Form Approved
OBM No. 0704-0188

1. Agency Use Only (Leave blank).	2. Report Date.	3. Report Type and Dates Covered.
	July 1992	Final - Journal Article

4. Title and Subtitle.
Least-squares time-delay estimation for transient signals in a multipath environment

5. Funding Numbers.
Program Element No. 0601153N
Project No. 3202
Task No. OHO
Accession No. DN257015
Work Unit No. 12440C

6. Author(s).
R. J. Vaccaro*, C. S. Ramalingam*, D. W. Tufts*, and R. L. Field

7. Performing Organization Name(s) and Address(es).
Naval Research Laboratory
Acoustics Division
Stennis Space Center, MS 39529-5004

8. Performing Organization Report Number.
JA 244:048:91

9. Sponsoring/Monitoring Agency Name(s) and Address(es).
Naval Research Laboratory
Center for Environmental Acoustics
Stennis Space Center, MS 39529-5004

10. Sponsoring/Monitoring Agency Report Number.
JA 244:048:91

11. Supplementary Notes.
Published in J. Acoust. Soc. Am.
*The University of Rhode Island, Kingston, RI 39529-0288

12a. Distribution/Availability Statement.
Approved for public release; distribution is unlimited.

12b. Distribution Code.

13. Abstract *(Maximum 200 words).*

The problem of estimating the arrival times of overlapping ocean-acoustic signals from a noisy received waveform that consists of scaled and delayed replicas of a deterministic transient signal is considered. It is assumed that the transmitted signal and the number of paths in the multipath environment are known, and an algorithm is developed that gives least-squares estimates of the amplitude and time delay of each path. A method is given to ensure that the global minimum of the error surface is found in spite of the existence of numerous local minima. The algorithm is then extended to the case in which the transmitted signal is not known precisely, but is assumed to belong to a parametric class of signals. The extended algorithm additionally obtains the parameters that characterize the transmitted signal. The algorithm is demonstrated on the class of signals consisting of gated sinusoids, using both simulated and experimental data.

14. Subject Terms.	15. Number of Pages.
Transients, distributed sensors, coherence, detection, classification	9
	16. Price Code.

17. Security Classification of Report.	18. Security Classification of This Page.	19. Security Classification of Abstract.	20. Limitation of Abstract.
Unclassified	Unclassified	Unclassified	SAR

Least-squares time-delay estimation for transient signals in a multipath environment

R. J. Vaccaro, C. S. Ramalingam, and D. W Tufts
Department of Electrical Engineering, The University of Rhode Island, Kingston, Rhode Island 02881

R. L. Field
Naval Research Laboratory Detachment, Stennis Space Center, Mississippi 39529-5004

(Received 15 June 1991; revised 1 October 1991; accepted 9 February 1992)

The problem of estimating the arrival times of overlapping ocean-acoustic signals from a noisy received waveform that consists of scaled and delayed replicas of a deterministic transient signal is considered. It is assumed that the transmitted signal and the number of paths in the multipath environment are known, and an algorithm is developed that gives least-squares estimates of the amplitude and time delay of each path. A method is given to ensure that the global minimum of the error surface is found in spite of the existence of numerous local minima. The algorithm is then extended to the case in which the transmitted signal is not known precisely, but is assumed to belong to a parametric class of signals. The extended algorithm additionally obtains the parameters that characterize the transmitted signal. The algorithm is demonstrated on the class of signals consisting of gated sinusoids, using both simulated and experimental data.

PACS numbers: 43.60.Gk, 43.30Wi, 43.30.Ma

INTRODUCTION

Time delay estimation is a well-known problem occurring frequently in the fields of sonar, radar, and geophysics. In this problem, the received waveform consists of delayed and scaled replicas of the transmitted signal. This is the result of multiple reflections and attenuation of the signal in the channel.

The received waveform $r(t)$ can be described mathematically as

$$r(t) = \sum_{k=1}^{M} a_k s(t - \tau_k) + n(t), \quad 0 \leqslant t \leqslant T, \tag{1}$$

where $s(t)$ is the transmitted signal, a_k is the attenuation factor for path k, τ_k is the time delay for path k, M is the number of different paths, and $n(t)$ is the inevitable noise component corrupting the received signal. We say that $s(t)$ is a *deterministic* signal, which means that we do not assume any statistical properties for the signal such as its power spectral density. The signal waveform itself is used by the algorithms developed in this paper. For the purposes of this paper, we define a *transient signal* to be a signal whose duration is less than 2 s.

Even though Eq. (1) has been stated in continuous time, any practical implementation using digital processing techniques would deal only with discrete-time samples. Hence, in what follows we will consider only discrete-time signals. Further, we will assume that the noise is white Gaussian. We will also assume that the number of paths M is known. This latter assumption is not as restrictive as it appears to be, since in many cases the number of paths can be determined quite reasonably from the bathymetry of the channel, the approximate locations of the transmitter and receiver, the sound velocity profile, and a propagation model.[1]

The classical method for estimating the times of arrival is correlating the received waveform with the transmitted waveform.[2] The peaks in the correlator output give the estimates of the arrival times. It can be shown that if the signals are separated in time by more than the duration of the signal autocorrelation function, the correlator is equivalent to the maximum-likelihood estimator.[3] Other approaches are given in Refs. 4, 5, 6, and 7.

A completely different approach was recently proposed by Kirsteins.[8,9] The basic idea of this approach is to view the problem in the frequency domain. Since a delay in the time domain is equivalent to multiplication by an exponential in the frequency domain, the corresponding frequency domain problem is one of fitting weighted complex exponentials to the spectrum of the received signal. Utilizing an iterative method of fitting complex exponentials as in Refs. 10 and 11, this approach provides a way of estimating the times of arrival. However, this algorithm does not work if the Fourier transform of the source signal has any nulls (zeros) in the regions of its spectral support. Even if a spectral sample does not fall exactly on a null, a small sample value could cause serious numerical ill-conditioning of the algorithm. Hence we cannot apply this method for a source signal such as a gated sinusoid. More importantly, the delay estimates obtained *via* this method are biased and do not correspond to the true parameter estimates, as demonstrated in Sec. IV A. In addition, contiguous frequency samples have to be considered, which might not be desirable, as will be explained later.

All the above methods require the source signal to be known. In this paper, we develop an algorithm for the case of a known source, and then extend this algorithm to the case in which the source signal is not known precisely, but is as-

sumed to belong to a parametric class of signals. That is, we assume that the transmitted signal waveform depends on the values of certain constant parameters (i.e., the parameters are *not* random variables). Our objective, then, will be to obtain good estimates for the delays and at the same time extract the parameters of the source signal. An example of a parametric class of signals is the class of rectangular pulses. A signal in this class is characterized by three parameters: duration, amplitude, and starting point.

In the next section, we set up the least-squares problem to be solved. In Sec. II, we develop an algorithm assuming that the transmitted signal is known. Section III extends this algorithm to the case when the signal is unknown, but can be described by a set of parameters. In Sec. IV we give the results of the known signal and the unknown signal algorithms using simulated as well as experimental data. The transmitted signal is a gated sinusoid in these examples. In Sec. V we give a discussion of the results. Finally, Sec. VI contains the conclusions.

