), i.e.

$$\Delta c(z) \sim \frac{c^3(z)}{p_{\max}} \int_0^{p_{\max}} dp v^{-2}(z, p) r(\tau(z, p), p).$$

The output of this averaging formula is known in geophysical parlance as a *stack*. Stacks are widely used in seismic data-processing as noise-reduction and data compression devices — both functions are evident here.

We have implemented this post-inversion stack of the final reflectivity estimate, and display the results for the reflectivity corresponding to curve 3 in

Figure 9. When stacked with constant velocity $c \equiv 1$, the estimate of $\Delta c(z)$ is completely erroneous in phase and wrong by a factor of perhaps 3 in amplitude (Figure 13). When stacked with the final velocity estimate from Figure 9 (curve 3), the estimated $\Delta c(z)$ has essentially correct phases and amplitudes for the major excursions (Figure 14), as compared to the $\Delta c(z)$ actually used to generate the data. Of course, none of the subwavelength-scale variation in the true $\Delta c(z)$ could appear in the reflectivity $r(t, p)$ or in the stacked estimate, so a more interesting comparison is the convolutional-model seismogram generated using the velocity from Figure 9 and the stacked $\Delta c(z)$ from Figure 14. This is displayed in Figure 15, and the difference of Figure 15 and Figure 8, plotted on the same scale, in Figure 16.

In summary, we have described a preliminary implementation of the coherency optimization method, exhibited the results achievable with this rather crude code, and suggested some issues worthy of further examination. Given the rather large discretization errors in our present implementation, which play the role (at least) of data noise, the stability of the final estimates and the rate of convergence both appear quite satisfactory: to a limited extent, coherency optimization appears to produce reliable solutions of the velocity inversion problem.

Extensive noise studies, tuning, modifications along the lines suggested above, and direct comparison with output least-squares optimization will be reported in a future publication.

REFERENCES

- ADAMS, R. [1975]. *Sobolev Spaces*, Academic Press, New York.
- AKI, K. and RICHARDS, P. [1980]. *Quantitative Seismology: Theory and Methods*, Freeman, San Francisco.
- CHAN, T. [1985]. Deflated Lanczos procedures for solving nearly singular linear systems, research report YALEU/DCS/RR-403, Department of Computer Science, Yale University.
- CHAPMAN, C. [1978]. A new method for computing seismograms, *Geophys. J. Roy. Astr. Soc.* 54, 481-518.
- DENNIS, J. E. JR. and R. B. SCHNABEL [1983]. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs.
- GAUTHIER, O. , TARANTOLA, A. and VIRIEUX, J. [1986]. Two-dimensional nonlinear inversion of seismic waveforms, *Geophysics* 51, pp. 1387-1403.
- HADJEE, Y. and F. COLLINO [1988]. A geometrical approach to the a-priori study of the 1-d inverse problem, Institut Francais du Petrole preprint.
- LUENBERGER, D. G. [1973]. *Introduction to Linear and Nonlinear Programming*, Addison-Wesley.
- MCAULAY, A. [1985]. Prestack inversion with plane-layer point source modeling, *Geophysics* 50, pp. 77-89.
- MORA, P. [1986]. Nonlinear 2-d elastic inversion of multi-offset seismic data, Expanded abstract, 56th Annual International Meeting, Society of Exploration Geophysicists, Houston, pp. 533-537; also *Geophysics* 52.
- MORA, P. [1987]. Nonlinear 2-d elastic inversion of real data, Expanded abstract, 57th Annual International Meeting, Society of Exploration Geophysicists, New Orleans, pp. 430-432.
- ROBINSON, E. A. and S. TREITEL [1980]. *Geophysical Signal Analysis*, Prentice-Hall.
- SANTOSA, F. and W. SYMES [1988]. High frequency perturbational analysis for the point-source response of a layered acoustic medium, *J. Comp. Phys.*,

74, pp. 318-381.

- SANTOSA, F. and W. SYMES [1986]. An Analysis of Least-Squares Velocity Inversion, to appear in *Society of Exploration Geophysicists Monograph Series* (1988-89).
- STEIHAUG, T. [1981]. Quasi-Newton methods for large-scale nonlinear problems, Ph.D. Thesis, Yale University.
- SYMES, W. [1988a]. Velocity Inversion by Coherency Optimization: Analysis, preprint.
- SYMES, W. [1988b]. Velocity Inversion by Coherency Optimization, Technical Report 88-4, Department of Mathematical Sciences, Rice University.
- SYMES, W. [1988c]. Bandlimited velocity inversion: a model problem in reflection seismology, preprint.
- TAPIA, R. A. [1978]. Quasi-Newton methods for equality constrained optimization: equivalence of existing methods and a new implementation, *Non-linear Programming 3*, Academic Press, New York.
- TARANTOLA, A. [1987]. *Inverse Problem Theory*, Elsevier.
- TIKHONOV, A. N. and V. Y. ARSEININ [1977]. *Solution of ill-posed problems*, Winston, New York.
- TOINT, PH. L. [1981]. Towards an efficient sparsity exploiting Newton method for minimization, in *Sparse Matrices and their Uses*, I. S. Duff, ed., Academic Press, London, pp. 57-88.
- TREITEL, S., P. R. GUTOWSKI, and D. E. WAGNER [1982]. Plane-wave decomposition of seismograms, *Geophysics* 47, pp. 1375-1401.
- WATERS, K. H. [1981]. *Reflection Seismology: a Tool for Resource Exploration*, John Wiley and Sons, New York.

FIGURE CAPTIONS

1. Ricker Wavelet (second derivative of scaled Gaussian), peaked at 20 Hz.
2. Power spectrum (square of absolute value of Fourier Transform) for the wavelet of Figure 1.
3. Velocity model used to generate data for numerical experiments:
 - (1) original model (solid line);
 - (2) smoothed by projection into spline space (smooth background model - dotted line);
 - (3) difference between 1. and 2 (reflection series - dot/dashed line).

Note that the velocities are normalized against their surface ($z = 0.0$) values. Also, depth in this and following plots is given in *one-way (travel) time* at the surface velocity.
4. (a) cubic b-splines: 5th (solid) and 7th (dashed) of 12 nodes in interval; (b) indefinite integrals of splines in (a).
5. Convolutional model seismogram generated from model displayed in Figure 3, using the Ricker wavelet displayed in Figure 1. This type of display is quite common in seismology: the t-axis is vertical, and each curve is the graph of $S(t,p)$ for fixed p . The coherent reflection phases are quite evident in this type of plot.
6. Perturbation in the background velocity, used to generate Figures 7 and 8.
7. Scan of Mean-Square Error function along the line segment in model space defined by the reference (background) velocity from Figure 3 and the perturbation from Figure 6.
8. Scan of Mean-Square Incoherency, same segment as in Figure 7.
9. Results of Coherency Optimization for background velocity:
 - (0) target model, from Figure 3 (solid curve);
 - (1) a non-constant initial model (short dashes);
 - (2) result of 5 Gauss-Newton steps starting at constant model (long dashes);
 - (3) result of 10 Gauss-Newton steps starting at constant model (dash/cross);

(4) result of 5 Gauss-Newton steps starting at model (1) (short/long dashes).

10. Results using too many degrees of freedom in model (16 nodes):

- (0) target model (solid curve);
- (1) 5 Gauss-Newton steps (short dashes);
- (2) 10 Gauss-Newton steps (long dashes);
- (3) 30 Gauss-Newton steps (short/long dashes)

11. Two-way travel time curves at normal incidence ($p = 0.0$):

- (0) target model (solid curve);
- (1) curve 2, Figure 9 (short dashes);
- (2) curve 1, Figure 10 (long dashes);
- (3) curve 2, Figure 10 (short/long dashes).

12. Two-way travel-time errors:

- (0) between curves (0) and (1), Figure 11 (solid line);
- (1) between curves (0) and (2), Figure 11 (short dashes);
- (2) between curves (0) and (3), Figure 11 (long dashes).

13. Stack (solid line) of reflection series estimate corresponding to Figure 9, curve 2, at constant velocity, compared with target velocity perturbation (dashed line).

14. Stack (solid line) of reflection series estimate corresponding to Figure 9, curve 2, at velocity given by Figure 9, curve 2, compared with target velocity perturbation (dashed line).

15. Convolutional model seismogram produced using background velocity from Figure 9, curve 2, with stacked reflection series estimate from Figure 14.

