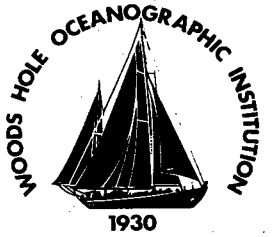
WHOI-87-3
Copy 1

Woods Hole Oceanographic Institution



Multipath Time-Delay Estimation via the EM Algorithm

by

Meir Feder and Ehud Weinstein

January 1987

Technical Report

Funding was provided by the Naval Underwater Systems Center under contract No. N00014-80-C-0381.

WHOI-87-3

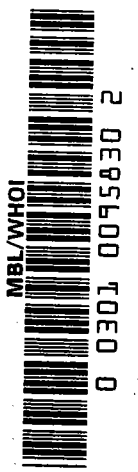
Multipath Time-Delay Estimation via the EM Algorithm

by

Meir Feder and Ehud Weinstein

Woods Hole Oceanographic Institution
Woods Hole, Massachusetts 02543

January 1987



Technical Report

Funding was provided by the Naval Underwater Systems Center under contract No. N00014-80-C-0381.

 This report should be cited as: Woods Hole Oceanog. Inst. Tech. Rept., WHOI-87-3.

Approved for publication; distribution unlimited.

Approved for Distribution:

Robert C. Spindel, Chairman
Department of Ocean Engineering

Administrative Information

This is the first of two reports to be submitted under contract number N00014-80-C-0381.

Abstract

We consider the application of the EM algorithm to the multipath time delay estimation problem. The algorithm is developed for the case of deterministic (known) signals, as well as for the case of wide-sense stationary Gaussian signals.

1 Introduction

Sounds radiated from an acoustic point source in the ocean arrive at the receiver over more than one path. Given prior knowledge of the geophysical parameters characterizing the channel between source and receiver, the observed multipath delay information can be used to estimate the location (i.e., range and depth) of the source relative to the receiver. In this report we consider the application of the EM algorithm to the multipath time delay estimation.

2 Maximum Likelihood Estimation and the EM Algorithm

Let $\underline{Y}$ denote the data vector possessing the probability density $f_{\underline{Y}}(\underline{y};\underline{\theta})$ indexed by the parameter vector $\underline{\theta} \in \Theta$. Θ is a subset of the Euclidean K-space. Given an observed $\underline{y}$, the ML estimate $\hat{\underline{\theta}}_{ML}$ is the value of $\underline{\theta}$ that maximizes the log-likelihood, that is

$$\max_{\underline{\theta}\in\Theta} \log f_{\underline{Y}}(\underline{y};\underline{\theta}) \Longrightarrow \hat{\underline{\theta}}_{ML} \tag{1}$$

Suppose the data vector $\underline{Y}$ can be viewed as being incomplete, and we can specify some "complete" data $\underline{X}$ related to $\underline{Y}$ by

$$H(\underline{X}) = \underline{Y} \tag{2}$$

where $H(\cdot)$ is a non-invertible (many to one) transformation. In the multi-path time delay estimation problem, the "complete" data could be the observation of the various paths *separately*, where the observed (incomplete) data is the *sum* of the signal contributions from the various paths.

The EM algorithm is directed at finding the solution to (1); however it does so by making an essential use of the complete data specification. The algorithm is basically an iterative method. It starts with an initial guess $\underline{\theta}^{(0)}$, and let $\underline{\theta}^{(n+1)}$ be defined inductively by

$$\max_{\underline{\theta}\in\Theta} E\left\{\log f_{\underline{X}}(\underline{x};\underline{\theta})/\underline{Y}=\underline{y};\underline{\theta}^{(n)}\right\} \Longrightarrow \underline{\theta}^{(n+1)} \tag{3}$$

where $f_{\underline{X}}(\underline{x};\underline{\theta})$ is the probability density of $\underline{X}$, and $E\left\{\cdot/\underline{Y}=\underline{y};\underline{\theta}^{(n)}\right\}$ denotes the conditional expectation given $\underline{y}$, computed using the parameter value $\underline{\theta}^{(n)}$. The heuristic idea here is that we would like to choose $\underline{\theta}$ that maximizes $\log f_{\underline{X}}(\underline{x};\underline{\theta})$, the log-likelihood of the complete data. However, since $\log f_{\underline{X}}(\underline{x};\underline{\theta})$ is not available to us (because the complete data is not available), we maximize instead its expectation, given the observed data $\underline{y}$. Since we used the current estimate $\underline{\theta}^{(n)}$ rather then the actual value of $\underline{\theta}$ which is unknown, the conditional expectation is not exact. Thus the algorithm iterates, using each new parameter estimate to improve the conditional expectation on the next iteration cycle and thus to improve the next parameter estimate. In Appendix A we give more comprehensive description of the EM algorithm.