I. THE LEAST-SQUARES ESTIMATOR

Assuming the source signal to belong to a parametric class of signals, the sampled received waveform can be modeled as

$$r[n] = \sum_{k=1}^{M} a_k s[n - \tau_k;\boldsymbol{\theta}] + w[n], \quad 0 \leqslant n \leqslant N-1, \tag{2}$$

where $\boldsymbol{\theta}$ is the vector of parameters that characterize the source signal. If the noise $w[n]$ is white Gaussian, then the least-squares estimator is also the maximum likelihood estimator.[12] The squared error function is given by

$$E(\boldsymbol{\theta},\tau_k,a_k) = \sum_{n=0}^{N-1} \left[r[n] - \sum_{k=1}^{M} a_k s[n - \tau_k;\boldsymbol{\theta}] \right]^2, \tag{3}$$

and the objective is to minimize E with respect to all of its arguments.

In the above time-domain formulation of the problem, we are restricted to values of τ_k that are integer multiples of the basic sampling interval. If a more accurate estimate is required, then one has to resort to interpolation. This inconvenience can be avoided by transforming the problem to the frequency domain, where τ_k can take on a continuum of values.

The data record in (1) exists in the time interval $[0,T]$. If this data record is set to zero for $t<0$ and $t>T$, then the summand in Eq. (3) will be zero outside the range $0 \leqslant n \leqslant N-1$. Thus we can extend the summation range from minus infinity to plus infinity in (3) and invoke Parseval's theorem to get

$$E(\boldsymbol{\theta},\tau_k,a_k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| R(\omega) - S(\omega;\boldsymbol{\theta}) \sum_{k=1}^{M} a_k e^{-j\tau_k \omega} \right|^2 d\omega, \tag{4}$$

where $R(\omega)$ and $S(\omega;\boldsymbol{\theta})$ are the discrete-time Fourier transforms of $r[n]$ and $s[n;\boldsymbol{\theta}]$, respectively.

The problem, now, is to approximate $R(\omega)$ by a weighted sum of complex exponentials. For computer implementation, we sample in the frequency domain at intervals of $\Delta\omega$. Hence,

$$e^{-j\tau_k \omega} \rightarrow e^{-j\tau_k n\Delta\omega} = e^{j\lambda_k n},$$

where

$$\lambda_k = -\tau_k \Delta\omega. \tag{5}$$

The integrals are then replaced by sums and the squared error function is given approximately by (after omitting the scale factor $1/2\pi$)

$$E(\boldsymbol{\theta},\lambda_k,a_k) = \sum_{n=0}^{N-1} \left| R[n] - S[n;\boldsymbol{\theta}] \sum_{k=1}^{M} a_k e^{j\lambda_k n} \right|^2, \tag{6}$$

where $S[n;\boldsymbol{\theta}] = S(n\Delta\omega;\boldsymbol{\theta})$ and $R[n] = R(n\Delta\omega)$. The above error expression will be used to develop both the known signal and the unknown signal algorithms. For the known signal case, the parameter vector $\boldsymbol{\theta}$ in Eq. (6) is fixed and is not included in the minimization, whereas for the unknown signal case Eq. (6) is minimized with respect to all the parameters.

II. THE KNOWN SIGNAL ALGORITHM FOR NARROWBAND SIGNALS

In this section we consider the case when the source signal is known. In the next section, we show how the known-signal algorithm can be used iteratively in the case when the signal is not known.

When the signal is narrow band, most of the energy of the signal is concentrated in the passband. For example, if the signal were a gated sinusoid, most of its energy is concentrated near the main lobe in the frequency domain. Recall that the spectrum of a gated sinusoid is a sinc function centered at the frequency of the sinusoid. In this case, we do not have to include all frequency points in the minimization but consider only those values near the main lobe. Going even further, one need not again include *all* frequency points near the main lobe, but *only those which have significant signal energy*. That is, one can threshold the signal spectrum and consider only those points which are above this value. This has the advantage of working only with those frequency points having good signal-to-noise ratios. Also, since the received and the modeled signal are real, only half the signal spectrum needs to be considered because of its conjugate symmetry. The expression for the error in Eq. (6) is thus rewritten as,

$$E(\boldsymbol{\lambda},\tilde{\mathbf{a}}) = \sum_{n \in \mathcal{N}} \left| R[n] - S[n] \sum_{k=1}^{M} a_k e^{j\lambda_k n} \right|^2, \tag{7}$$

where $\boldsymbol{\lambda}$ is the vector of scaled delays [see Eq. (5)] and $\tilde{\mathbf{a}}$ is a vector of amplitudes. (The tildes are used now to avoid cumbersome notation in subsequent developments.) The notation $n \in \mathcal{N}$ means that the summation is over those values of n which belong to the set $\mathcal{N}$, and not necessarily over contiguous values. The set $\mathcal{N}$ consists of those frequency points at which the magnitude of the signal spectrum exceeds a threshold. A rule-of-thumb is to set this threshold to roughly one-twentieth to one-tenth the peak magnitude of the signal

spectrum. Since matrix notation leads to succinct expressions, let us define the following:

$$\boldsymbol{\lambda} = (\lambda_1 \quad \lambda_2 \quad \cdots \quad \lambda_M)^T,$$

$$\tilde{\mathbf{r}} = (R[q_1] \quad R[q_2] \quad \cdots \quad R[q_N])^T,$$

$$\tilde{\mathbf{a}} = (a_1 \quad a_2 \quad \cdots \quad a_M)^T,$$

$$\mathbf{S} = \text{diag}(S[q_1] \quad S[q_2] \quad \cdots \quad S[q_N]),$$

$$\mathbf{A}(\boldsymbol{\lambda}) = \begin{pmatrix} e^{j\lambda_1 q_1} & e^{j\lambda_2 q_1} & \cdots & e^{j\lambda_M q_1} \\ e^{j\lambda_1 q_2} & e^{j\lambda_2 q_2} & \cdots & e^{j\lambda_M q_2} \\ \vdots & \vdots & \vdots & \vdots \\ e^{j\lambda_1 q_N} & e^{j\lambda_2 q_N} & \cdots & e^{j\lambda_M q_N} \end{pmatrix},$$

$$\tilde{\mathbf{P}}(\boldsymbol{\lambda}) = \mathbf{S}\mathbf{A}(\boldsymbol{\lambda}), \tag{8}$$

where $q_k \in \mathcal{V}$, $k = 1,2,\ldots,N$ denote the frequency samples that are selected. Hence Eq. (7) can be recast as

$$E(\boldsymbol{\lambda},\tilde{\mathbf{a}}) = \|\tilde{\mathbf{r}} - \tilde{\mathbf{P}}(\boldsymbol{\lambda})\tilde{\mathbf{a}}\|^2. \tag{9}$$

This is the error expression considered in Ref. 9. However, in that work the amplitudes were allowed to be complex. As mentioned in the Introduction, this leads to a biased estimate of the delay parameters if the SNR were not sufficiently high (we show this in Sec. IV A). But an important advantage of allowing complex-valued amplitudes is that the corresponding error surface is reasonably smooth, making it easier to find its global minimum. On the other hand, while the global minimum of the error surface defined by Eq. (9) with the amplitudes constrained to be real yields the true delay parameters, finding this minimum is extremely difficult as the surface is characterized by numerous closely spaced local minima. We now outline a procedure that leads us to this global minimum, starting with the error surface corresponding to the complex amplitudes.