16. Difference of Figures 15 and 5.

FIGURE 1

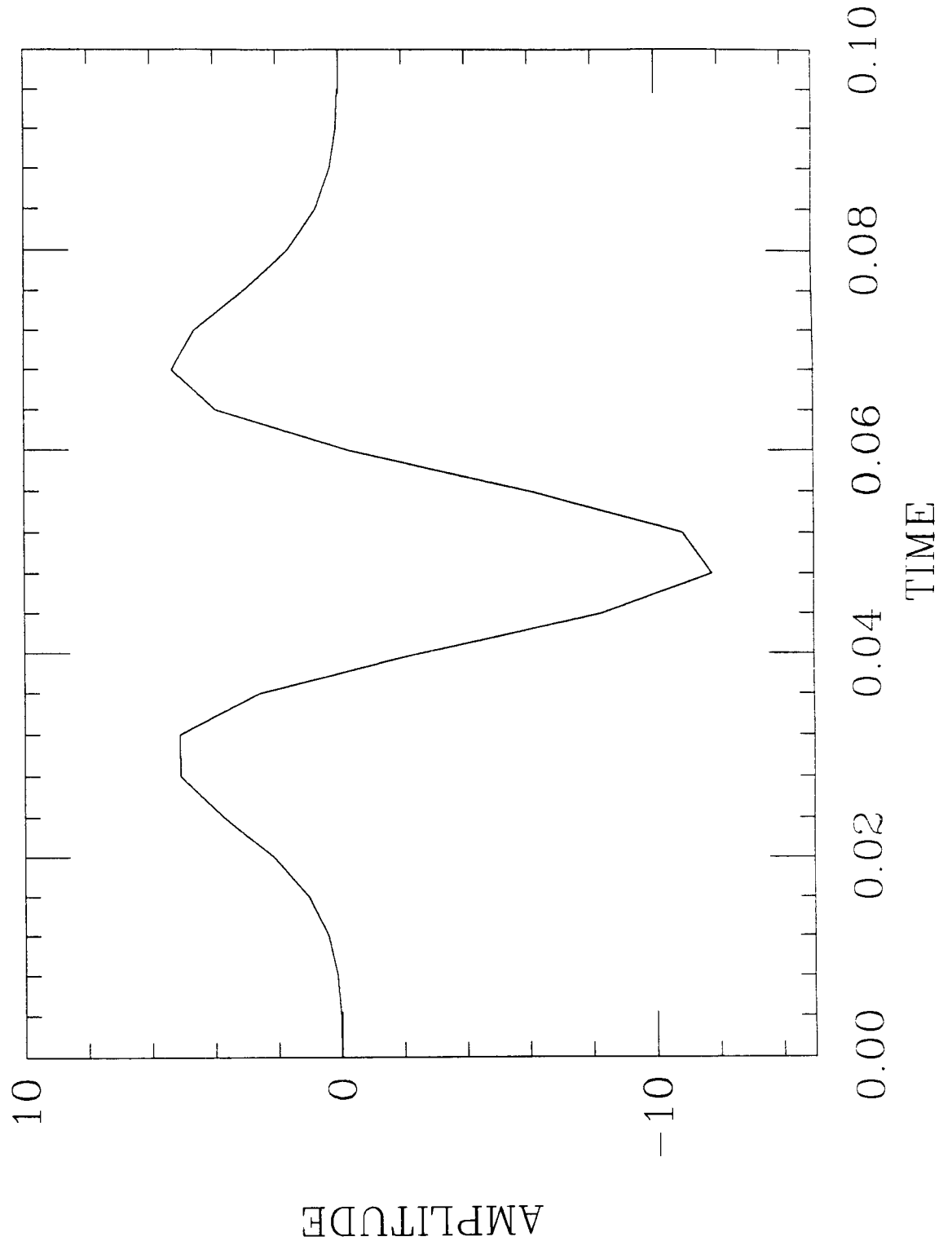


FIGURE 2

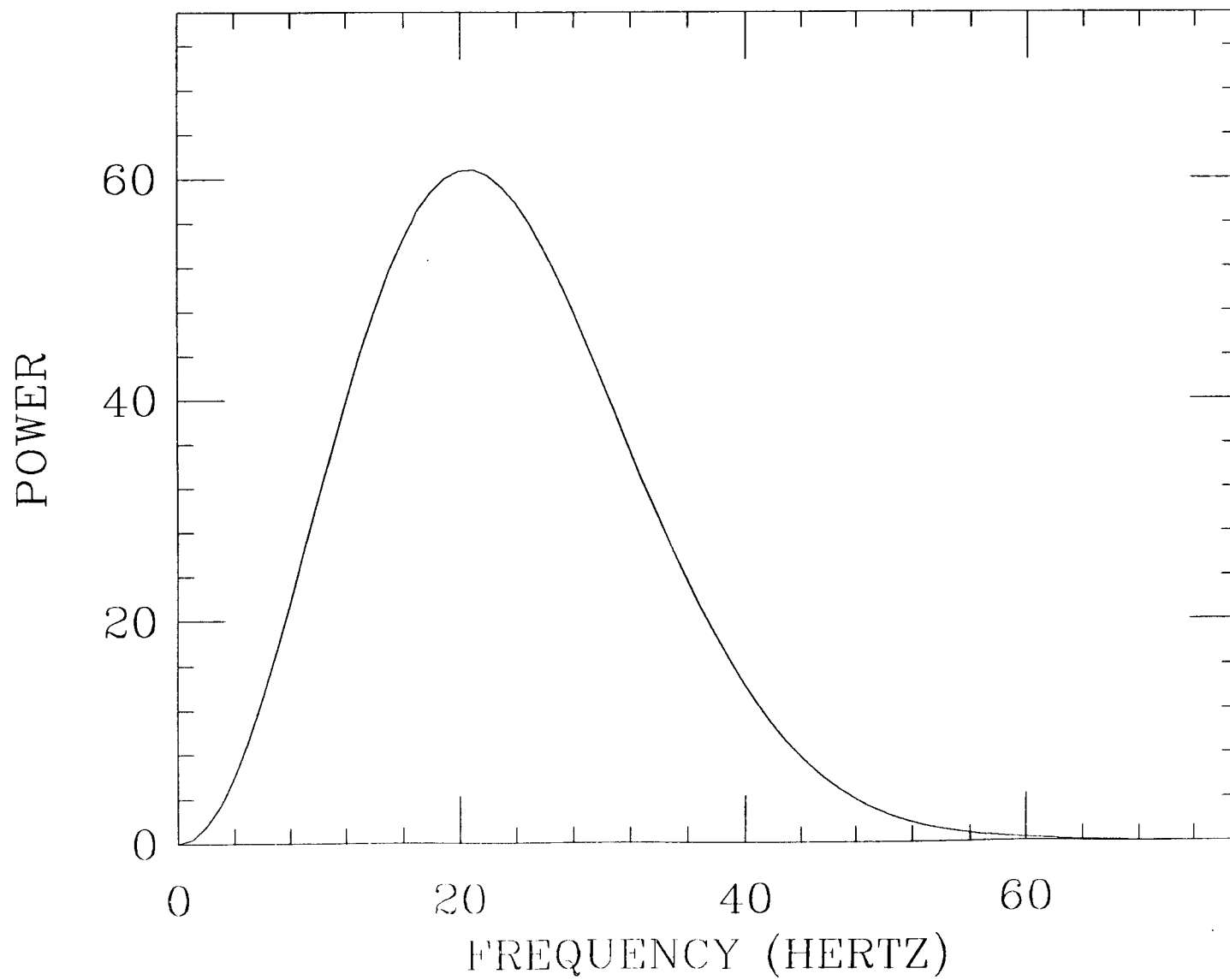