The EM algorithm was first presented by Dempster et al. in [1]. The algorithm was suggested before, however not in its general form, by several authors e.g. [2], [3], [4]. Instances of the EM algorithm has been suggested for some signal processing problems e.g [5] [6]. The most important work has been done by Musicus [7],[8], who independently suggested a general class of iterative algorithms, some of them coincide with the EM algorithm, and applied them to several signal processing problems.

In [1] it is shown that each iteration increases the likelihood, however there is an error in the convergence proof (theorem 2 of [1]), pointed out Wu, [9]. The proper conditions that guarantee the convergence of the algorithm to a stationary point of the likelihood are

given in [9].

The rate of convergence of the algorithm is exponential, depending on the fraction of the covariance of the "complete" data that can be predicted using the observed data (theorem 4 of [1]). If that fraction is small, the rate of convergence tends to be slow, in which case one could use standard numerical methods to accelerate the algorithm.

We note that the EM algorithm is not uniquely defined. The transformation $H(\cdot)$ relating $\underline{X}$ to $\underline{Y}$ can be ***any*** non-invertible transformation. Obviously, there are many possible "complete" data specifications that will generate the observed data. thus the EM algorithm can be implemented in many possible ways. The way $H(\cdot)$ is specified (i.e the choice of the "complete" data) may critically affect the complexity and the rate of convergence of the algorithm. It may also affect the convergence point, leading to a different stationary point for different choices of "complete data". Thus, the choice of "complete" data is an important factor in designing an EM algorithm for a problem; an unfortunate choice of $H(\cdot)$ may yield a completely useless algorithm.

We shall proceed as follows: First we develop the EM algorithm for the Linear-Gaussian case. This case covers a wide range of applications. Then we show that for the problem of interest here, there is a natural choice of the "complete" data, leading to a surprisingly simple algorithm to extract the ML estimates. We will conclude by presenting preliminary simulation results for the deterministic and the stochastic cases.

3 The Linear Gaussian Case

Suppose that $\underline{Y} = H\underline{X}$, where H is a $m \times n$ matrix ($m < n$), and $\underline{X}$ possesses the following multivariate Gaussian probability density:

$$f_{\underline{X}}(\underline{x};\underline{\theta}) = \left[\det(\frac{2\pi}{\lambda}\Lambda(\underline{\theta}))\right]^{-\lambda/2} \exp\left[-\frac{\lambda}{2}(\underline{x}-\underline{m}(\underline{\theta}))^{\dagger}\Lambda^{-1}(\underline{\theta})(\underline{x}-\underline{m}(\underline{\theta}))\right] \tag{4}$$

where $\lambda = 1$ if $\underline{X}$ is real valued, $\lambda = 2$ if $\underline{X}$ is complexed valued, and $\dagger$ denotes the conjugate transpose operation. We shall refer to this case as the Linear Gaussian case. Taking the logarithm of (4), we get

$$\begin{aligned}\log f_{\underline{X}}(\underline{x};\underline{\theta}) &= -\frac{\lambda}{2}\log\det(\frac{2\pi}{\lambda}\Lambda(\underline{\theta})) - \frac{\lambda}{2}(\underline{x}-\underline{m}(\underline{\theta}))^{\dagger}\Lambda^{-1}(\underline{\theta})(\underline{x}-\underline{m}(\underline{\theta})) \\ &= -\frac{\lambda}{2}\log\det(\frac{2\pi}{\lambda}\Lambda(\underline{\theta})) - \frac{\lambda}{2}\underline{m}^{\dagger}(\underline{\theta})\Lambda^{-1}(\underline{\theta})\underline{m}(\underline{\theta}) \\ &+ \frac{\lambda}{2}\underline{x}^{\dagger}\Lambda^{-1}(\underline{\theta})\underline{m}(\underline{\theta}) + \frac{\lambda}{2}\underline{m}^{\dagger}(\underline{\theta})\Lambda^{-1}(\underline{\theta})\underline{x} - \frac{\lambda}{2}tr(\Lambda^{-1}(\underline{\theta})\underline{x}\underline{x}^{\dagger})\end{aligned} \tag{5}$$