The key step that helps us to achieve the transition from the biased error surface (due to the complex amplitudes) to the true error surface (due to the real amplitudes) is to rewrite Eq. (9) explicitly in terms of real and imaginary parts of each of the terms involved, and add a penalty term to the imaginary part of the complex amplitude. That is, let

$$E_\alpha(\boldsymbol{\lambda},\mathbf{a}) = \left\| \begin{pmatrix} \mathbf{r}_r \\ \mathbf{r}_i \end{pmatrix} - \begin{pmatrix} \mathbf{P}_r & -\mathbf{P}_i \\ \mathbf{P}_i & \mathbf{P}_r \end{pmatrix} \begin{pmatrix} \mathbf{a}_r \\ \mathbf{a}_i \end{pmatrix} \right\|^2 + \alpha\|\mathbf{a}_i\|^2, \tag{10}$$

where

$$\mathbf{r}_r = \text{Re}\{\tilde{\mathbf{r}}\}, \quad \mathbf{r}_i = \text{Im}\{\tilde{\mathbf{r}}\},$$

$$\mathbf{P}_r = \text{Re}\{\tilde{\mathbf{P}}(\boldsymbol{\lambda})\}, \quad \mathbf{P}_i = \text{Im}\{\tilde{\mathbf{P}}(\boldsymbol{\lambda})\},$$

$$\mathbf{a}_r = \text{Re}\{\tilde{\mathbf{a}}\}, \quad \mathbf{a}_i = \text{Im}\{\tilde{\mathbf{a}}\}.$$

Rewriting the right-hand side of Eq. (10) as a single norm, we get

$$E_\alpha(\boldsymbol{\lambda},\mathbf{a}) = \left\| \begin{pmatrix} \mathbf{r}_r \\ \mathbf{r}_i \\ \mathbf{0} \end{pmatrix} - \begin{pmatrix} \mathbf{P}_r & -\mathbf{P}_i \\ \mathbf{P}_i & \mathbf{P}_r \\ \mathbf{0} & \alpha\mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{a}_r \\ \mathbf{a}_i \end{pmatrix} \right\|^2 = \|\mathbf{r} - \mathbf{P}(\boldsymbol{\lambda})\mathbf{a}\|^2, \tag{11}$$

where

$$\mathbf{P}(\boldsymbol{\lambda}) = \begin{pmatrix} \mathbf{P}_r & -\mathbf{P}_i \\ \mathbf{P}_i & \mathbf{P}_r \\ \mathbf{0} & \alpha\mathbf{I} \end{pmatrix}, \quad \mathbf{r} = \begin{pmatrix} \mathbf{r}_r \\ \mathbf{r}_i \\ \mathbf{0} \end{pmatrix}, \quad \mathbf{a} = \begin{pmatrix} \mathbf{a}_r \\ \mathbf{a}_i \end{pmatrix}. \tag{12}$$

It is easy to see that for $\alpha = 0$, the error function in Eq. (11) allows the amplitudes to be complex since there is no penalty attached to the imaginary part. As α increases, the penalty on the imaginary part increases, and in the limit as $\alpha \to \infty$, the amplitudes are constrained to be purely real. Note that the model for the channel assumes the amplitudes to be real.

With the above modification of the error expression, the procedure to minimize $E_\alpha(\boldsymbol{\lambda},\mathbf{a})$ is as follows: Initially, α is set equal to zero. Then $E_0(\boldsymbol{\lambda},\mathbf{a})$ is minimized with respect to $\boldsymbol{\lambda}$ and $\mathbf{a}$. The minimum yields a biased estimate of $\boldsymbol{\lambda}$, but one which is reasonably close to the true value. To estimate the true delay values, α is increased gradually and the minimum of $E_\alpha(\boldsymbol{\lambda},\mathbf{a})$ is found for each α. The minimum of $E_\infty(\boldsymbol{\lambda},\mathbf{a})$ yields the final parameter estimates. This method of using a quadratic penalty function to impose a nonlinear constraint is known to converge to a minimum of the true error surface.[13]

For any fixed $\boldsymbol{\lambda}$ and α, the best $\mathbf{a}$ which minimizes $E_\alpha(\boldsymbol{\lambda},\mathbf{a})$ is given by

$$\mathbf{a} = (\mathbf{P}^H\mathbf{P})^{-1}\mathbf{P}^H\mathbf{r}. \tag{13}$$

Substituting this value of $\mathbf{a}$ in Eq. (11) we get

$$E_\alpha(\boldsymbol{\lambda}) = \|[\mathbf{I} - \mathbf{P}(\mathbf{P}^H\mathbf{P})^{-1}\mathbf{P}^H]\mathbf{r}\|^2 = \|\mathbf{P}^\perp\mathbf{r}\|^2, \tag{14}$$

where $\mathbf{P}^\perp = \mathbf{I} - \mathbf{P}(\mathbf{P}^H\mathbf{P})^{-1}\mathbf{P}^H$, and now the error is a function of the delay parameter vector $\boldsymbol{\lambda}$ only. In the next subsection, we present a Gauss–Newton approach to find the minimum of $E_\alpha(\boldsymbol{\lambda})$.

A. A Gauss–Newton approach to minimizing $E_\alpha(\boldsymbol{\lambda})$

A common algorithm for minimizing an objective function, which is expressed as a squared norm, is the Gauss-Newton algorithm.[13,14] This algorithm uses a first-order perturbation expansion to convert the nonlinear minimization problem to a linear one. A sequence of linear problems are then solved until the solutions converge. In this section, we derive the formulas needed to implement the Gauss–Newton algorithm for the error function in Eq. (11).