FIGURE 3

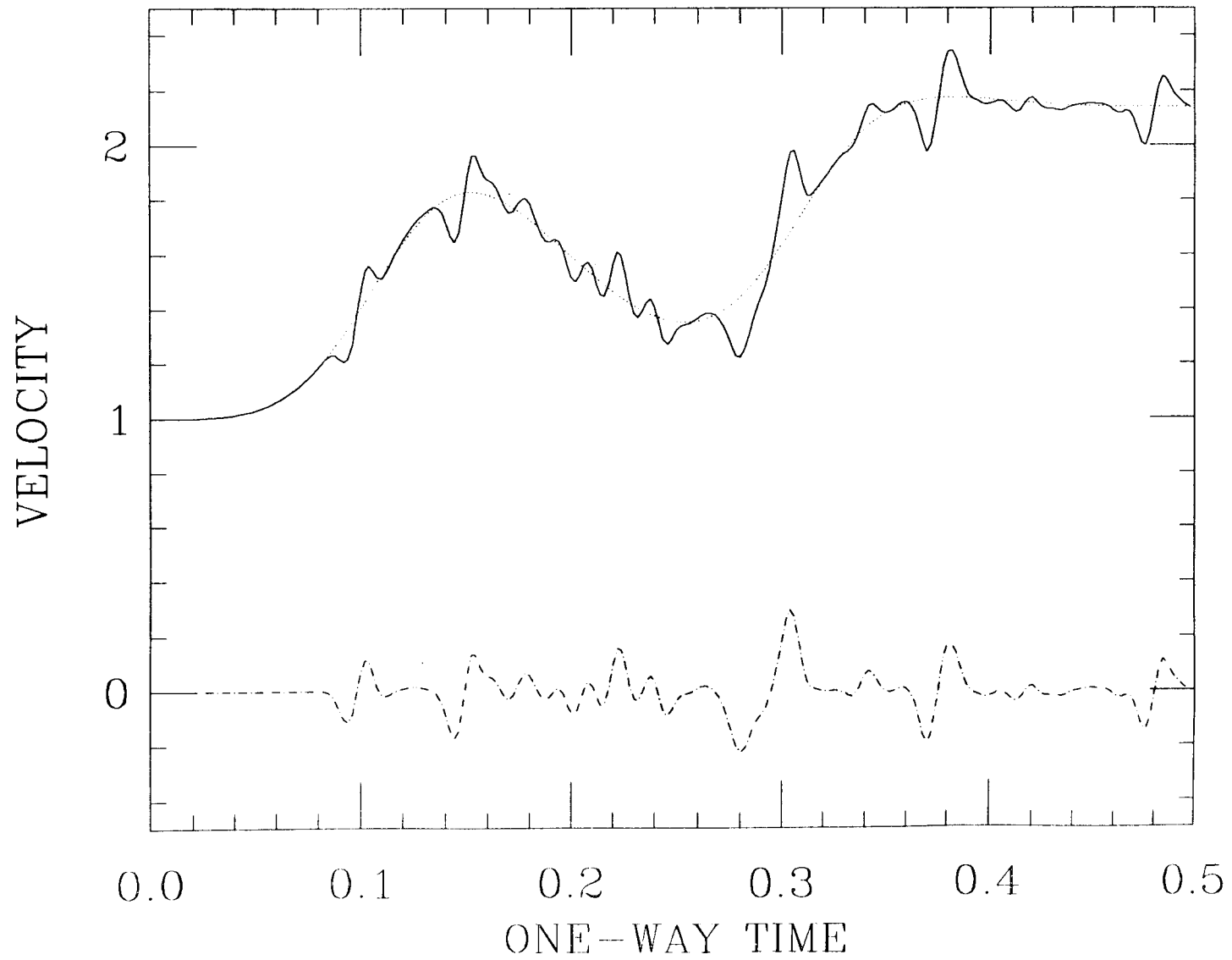


FIGURE 4 A

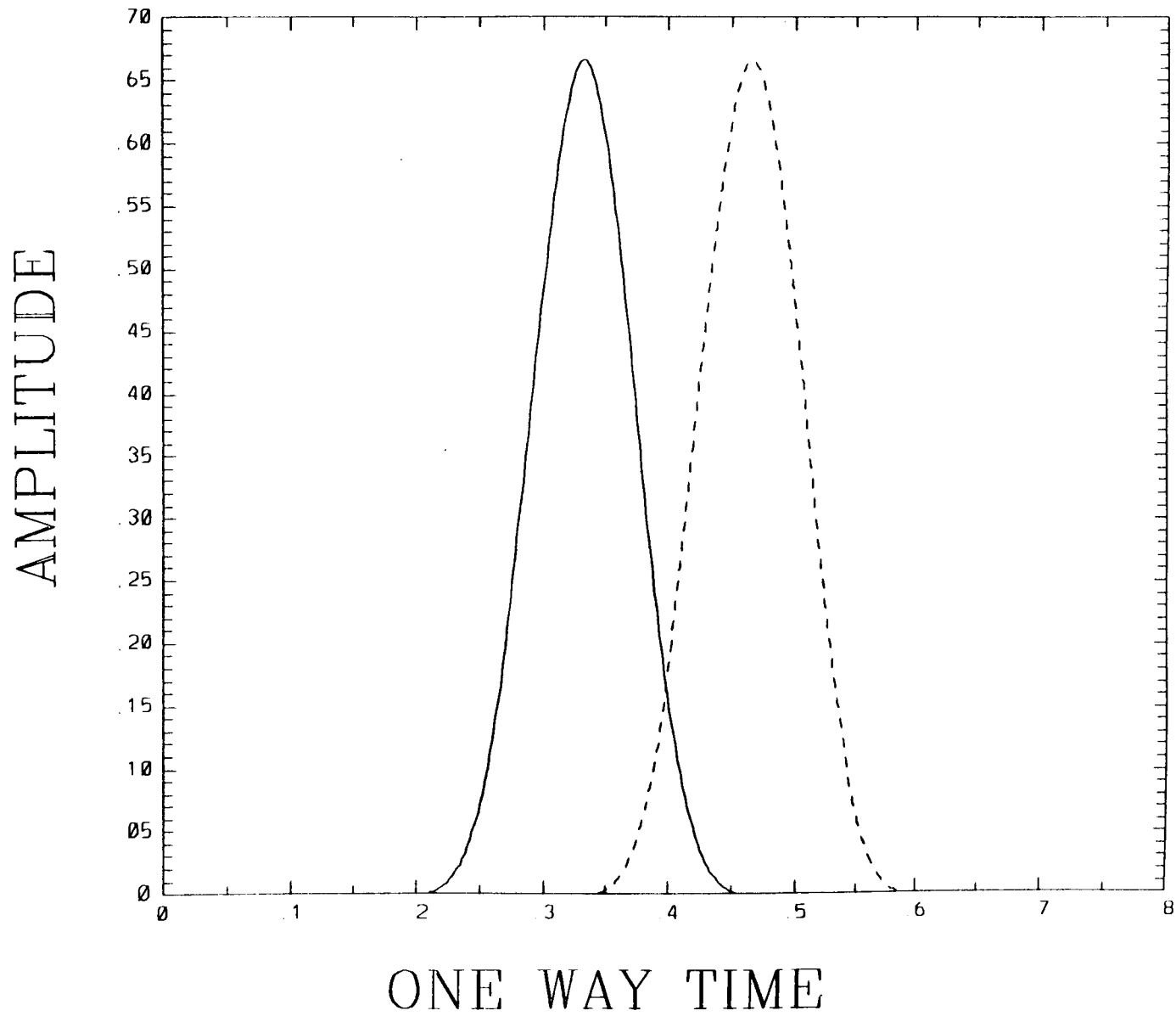


FIGURE 4 B