where $tr(\cdot)$ stands for the trace of the bracketed matrix. Thus

$$\begin{aligned}E\left\{\log f_{\underline{X}}(\underline{x};\underline{\theta})/\underline{y};\underline{\theta}^{(n)}\right\} &= -\frac{\lambda}{2}\log\det(\frac{2\pi}{\lambda}\Lambda(\underline{\theta})) - \frac{\lambda}{2}\underline{m}^{\dagger}(\underline{\theta})\Lambda^{-1}(\underline{\theta})\underline{m}(\underline{\theta}) \\ &+ \frac{\lambda}{2}(\underline{x}^{(n)})^{\dagger}\Lambda^{-1}(\underline{\theta})\underline{m}(\underline{\theta}) + \frac{\lambda}{2}\underline{m}^{\dagger}(\underline{\theta})\Lambda^{-1}(\underline{\theta})\underline{x}^{(n)} - \frac{\lambda}{2}tr(\Lambda^{-1}(\underline{\theta})\Psi^{(n)})\end{aligned} \tag{6}$$

where $\underline{x}^{(n)} = E\left\{\underline{x}/\underline{Y} = \underline{y};\underline{\theta}^{(n)}\right\}$ and $\Psi^{(n)} = E\left\{\underline{x}\underline{x}^{\dagger}/\underline{Y} = \underline{y};\underline{\theta}^{(n)}\right\}$.

Since $\underline{X}$ and $\underline{Y}$ are jointly Gaussian, these conditional expectations are readily available in the literature (e.g. [10], chap. 5)

$$\underline{x}^{(n)} = \underline{m}(\underline{\theta}^{(n)}) + \Gamma(\underline{\theta}^{(n)})\left[\underline{y} - H\cdot\underline{m}(\underline{\theta}^{(n)})\right] \tag{7}$$

$$\Psi^{(n)} = \left[I - \Gamma(\underline{\theta}^{(n)})\cdot H\right]\Lambda(\underline{\theta}^{(n)}) + (\underline{x}^{(n)})(\underline{x}^{(n)})^{\dagger} \tag{8}$$

where I is the identity matrix and $\Gamma(\underline{\theta})$ is the "Kalman gain" defined by

$$\Gamma(\underline{\theta}) = \Lambda(\underline{\theta})H^{\dagger}\left[H\Lambda(\underline{\theta})H^{\dagger}\right]^{-1} \tag{9}$$

The EM algorithm is completely specified by eqs. (6) - (9); the algorithm iterates between calculating $\underline{x}^{(n)}$ and $\Psi^{(n)}$ and maximizing the expression in (6) with respect to $\underline{\theta}$. Each iteration increases the likelihood. We observe that $E\left\{\log f_{\underline{X}}(\underline{x};\underline{\theta})/\underline{y};\underline{\theta}^{(n)}\right\}$ (Eq. 6) and $\log f_{\underline{X}}(\underline{x};\underline{\theta})$ (Eq. 5), have the same dependence on $\underline{\theta}$. Maximizing (6) with respect to $\underline{\theta}$ is the same as maximizing (5) with respect to $\underline{\theta}$. Hence, the EM algorithm essentially requires the ML solution in the $\underline{X}$ model which might be significantly simpler than the direct ML solution in the $\underline{Y}$ model.

4 Application to Multipath Time-Delay Estimation

Let the mathematical model characterizing the observed signal $y(t)$ be given by

$$y(t) = \sum_{k=1}^{N} \alpha_k s(t - \tau_k) + n(t) \quad T_i \leq t \leq T_f \tag{10}$$

where $s(t)$ is the source signal and $n(t)$ is the additive noise. Assuming that N, the number of paths, is known a-priori, the vector unknown parameters to be estimated is

$$\underline{\theta} = (\tau_1,\ \tau_2, \cdots, \tau_N,\ \alpha_1,\ \alpha_2, \cdots, \alpha_N)^T \tag{11}$$

We shall now consider the estimation of the components of $\underline{\theta}$ for the case of deterministic signals and for the case of stochastic Gaussian signals separately.

4.1 Deterministic Signals

Consider the signal model of (10) under the following assumptions:

1) $s(t)$ is a deterministic (known) signal. We can assume, without any loss of generality, that $\int_{T_i}^{T_f} s^2(t)dt = 1$.

2) $n(t)$ is a sample function from a zero-mean white Gaussian process with (double-sided) spectral level of N_0 watts/Herz.