We begin by considering the value of the error function at an increment $\Delta\boldsymbol{\lambda}$ away from the nominal value of $\boldsymbol{\lambda}$, i.e.,

$$E_\alpha(\boldsymbol{\lambda} + \Delta\boldsymbol{\lambda}) = \|[\mathbf{I} - \mathbf{P}(\boldsymbol{\lambda} + \Delta\boldsymbol{\lambda})(\mathbf{P}^H(\boldsymbol{\lambda} + \Delta\boldsymbol{\lambda}) \times \mathbf{P}(\boldsymbol{\lambda} + \Delta\boldsymbol{\lambda}))^{-1}\mathbf{P}^H(\boldsymbol{\lambda} + \Delta\boldsymbol{\lambda})]\mathbf{r}\|^2. \tag{15}$$

To simplify the above equation, we consider a first-order perturbation expansion of $\mathbf{P}(\boldsymbol{\lambda} + \Delta\boldsymbol{\lambda})$:

$$\mathbf{P}(\boldsymbol{\lambda} + \Delta\boldsymbol{\lambda}) \doteq \mathbf{P} + \Delta\mathbf{P}, \tag{16}$$

where $\doteq$ means "equal to first order," and $\Delta\mathbf{P}$ is a matrix which is a linear function of $\Delta\boldsymbol{\lambda}$. Substituting this expansion of $\mathbf{P}(\boldsymbol{\lambda} + \Delta\boldsymbol{\lambda})$ into Eq. (15) and retaining only first order terms (see Ref. 15 for details) we get

$$E_\alpha(\lambda+\Delta\lambda)=\|\mathbf{P}^\perp\mathbf{r}-\mathbf{P}^\perp(\Delta\mathbf{P})(\mathbf{P}^H\mathbf{P})^{-1}\mathbf{P}^H\mathbf{r}-\mathbf{P}(\mathbf{P}^H\mathbf{P})^{-1}(\Delta\mathbf{P})^H\mathbf{P}^\perp\mathbf{r}\|^2, \quad (17)$$

where the superscript H denotes complex-conjugate transpose. Using the definition of $\mathbf{P}$ from Eq. (12) it can be verified that

$$\Delta\mathbf{P}=-\underbrace{\begin{pmatrix}\mathbf{QP}_i & \mathbf{QP}_r\\ -\mathbf{QP}_r & \mathbf{QP}_i\\ \mathbf{0} & \mathbf{0}\end{pmatrix}}_{\mathbf{S}}\begin{pmatrix}\Delta\mathbf{\Lambda} & \mathbf{0}\\ \mathbf{0} & \Delta\mathbf{\Lambda}\end{pmatrix}, \quad (18)$$

where

$$\mathbf{Q}=\mathrm{diag}\{q_1 \quad q_2 \quad \cdots \quad q_N\},$$
$$\Delta\lambda=(\Delta\lambda_1 \quad \Delta\lambda_2 \quad \cdots \quad \Delta\lambda_M)^T,$$
$$\Delta\mathbf{\Lambda}=\mathrm{diag}\{\Delta\lambda\}.$$

Since

$$\mathrm{diag}\{x_1 \quad x_2 \quad \cdots \quad x_L\}\begin{pmatrix}y_1\\ y_2\\ \vdots\\ y_L\end{pmatrix}=\mathrm{diag}\{y_1 \quad y_2 \quad \cdots \quad y_L\}\begin{pmatrix}x_1\\ x_2\\ \vdots\\ x_L\end{pmatrix},$$

we can pull $\mathrm{diag}\{\Delta\lambda;\Delta\lambda\}$ to the end in the second and third terms on the right-hand side of Eq. (17), and further simplify to get

$$E_\alpha(\lambda+\Delta\lambda)=\|\mathbf{x}-\mathbf{B}\Delta\lambda\|^2, \quad (19)$$

where

$$\mathbf{x}=\mathbf{P}^\perp\mathbf{r},$$
$$\mathbf{B}'=\mathbf{P}^\perp\mathbf{S}\cdot\mathrm{diag}\{(\mathbf{P}^H\mathbf{P})^{-1}\mathbf{P}^H\mathbf{r}\}+\mathbf{P}(\mathbf{P}^H\mathbf{P})^{-1}\cdot\mathrm{diag}\{\mathbf{S}^T\mathbf{P}^\perp\mathbf{r}\},$$
$$\mathbf{B}=\mathbf{B}'(:,1{:}M)+\mathbf{B}'(:,(M+1){:}2M).$$

(Note that we have used standard MATLAB notation in the definition of $\mathbf{B}$, which is derived from $\mathbf{B}'$ by summing the first M columns with its second M columns.) From Eq. (19) we can solve for the best $\Delta\lambda$ (in the least-squares sense) as

$$\Delta\lambda=\mathbf{B}^{\#}\mathbf{x}=(\mathbf{B}^H\mathbf{B})^{-1}\mathbf{B}^H\mathbf{x}. \quad (20)$$

Now, λ is replaced by $\lambda+\Delta\lambda$ and the procedure is repeated till $\|\Delta\lambda\|_2<\epsilon$. In many cases, it might be desirable not to make large changes in λ to avoid oscillations. If so, the least-squares solution in Eq. (20) is replaced by a constrained least-squares solution,[16] where $\|\Delta\lambda\|_2$ is restricted to not to exceed a specified value, say δ.

An initial value for the parameter vector λ must be estimated before (20) can be used to obtain improved parameter estimates. A standard procedure for generating initial estimates is the *coordinate descent algorithm* which is described in Refs. 17 and 18. We used the coordinate descent algorithm on $E_\alpha(\lambda)$ given in (14) with $\alpha=0$ to obtain initial estimates for the known signal algorithm.

III. THE UNKNOWN SIGNAL ALGORITHM

We now address the problem of time-delay estimation when the source signal is unknown. In this case, the source parameter vector θ will also enter the minimization in Eq. (3). In the algorithm outlined below, we assume the source to belong to a parametric class of signals, viz., gated sinusoids of unknown frequency and duration.

First, we observe that the source signal can completely be specified (except for a real-valued scale factor) by its frequency, duration, starting point, and phase. If the frequency and duration are such that there are many cycles of the signal present, then we need not precisely determine the phase, which can be set equal to zero. Also, since all time delays are relative to the source pulse, the starting point can be assumed to be at $t=0$, without loss of generality.

With the above assumptions, the overall problem reduces to minimizing

$$\min_{\lambda,a,f,d}\sum_{n=1}\left|R[n]-S[n;f,d]\sum_{k=1}^{M}a_k e^{j\lambda_k n}\right|^2,$$

where f and d are the frequency and duration of the transmitted signal, respectively. This expression is not easy to minimize with respect to all the parameters simultaneously. Instead, we resort to an iterative minimization technique outlined below.

(1) Obtain initial estimates of f^0 and d^0, i.e., of the unknown frequency and duration, respectively.

(2) Use f^i and d^i in the known signal algorithm to estimate λ^i and $\mathbf{a}^i$.

(3) Using the estimated λ^i and $\mathbf{a}^i$, calculate f^{i+1} and d^{i+1}.

(4) Check for convergence to return to step 2.

We now show how to calculate f^{i+1} and d^{i+1} mentioned in step 3. Consider the error function for a given λ and $\mathbf{a}$:

$$E(f,d)=\sum_{n=1}\left|R[n]-S[n;f,d]\sum_{k=1}^{M}a_k e^{j\lambda_k n}\right|^2.$$

Here, $E(f,d)$ is nonlinear in f and d and an analytical minimization is again very difficult. We simplify the minimization by taking advantage of the fact that d is a discrete variable since the signal is sampled, i.e., only finite number of values for d are possible. Hence, $E(f,d)$ can be minimized as follows:

(1) Carry a search over values of d near the previous estimate.

(2) For each d, find the best value of f using a gradient descent technique.

(3) Repeat step 2 until minimum is found.