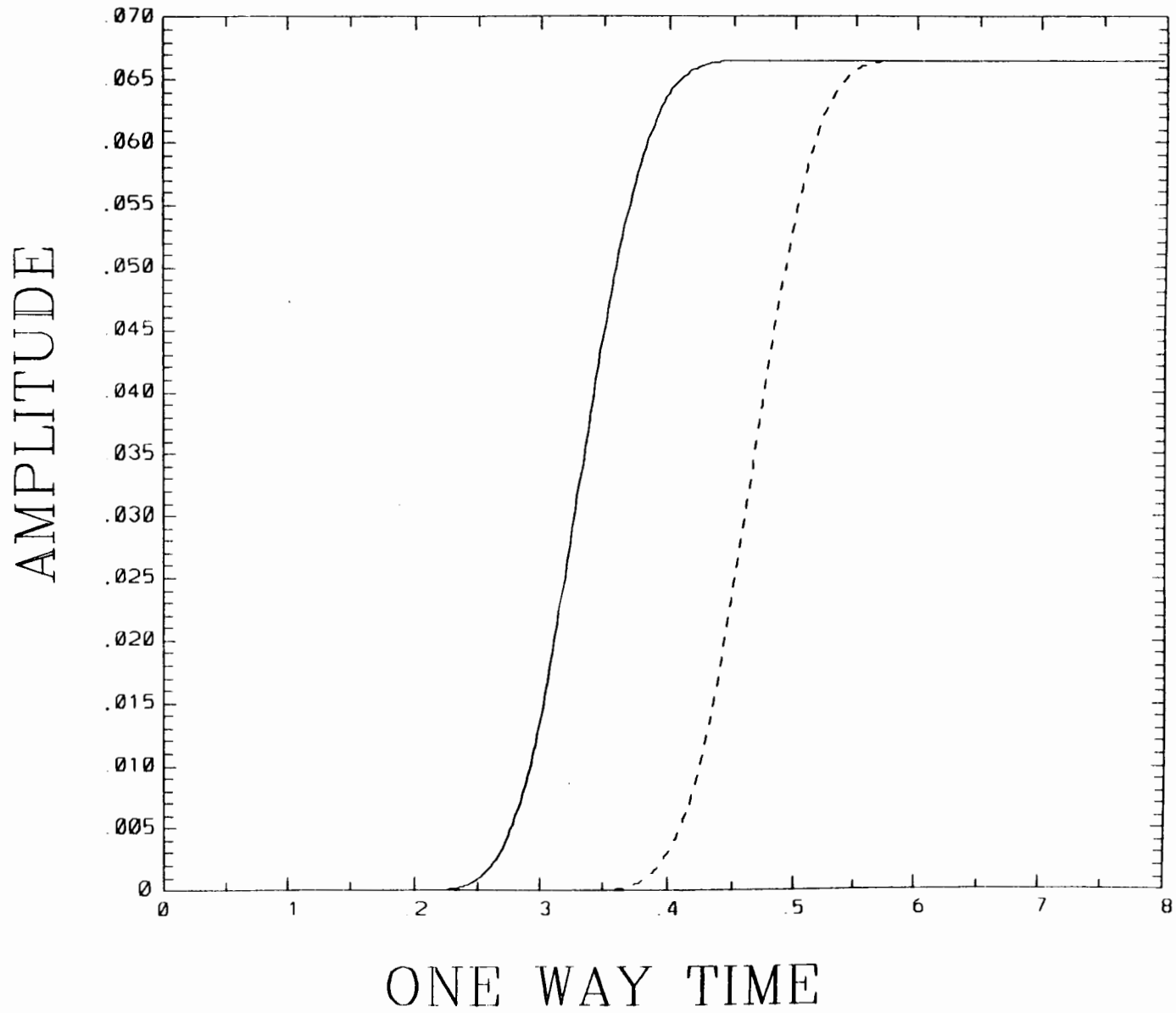


FIGURE 5

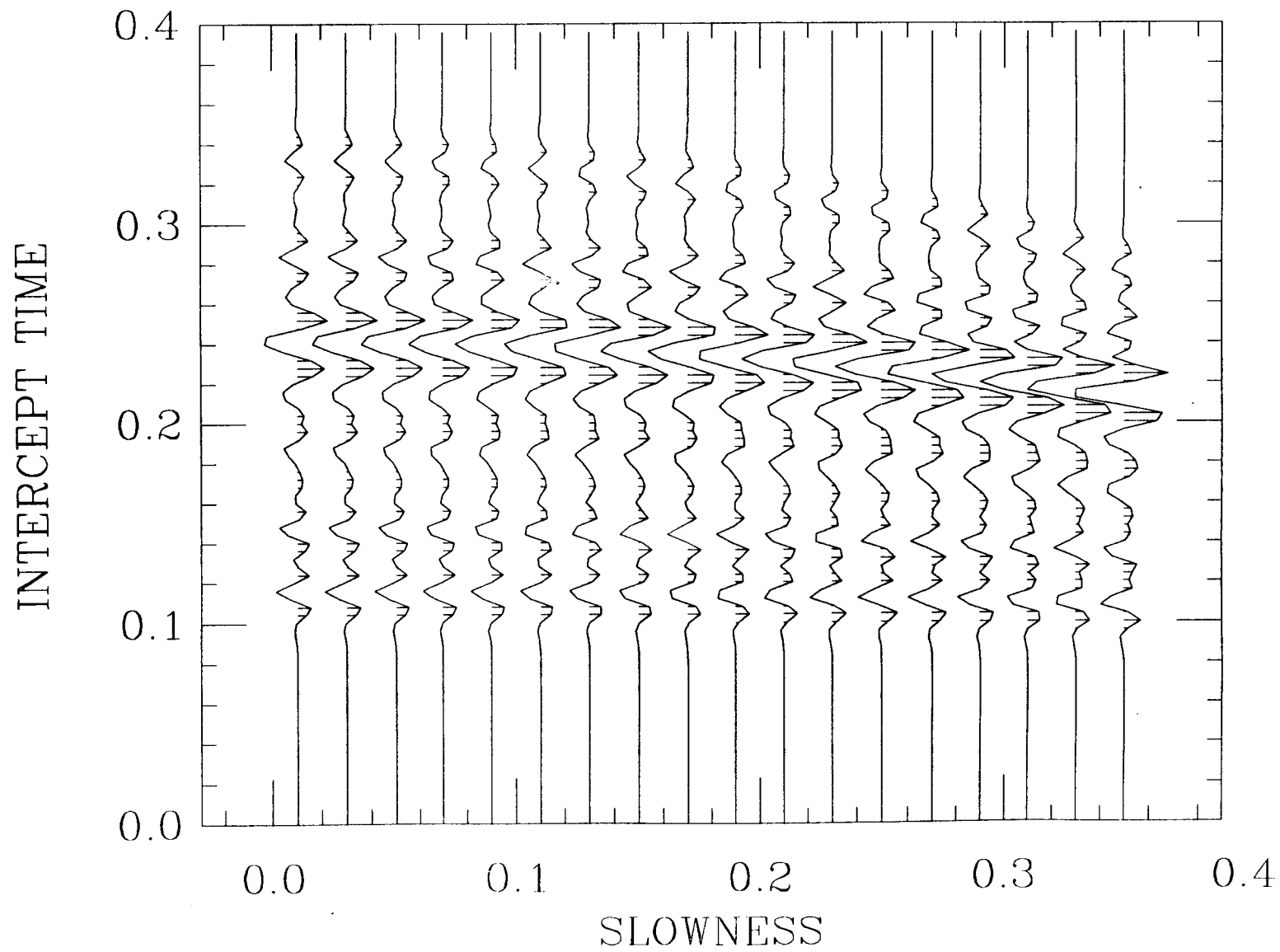


FIGURE 6