The final question is how to obtain f^0 and d^0. In our model we assume that the first path is the least attenuated one and that its amplitude is greater than that of the noise. There may or may not be pulses from other paths overlap-

ping with the first. With these assumptions we get rough initial guesses for f^0 and d^0.

(i) Obtain the envelope of the signal.

(ii) Obtain the high-amplitude portion of the received waveform by finding that part of the signal envelope which exceeds a threshold. Let (t_1, t_2) be the interval in which the envelope exceeds the threshold. Choose $d^0 = (t_2 - t_1)/2$.

(iii) f^0 is obtained using any standard frequency estimator on the signal in the interval (t_1, t_2).

This completes the description of the unknown source algorithm.

IV. ALGORITHM PERFORMANCE

In this section we show the performance of the algorithms described in the previous sections using synthetic signals as well as experimental data. We first present a one-path example to illustrate the nature of the $E_0(\lambda)$ and $E_\infty(\lambda)$ error surfaces. We then demonstrate the effect of increasing α on $E_\alpha(\lambda)$. Finally, results on experimental data using the known- and the unknown-signal algorithms are presented.

A. A one-path example to compare error surfaces

As mentioned in Sec. II, the true parameter estimates are obtained from the global minimum of $E_\infty(\lambda)$. However, its minimum is difficult to find. On the other hand, the error surface $E_0(\lambda)$ is easier to work with, but its global minimum yields an increasingly biased estimate of λ with decreasing SNR. We now illustrate the nature of $E_0(\lambda)$ and $E_\infty(\lambda)$ error surfaces with a one-path example to highlight the difficulties involved.

A synthetic received signal is considered. It was constructed by delaying a 244-Hz, 40-ms duration sinusoid by 50 ms and adding computer generated white Gaussian noise. The error surface (plotted as a function of *relative* delay) should have a minimum at $t = 0.05$ s. We consider two cases, viz., high and low SNRs. (We use the terms "high" and "low" SNR in a qualitative sense.) Figure 1 shows the high SNR signal. In Fig. 2 we show the corresponding error surfaces, $E_0(\lambda)$ and $E_\infty(\lambda)$. The smooth error surface is $E_0(\lambda)$, while the sinusoidal surface is $E_\infty(\lambda)$. Both have the same global minimum, but $E_0(\lambda)$ is easier to minimize, since it is unimodal. However, when the SNR is not high enough, the global minima of these two surfaces can be distinctly different. Figure 3 shows the low SNR received waveform, with the corresponding error surfaces in Fig. 4. In this example, the minimum of $E_0(\lambda)$ is approximately at $t = 0.049$ s, while that of $E_\infty(\lambda)$ is still at $t = 0.05$ s. It can be seen that if the time-delay obtained from the biased error surface is used, the true error can be quite large. This is more so in the M-path case. Hence, one has to minimize $E_\infty(\lambda)$ to obtain unbiased parameter estimates. This fact was first noted in Ref. 19.

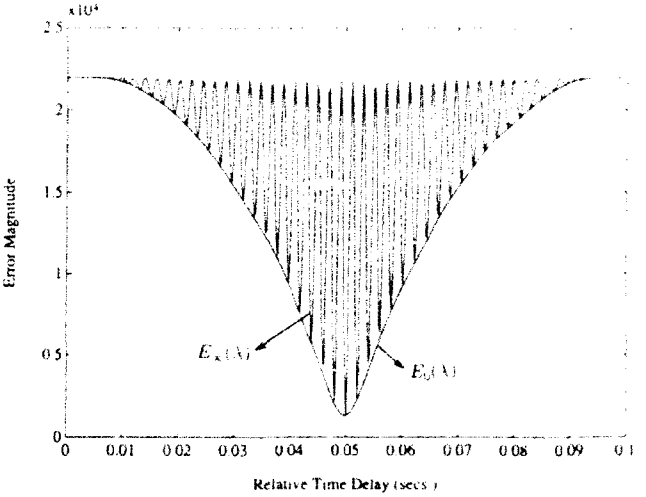

FIG. 2. Error surfaces $E_0(\lambda)$ and $E_\infty(\lambda)$ for the high-SNR received signal of Fig. 1.

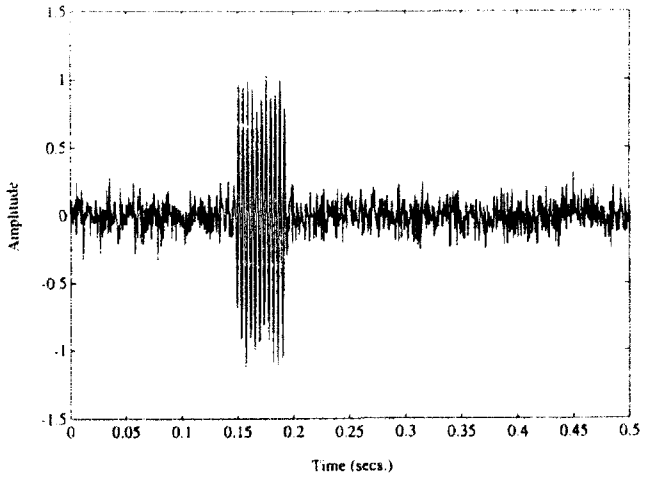

FIG. 1. Received signal at high SNR for a synthetic one-path example.

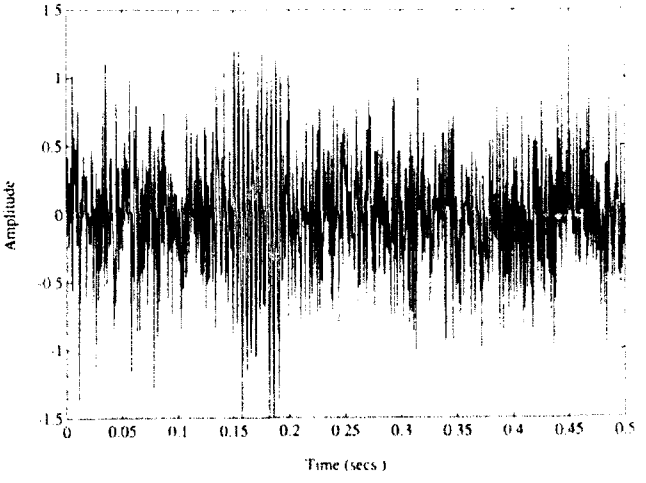

FIG. 3. Received signal at low SNR for a synthetic one-path example.

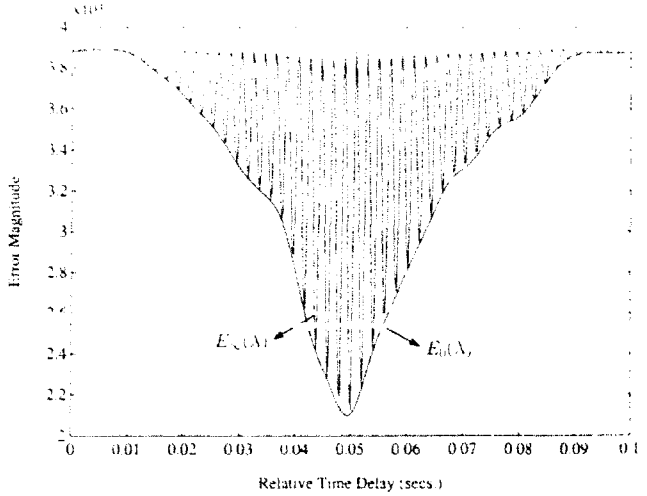

FIG. 4. Error surfaces $E_0(\lambda)$ and $E_\infty(\lambda)$ for the low-SNR received signal of Fig. 3.

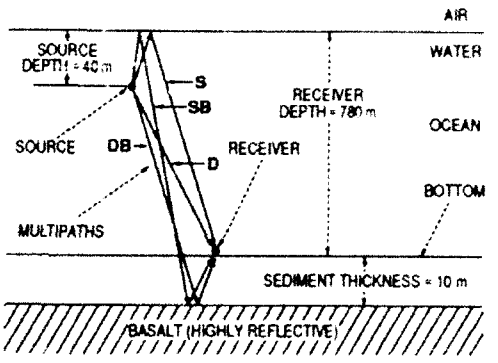

FIG. 6. The geometry of the channel used for the experiments.

B. The effect of increasing α

In order to obtain unbiased estimates, we need to go from the minimum of $E_0(\lambda)$ to the true minimum, i.e., that of $E_\infty(\lambda)$. This is achieved by increasing α. In Fig. 5, the smooth error surface corresponds to $\alpha = 0$, i.e., for which the amplitudes are complex. As α increases, the norm of $\mathbf{a}_i$ decreases and $E_\alpha(\lambda)$ is gradually transformed to the true error surface, as can be seen from Fig. 5.

C. Performance with experimental data

1. The experiment

Transient data were gathered in the Atlantic Ocean on a bottom-mounted receiver in 780 m of water. The experimental geometry is shown in Fig. 6.

The acoustic source was at a depth of 40 m and transmitted a 244-Hz gated sinusoid of 40-ms duration. The source signature was recorded using a hydrophone mounted on the source. The signature is shown in Fig. 7. The pulse was transmitted as the source ship drifted over the bottom receiver shown in Fig. 6. The horizontal range is estimated to be 100 m.

The ocean bottom is characterized by a thin sediment layer over a highly reflecting basalt as shown in Fig. 6. The sediment varies in thickness from 0 to 20 m. For this problem, a 10-m sediment thickness was chosen. This environment was modeled with a fast field program, SAFARI.[20,21] A broadband Gaussian pulse is transmitted in the model. The model predicts four paths shown in Fig. 6 (note that Fig. 6 is an artist's sketch of the four paths). Path D is the direct path, path DB is the direct path reflection off the basalt, path S is surface reflected, and path SB is the path reflected from the ocean surface and the basalt. The model ocean impulse response is shown in Fig. 8 with the four paths labeled. Pressure release surfaces such as the air/ocean interface cause a 180° phase shift in the reflected signal, causing the negative peaks seen in Fig. 8. The received signal is shown in Fig. 9.

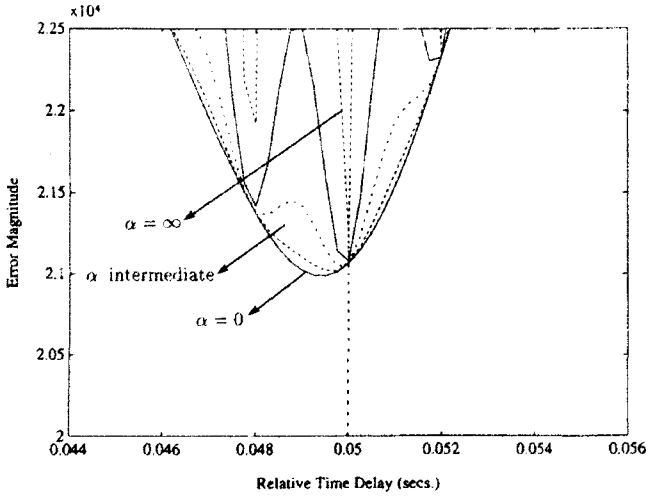

FIG. 5. Effect of increasing α on the constrained error surface $E_\alpha(\lambda)$.

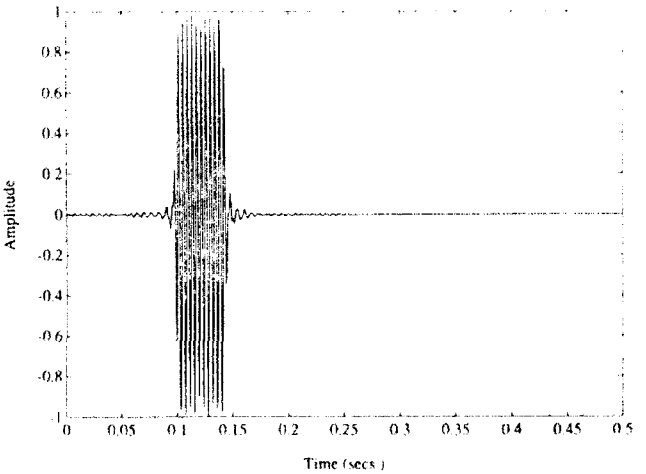

FIG. 7. The transmitted (source) signal: a 244-Hz gated sinusoid.

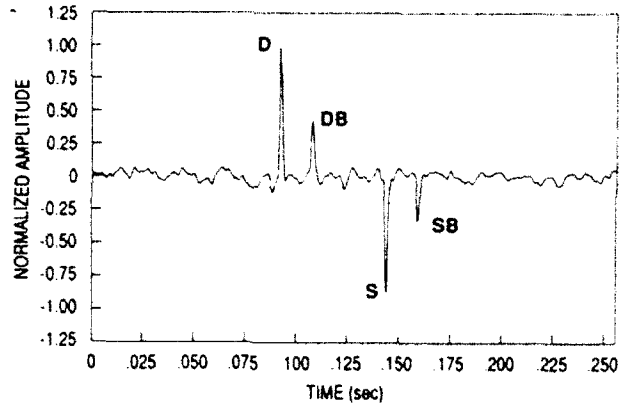

FIG. 8. Broadband Gaussian pulse propagated with SAFARI, source depth = 40 m, range = 0.1 km.

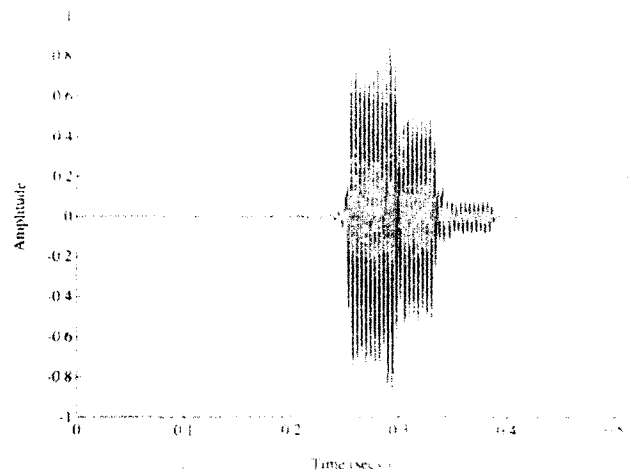

FIG. 10. The reconstructed received signal using channel parameters estimated by the known-signal algorithm.