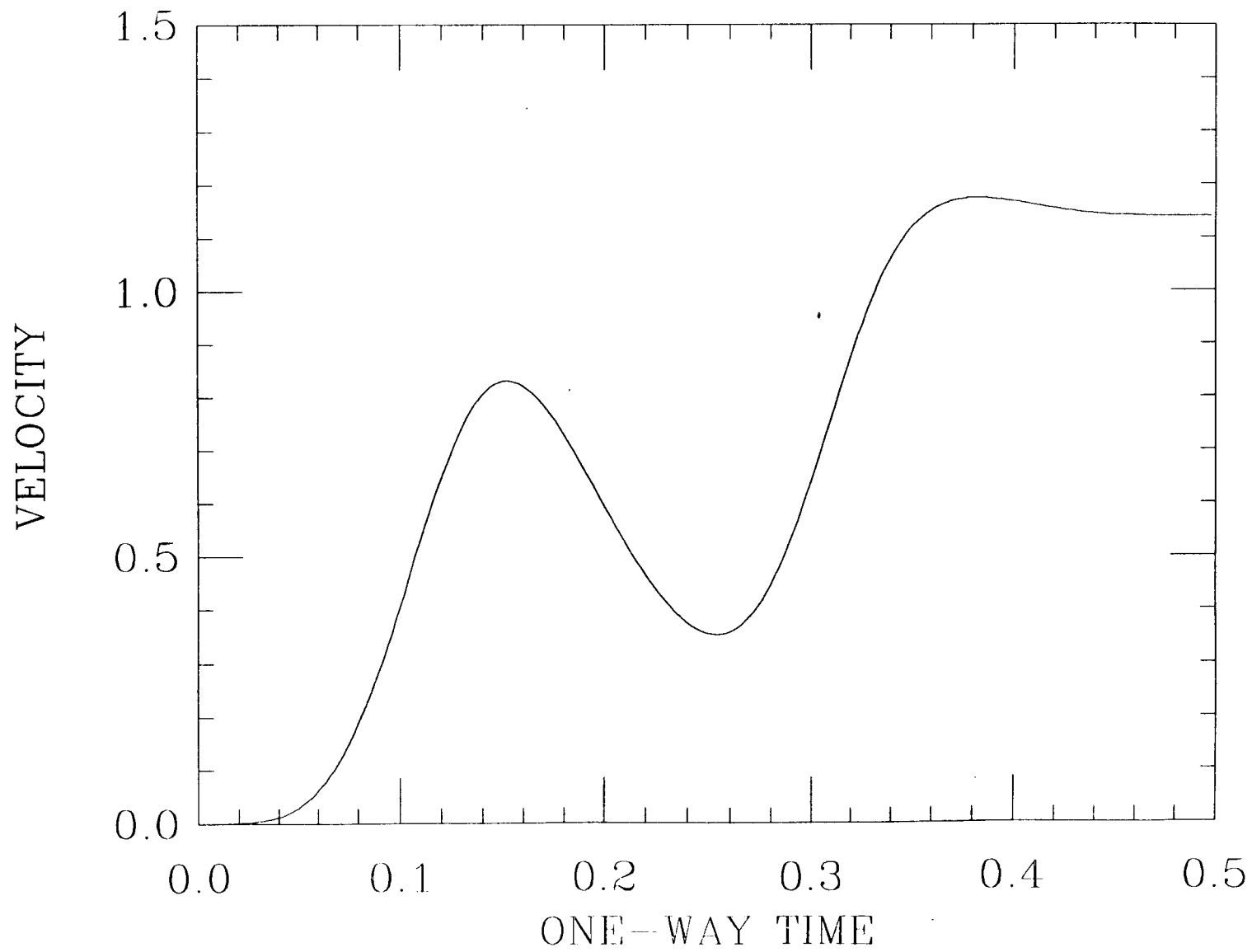


FIGURE 7

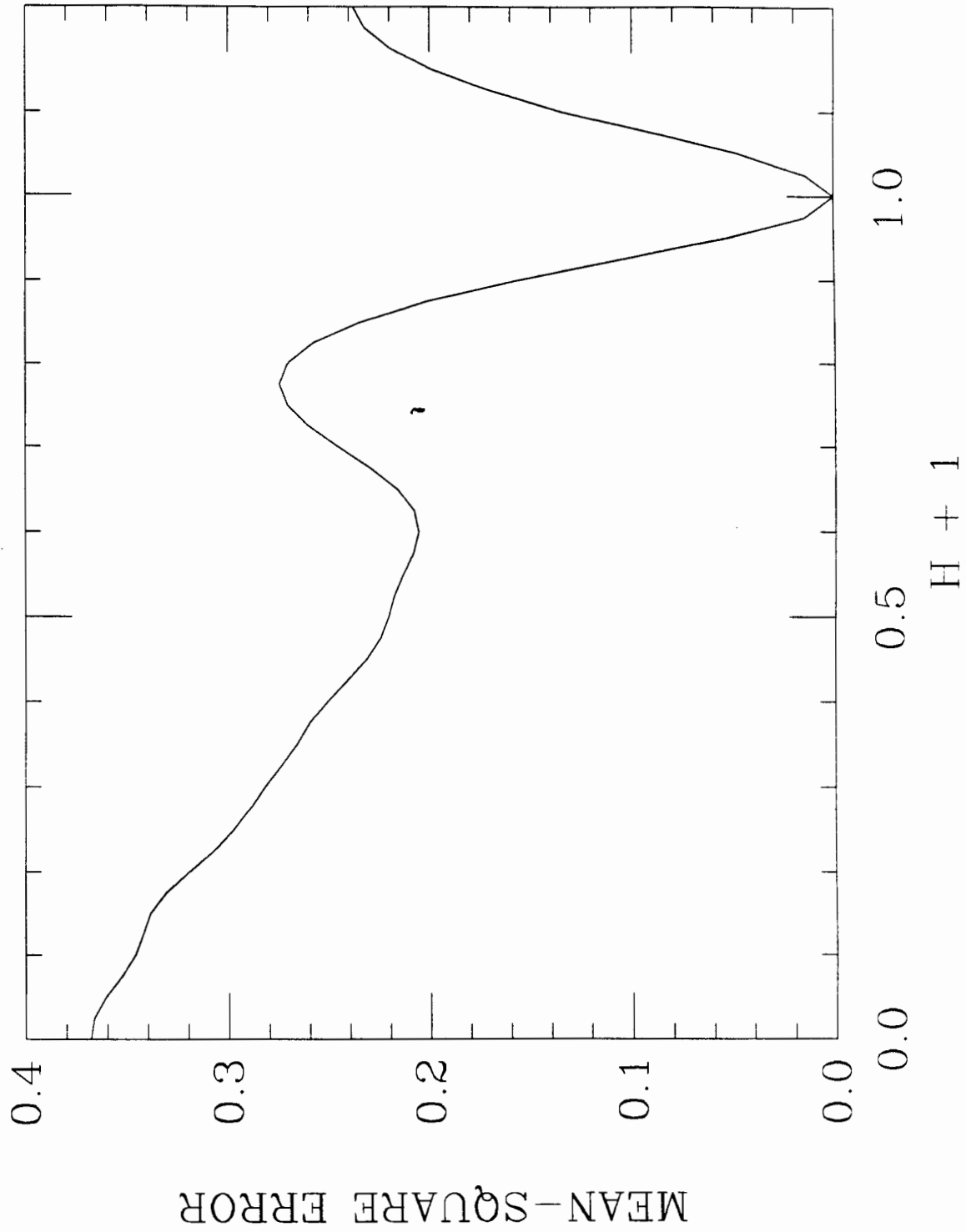


FIGURE 8

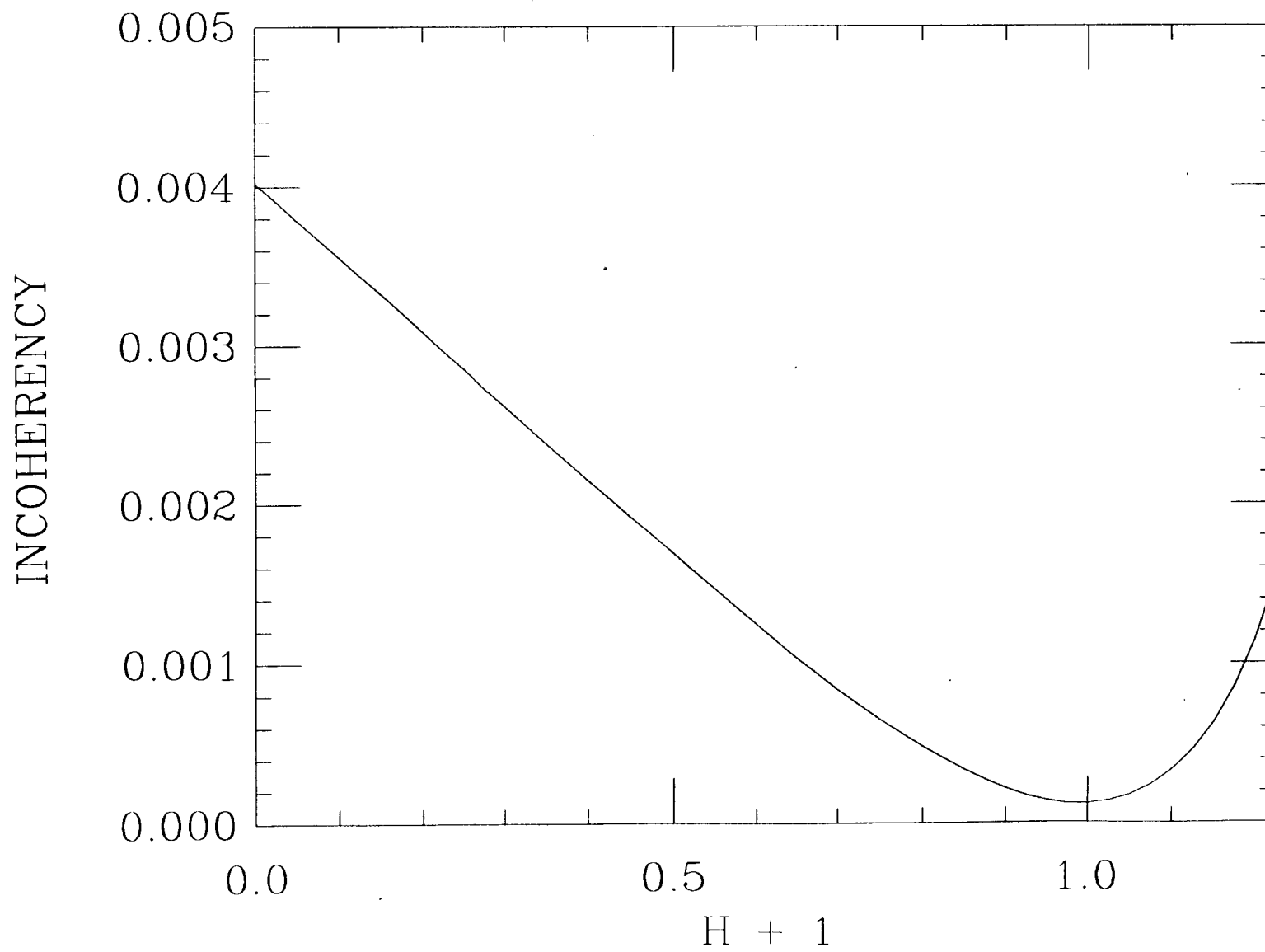


FIGURE 9

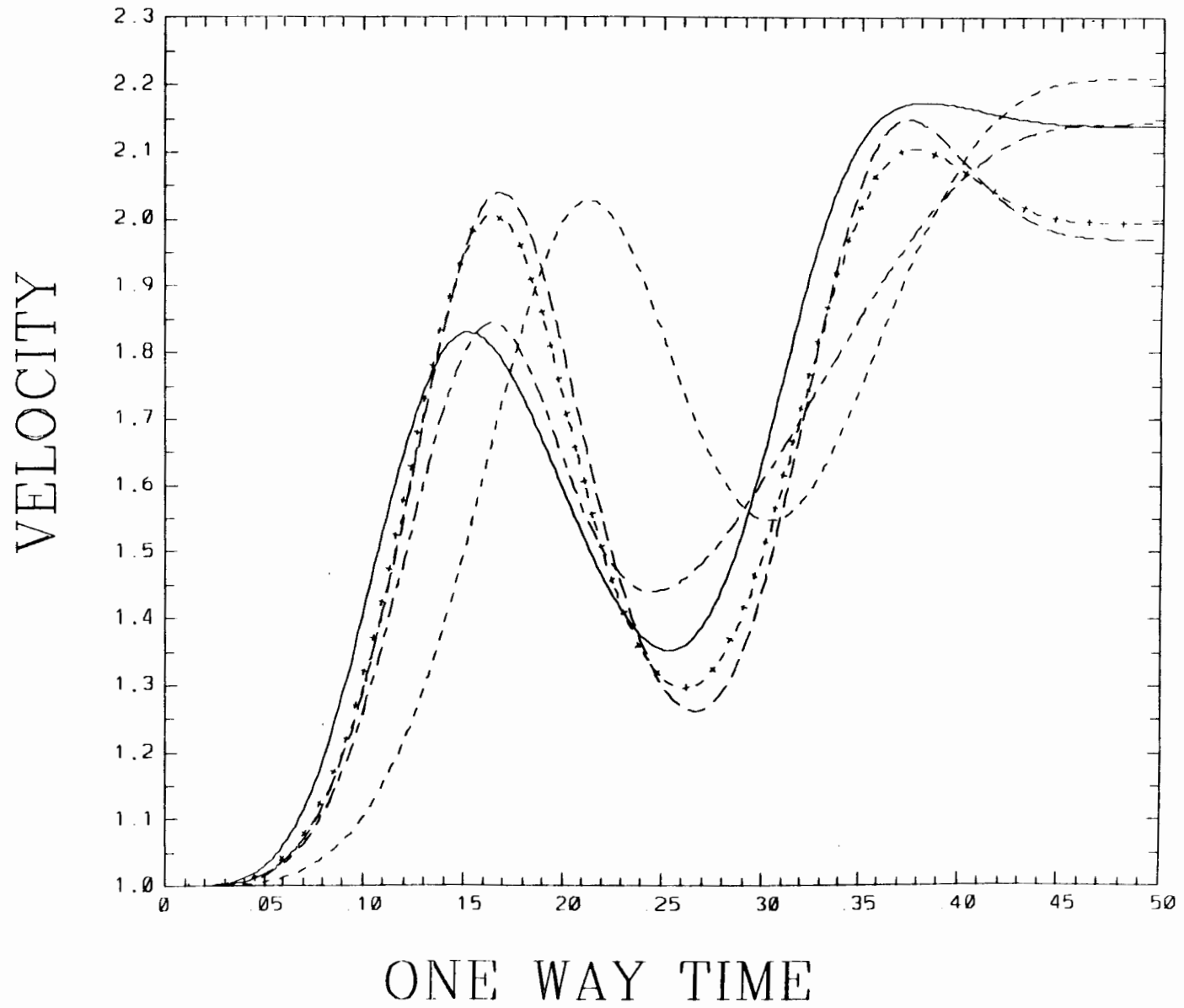


FIGURE 10

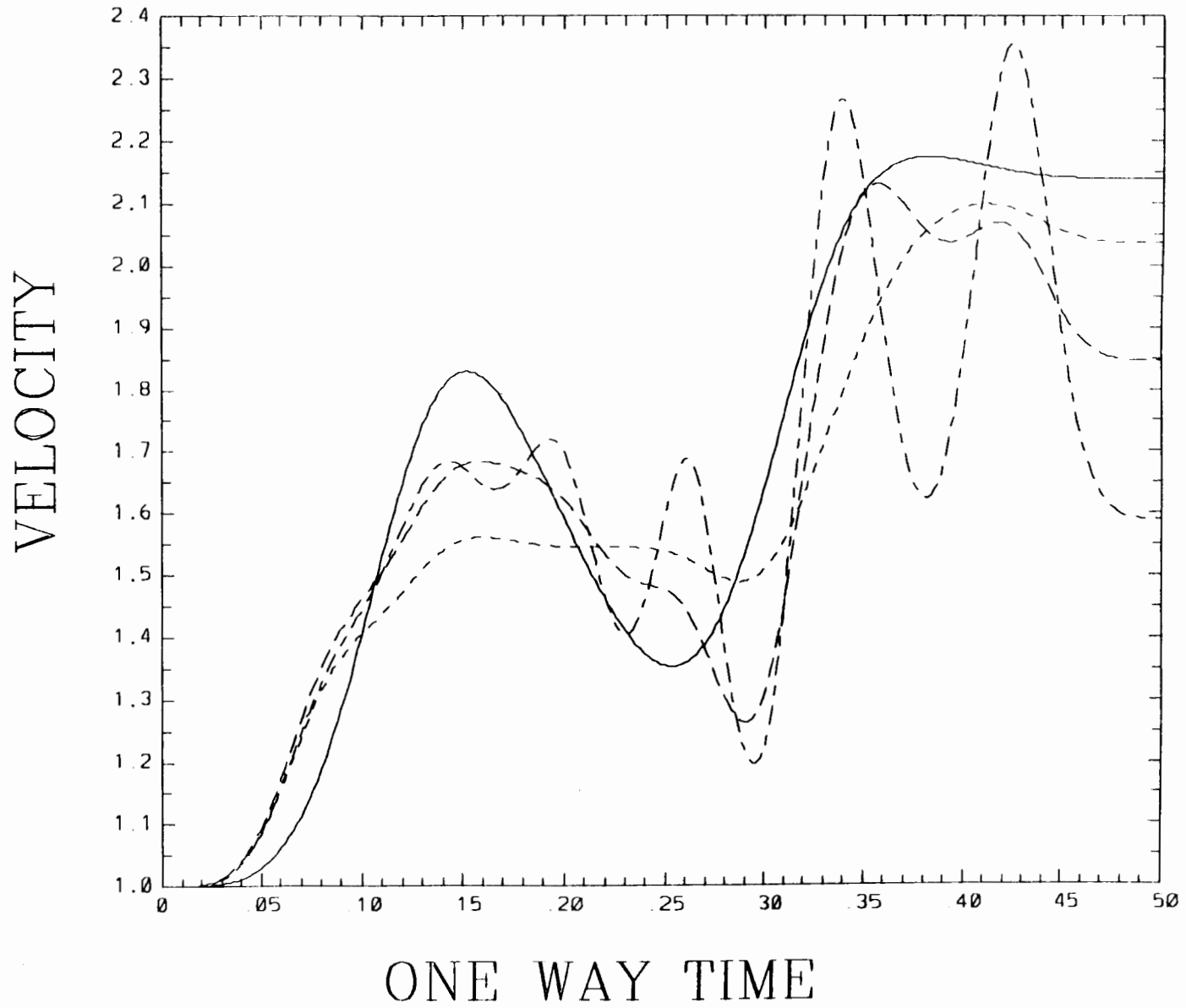


FIGURE 11