2. The known signal algorithm

Using the known signal algorithm presented in Sec. II, the time delays and amplitudes were estimated. With these parameters, the received signal was reconstructed (see Fig. 10). The residual error between the received signal and the reconstructed signal is shown in Fig. 11. It is seen that the fit is very good, indicating the accuracy of the estimates. The estimated parameters are shown in Table I in the column labeled "K.S. estimate." Note that the signs of the estimated amplitudes need not necessarily agree with that predicted by the channel model (see Fig. 8), which shows that the first two amplitudes are positive while the second two are negative. The signs of the amplitudes were not constrained in this algorithm; however, such a constraint would be easy to impose using an additional penalty function.

3. The unknown signal algorithm

Next we applied the unknown signal algorithm outlined in the previous section and estimated the parameters of the transmitted signal, as well as the delays and amplitudes in the received signal. These estimates, along with those obtained from the known signal algorithm, are presented in Table I.

The estimated source parameters agree quite closely with the nominal values (244 Hz and 40 ms) of the actual source signal. The channel parameters, too, agree with those obtained by the known signal algorithm, although they are not in the same order. The first three paths identified by the known signal algorithm correspond to paths 1, 3, and 4 found by the unknown signal algorithm. There are many different parameter sets for this problem which all give reasonably good fits to the data. The efficacy of the unknown signal algorithm can be judged by reconstructing a signal using the estimated parameters, and computing the residual error between the received and the reconstructed signals. The reconstructed signal is formed by convolving the modeled source signal (a zero-phase, 245-Hz sinusoid of 43-ms duration) with the estimated channel. Convolution is accomplished by multiplication in the frequency domain fol-

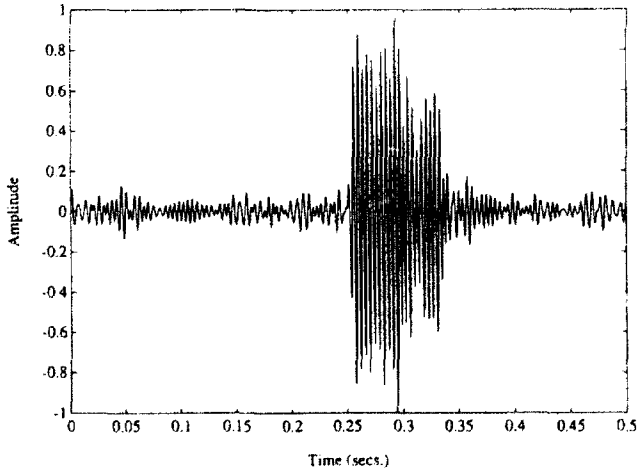

FIG. 9. A record of experimental data containing a received signal with four overlapping paths.

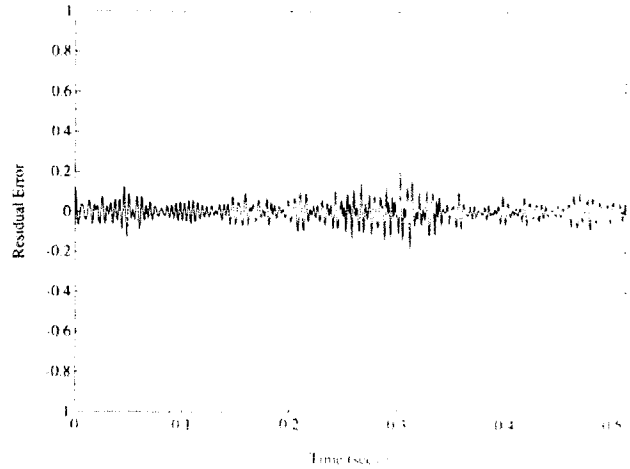

FIG. 11. The residual error (received signal minus reconstructed signal) for the known-signal algorithm.

TABLE I. Parameter estimates using the known signal and the unknown signal algorithms.

	K.S. estimate	U.S. estimate
f (Hz)		245
D (ms)		43
τ_1	0.1544[a]	0.2547[b]
$\tau_2 - \tau_1$	0.0360	0.0198
$\tau_3 - \tau_1$	0.0453	0.0360
$\tau_4 - \tau_1$	0.0898	0.0435
a_1	0.7913	0.8287
a_2	0.4925	−0.1597
a_3	−0.1724	0.4926
a_4	0.0864	0.1810

[a] Relative to the source pulse shown in Fig. 7.
[b] Relative to estimated source pulse starting at $t = 0$.

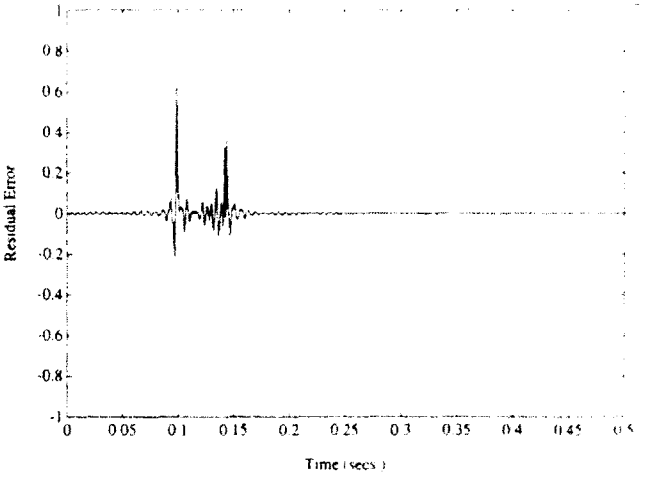

FIG. 13. The residual error between the source signal of Fig. 7. and the optimum zero-phase sinusoid parametric model.

lowed by an inverse DFT. The corresponding residual error is shown in Fig. 12. The fit in this case is poorer compared to the known signal algorithm. The reason for this is explained in the next section.

V. DISCUSSION

The biased error surface corresponding to the complex amplitudes is easier to minimize as it is reasonably smooth. For the one-path example of Sec. IV A, the surface was seen to be unimodal, i.e., has no local minima. In higher dimensions, the surface is still reasonably smooth as opposed to the $\alpha = \infty$ case, *but is no longer unimodal.* Hence, one has to begin any minimizing routine with a reasonable initial estimate for $\boldsymbol{\lambda}$ to reach the minimum of $E_0(\boldsymbol{\lambda})$, to avoid getting stuck in a local minimum. We have found that the coordinate descent algorithm provides good initial estimates. The error surface being smooth for $\alpha = 0$ is not true for an arbitrary signal. If, for example, the source signal consisted of multiple sinusoids, even for $\alpha = 0$ the error surface would possess numerous local minima. We are currently developing a technique which will overcome this problem by partitioning the frequency axis into bins. An algorithm similar to that given in this paper is used in each bin.

The rather large residual error obtained from the unknown signal algorithm is now explained. This poor fit is not due to any inadequacy in the unknown signal algorithm *per se*, but due to the fact that the actual source could not accurately be modeled as a perfect sinusoid. To demonstrate this, we ran the unknown signal algorithm by considering the recorded source to be the received signal and tried to model it by a sinusoid with $M = 1$. The estimated frequency and duration for the source were found to be $f = 243$ Hz and $d = 44$ ms, respectively. The residual error is shown in Fig. 13. There is considerable mismatch at the beginning and at the end of the source pulse due to gradual signal build up and decay. This, therefore, is the reason for the spikes in the residual error of Fig. 12, wherein the unknown signal algorithm assumed a perfectly sinusoidal source.

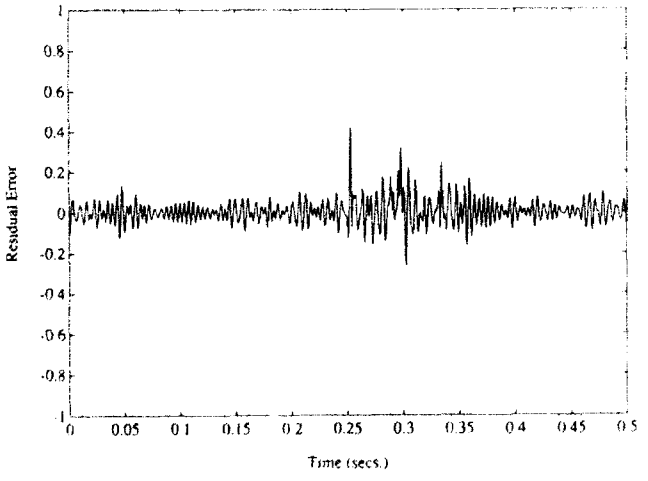

FIG. 12. The residual error (received signal minus reconstructed signal) for the unknown-signal algorithm.

VI. CONCLUSIONS AND FURTHER WORK

The residual error for the experimental data of Sec. IV C using the known signal algorithm was quite small. We tried the same procedure on a different experimental data set which used a chirp signal as the source pulse. The bottom reflecting surface was very rough in this case. The known signal algorithm was unable to improve the fitting error beyond a certain point. There still appeared to be considerable signal structure in the error residual, indicating scope for a better fit. We feel that this might be due to inadequacy of the channel model given in Eq. (1); further work is being carried out to verify this.

To summarize, in this paper we pointed out the importance of constraining the amplitudes to be real. But the error minimization in this case turned out to be difficult. The complex amplitude case is easier to solve but yields biased pa-

rameter estimates. We then presented an algorithm which finds the minimum of the true error surface starting with the biased error surface. When initial parameter estimates are obtained using coordinate descent, our algorithm finds the global minimum of the error surface in spite of the existence of numerous local minima. The source was assumed to be known, initially. Finally, even if the source were unknown, it was shown that it can be estimated if it belongs to a parametric class of signals.

ACKNOWLEDGMENTS

This work was supported by the ONR funded Acoustic Transients ARI, program element No. 0601153N, under NRL–SSC program management (Dr. Edward Franchi).

[1] R. Field, "Transient signal distortion in a multipath environment," in *Proceedings of Oceans '90* (IEEE, Piscataway, NJ, 1990), pp. 111–114.
[2] G. Carter, "Time delay estimation for passive sonar signal processing," IEEE Trans. Acoust. Speech Signal Process. **ASSP-29**, 463–470 (June 1981).
[3] J. E. Ehrenberg, T. E. Ewart, and R. D. Morris, "Signal Processing Techniques for Resolving Individual Pulses in a Multipath Signal," J. Acoust. Soc. Am. **63**, 1861–1865 (1978).
[4] S. Senamato and D. G. Childers, "Signal resolution via digital inverse filtering," IEEE Trans. Aerosp. Electron. Syst. **AES-8**, 633–640 (1972).
[5] B. M. Bell and T. E. Ewart, "Separating multipaths by global optimization of a multidimensional matched filter," IEEE Trans. Acoust. Speech Signal Process. **ASSP-34**, 1029–1037 (Oct. 1986).
[6] J. O. Smith and B. Friedlander, "Adaptive multipath delay estimation," IEEE Trans. Acoustics Speech Signal Process. **ASSP-33**, 812–822 (August 1985).
[7] R. Tremblay, G. Carter, and D. Lytle, "A practical approach to the estimation of amplitude and time-delay parameters of a composite signal," IEEE J. Oceanic Eng. **OE-12**, 273–278 (1987).
[8] I. P. Kirsteins, "High Resolution Time Delay Estimation," in *IEEE Proceedings ICASSP 87*, Dallas, TX (IEEE, New York, 1987), pp. 451–454.
[9] I. P. Kirsteins and A. H. Quazi, "Exact Maximum Likelihood Time Delay Estimation for Deterministic Signals," in *Proceedings of EUSIPCO-88*, Grenoble, France, Sept. 1988 (Elsevier, Amsterdam, 1988).
[10] A. G. Evans and R. Fischl, "Optimal least squares time-domain synthesis of recursive digital filters," IEEE Trans. Audio Electroacoust. **AU-21**, 61–65 (February 1973).
[11] R. Kumaresan, L. Scharf, and A. Shaw, "An algorithm for pole-zero modeling and spectral analysis," IEEE Trans. Acoust. Speech Signal Process. **ASSP-34**, 637–640 (June 1986).
[12] C. W. Helstrom, *Statistical Theory of Detection* (Pergamon, New York, 1960).
[13] P. Gill, W. Murray, and M. Wright, *Practical Optimization* (Academic, New York, 1981).
[14] J. Dennis and R. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations* (Prentice-Hall, Englewood Cliffs, NJ, 1983).
[15] E. Maragakis, "Time Delay Estimation in a Multipath Environment," Master's thesis, University of Rhode Island, 1990.
[16] G. Golub and C. V. Loan, *Matrix Computations* (Johns Hopkins U.P., Baltimore, MD, 1984).
[17] D. Luenberger, *Introduction to Linear and Nonlinear Programming* Reading, (Addison-Wesley, Reading, MA, 1973).
[18] J. Cadzow, "Signal processing via least squares error modeling," IEEE Signal Process. Mag. **7**, 12–31 (October 1990).
[19] R. Vaccaro, E. Maragakis, and R. Field, "Transient Signal Extraction in a Multipath Environment," in Proc. Oceans '90, Washington, DC, pp. 115–118 (September 1990).
[20] H. Schmidt, "SAFARI (Seismo-Acoustic Fast Field Algorithm for Range Independent Environment) User's Guide," Technical report, SACLANT Undersea Research Centre, Rep. SR-113 (1988).
[21] J. Leclere and R. Field, "Comparison of time-domain parabolic equation and measured ocean impulse responses," Proc. Oceans '90, pp. 125–128 (September 1990).