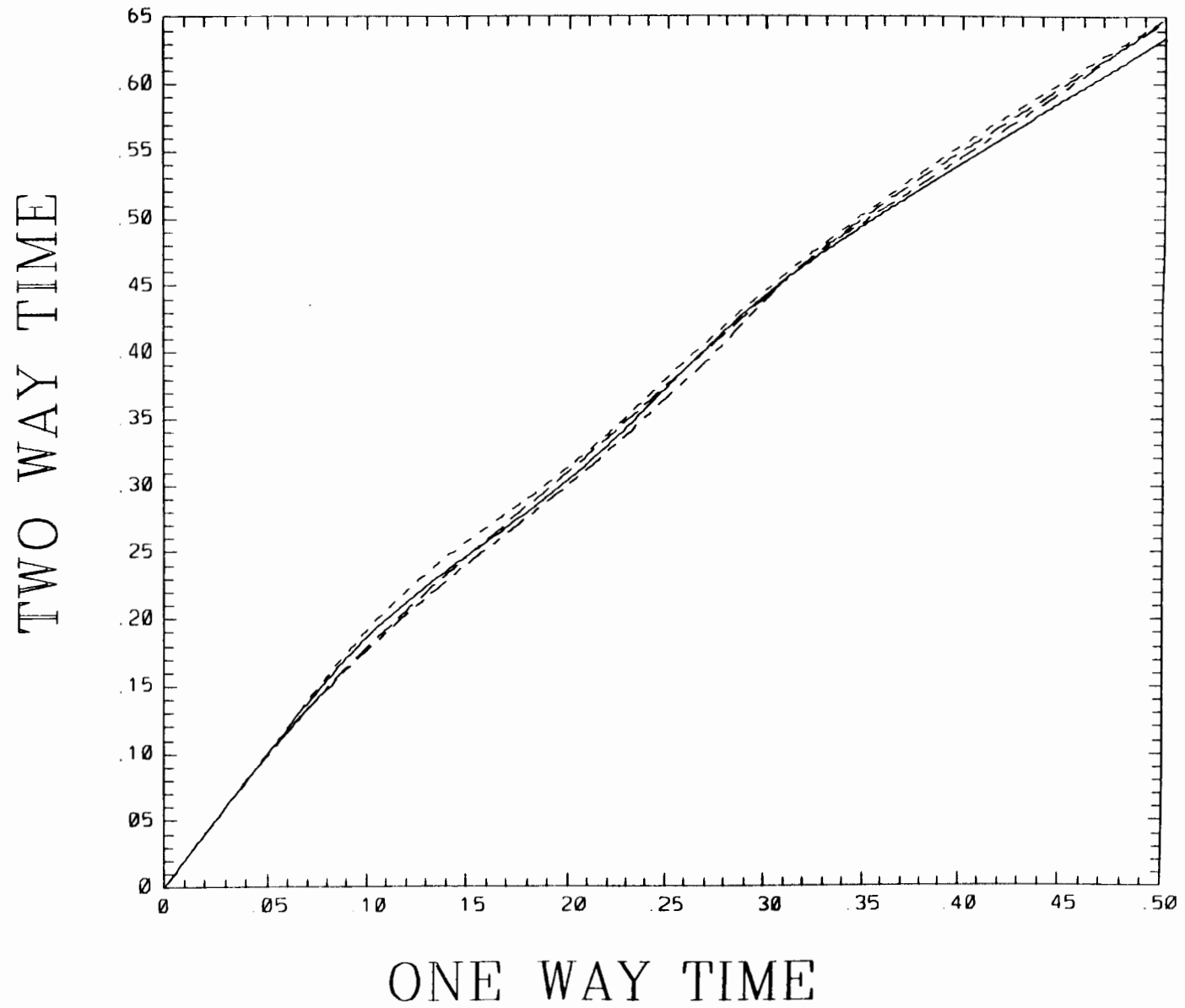


FIGURE 12

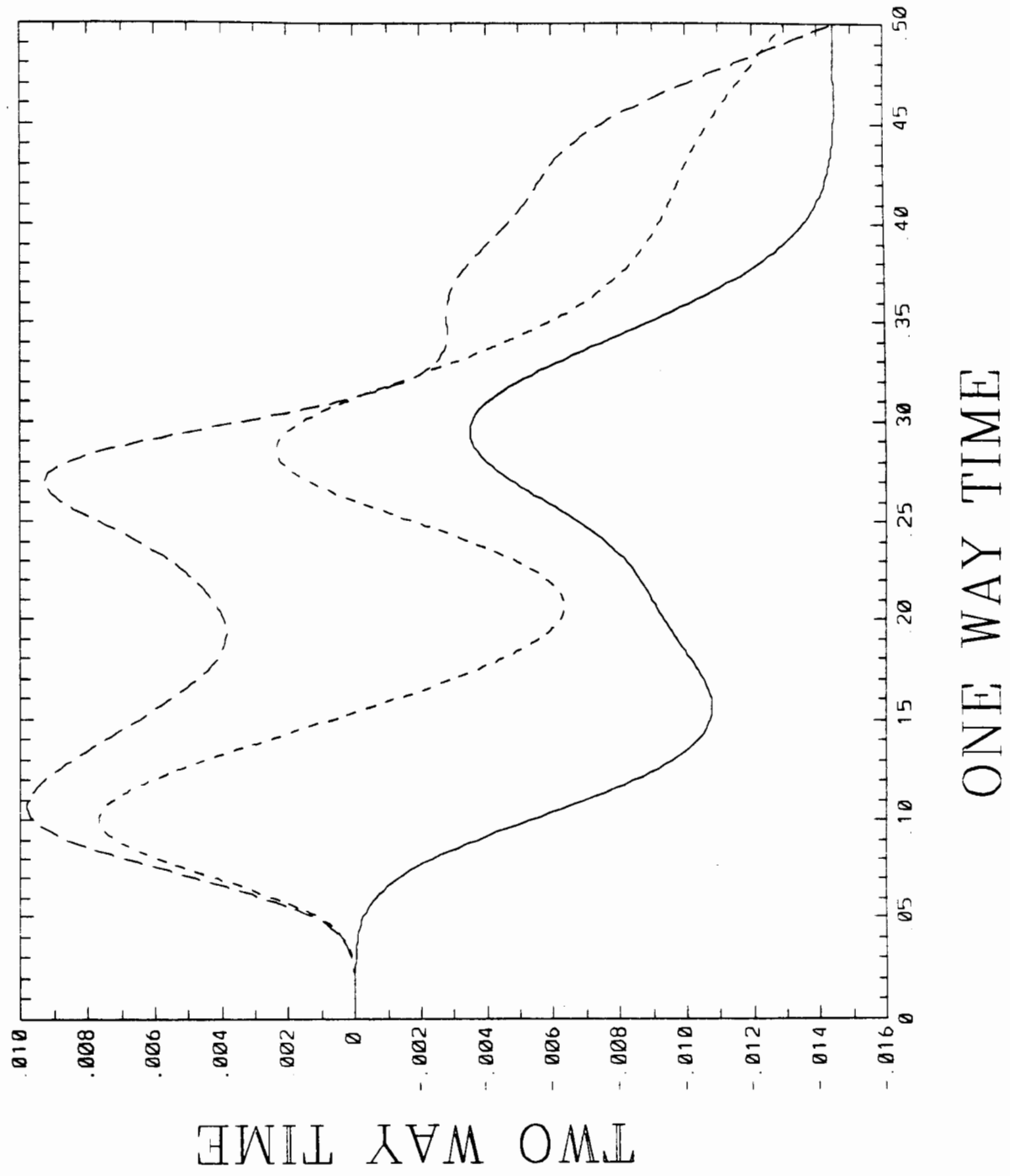


FIGURE 13

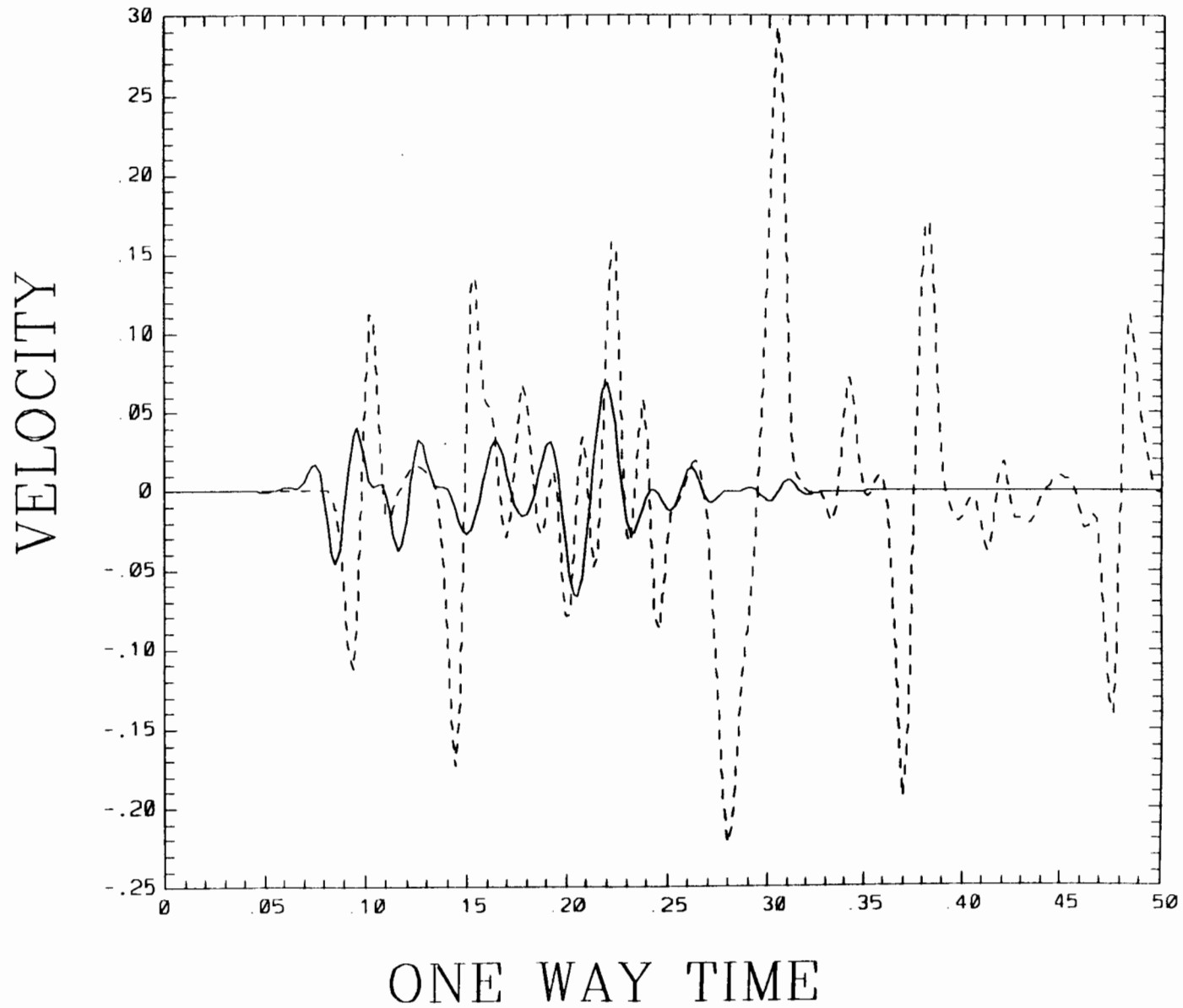


FIGURE 14

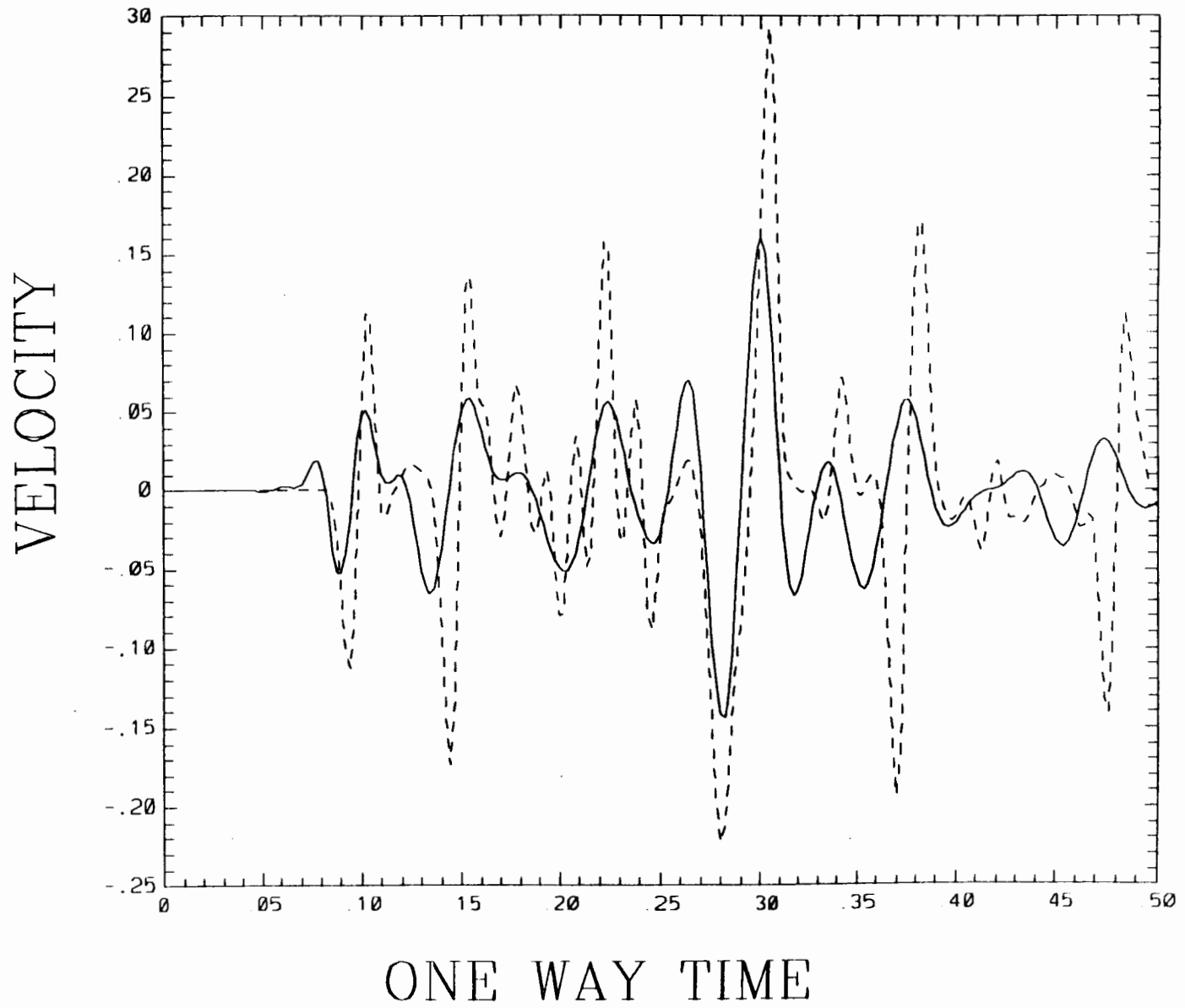


FIGURE 15

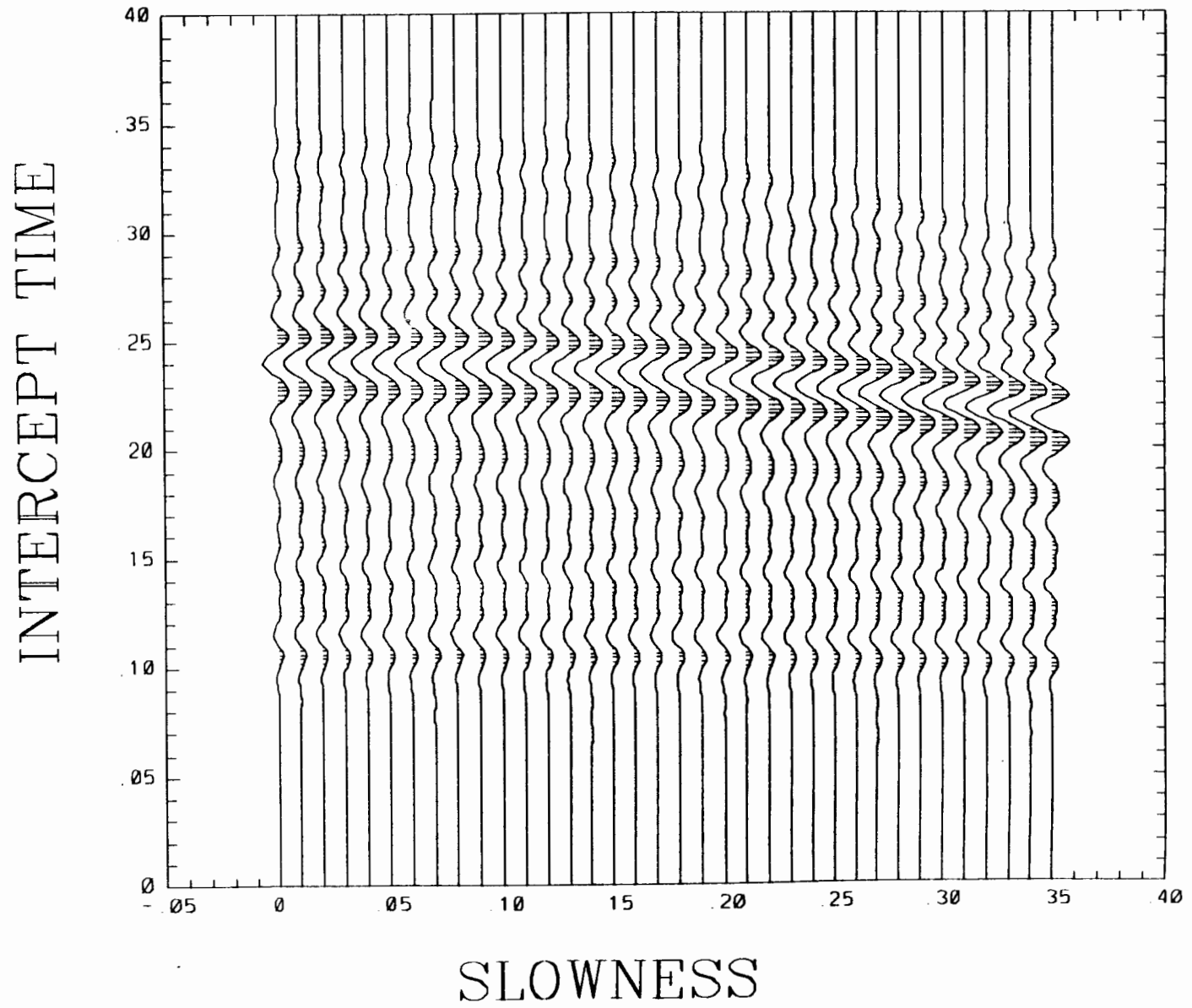


FIGURE 16