Under these assumptions, the log-likelihood function is given by, ([11] Chapter 2)

$$\log f_{\mathbf{Y}}(y;\underline{\theta}) = C - \frac{1}{2N_0}\int_{T_i}^{T_f}\left[y(t) - \sum_{k=1}^{N}\alpha_k s(t-\tau_k)\right]^2 dt \tag{12}$$

where C is a constant independent of the unknown parameters. To obtain the ML estimate of all the τ_k's and α_k's we must therefore solve the following optimization problem:

$$\min_{\substack{\tau_1,\tau_2,\cdots,\tau_N \\ \alpha_1,\alpha_2,\cdots,\alpha_N}} \int_{T_i}^{T_f}\left[y(t) - \sum_{k=1}^{N}\alpha_k s(t-\tau_k)\right]^2 dt \tag{13}$$

This is a complicated optimization problem in 2N unknowns. Of course, brute force can always be used to solve the problem, evaluating the objective function on a coarse grid to roughly locate the global minimum, and then applying the Gauss method or the Newton-Raphson or some other iterative gradient-search algorithm. However, when applied to the problem at hand, these methods tend to be very complex and computationally time consuming.

Having the EM algorithm in mind, we would like to simplify the optimization problem associated with the direct ML approach. To apply the algorithm to the problem at hand, the first step is to specify the "complete" data. A natural choice of "complete" data is obtained by decomposing $y(t)$ into

$$y(t) = \sum_{k=1}^{N} x_k(t) \tag{14}$$

where

$$x_k(t) = \alpha_k s(t-\tau_k) + n_k(t) \tag{15}$$

and the $n_k(t)$ are chosen to be mutually uncorrelated zero-mean Gaussian processes satisfying

$$\sum_{k=1}^{N} n_k(t) = n(t) \tag{16}$$

The "complete" data is

$$\underline{x}(t) = \Big(x_1(t),\ x_2(t), \cdots, x_N(t)\Big)^T \tag{17}$$

The relationship between the "complete" data and the observed ("incomplete") data is

$$y(t) = \mathbf{1}^T \underline{x}(t) \tag{18}$$

where $\mathbf{1}^T = (1,\ 1, \cdots, 1)$.

Since the "complete" data is Gaussian, and the relationship between $\underline{x}(t)$ and $y(t)$ is linear, the results developed for the Linear-Gaussian case can be applied here. The detailed derivation is given in Appendix B. The resulting algorithm is:

E-step For $k = 1, 2, \ldots N$, compute:

$$x_k^{(n)}(t) = \alpha_k^{(n)} s\left(t - \tau_k^{(n)}\right) + \beta_k \left[y(t) - \sum_{k=1}^{N} \alpha_k^{(n)} s\left(t - \tau_k^{(n)}\right)\right] \tag{19}$$

M-step for $k = 1, 2, \ldots N$

$$\min_{\tau_k, \alpha_k} \int_{T_i}^{T_f} \left[x_k^{(n)}(t) - \alpha_k s(t - \tau_k)\right]^2 dt \Longrightarrow \tau_k^{(n+1)}, \alpha_k^{(n+1)} \tag{20}$$

Perhaps the most striking feature of the algorithm is that it decouples the complicated multi-parameter optimization into N separate optimization. Each optimization is, in fact, the optimization associated with the ML estimation of the pair (τ_k, α_k) given separate observations of the k^{th} signal path. Thus, the complexity of the algorithm is unaffected by the assumed number of signal paths. As N increases, we have to increase the number of ML processors in parallel; however, each processor is maximized separately.

Since the algorithm is based on the EM method, it must converge to a stationary point of the log-likelihood function, where each iteration increases the likelihood.

We note that the β_k's must satisfy the constraint

$$\sum_{k=1}^{N} \beta_k = 1 \quad , \tag{21}$$

but otherwise they are arbitrary free variables in the algorithm. The choice of β_k's may be used to control the rate of convergence of the algorithm.

If the observation time $T = T_f - T_i$ is long compared with the signal duration and with the maximum expected delay, then

$$\int_{T_i}^{T_f} s^2(t - \tau_k)dt \approx \int_{T_i}^{T_f} s^2(t)dt = 1 \tag{22}$$

In this case the optimization required in the M-step (Eq. (20)) reduces to:

M-step

$$\max_{\tau_k} \int_{T_i}^{T_f} x_k^{(n)}(t)s(t - \tau_k)dt \Longrightarrow \tau_k^{(n+1)} \tag{23}$$

$$\alpha_k^{(n+1)} = \int_{T_i}^{T_f} x_k^{(n)}(t)s(t - \tau_k^{(n+1)})dt \tag{24}$$

The term to be maximized in (23) can be generated by passing $x_k^{(n)}(t)$ through a filter matched to $s(t)$. Thus, under the assumption of (22), the M-step of the algorithm consists of maximizing the outputs of N match filters in parallel.

4.2 Gaussian Signals

Consider the signal model of (10) under the following assumptions:

1) $s(t)$ is a sample function from a wide sense stationary zero mean Gaussian process with spectral density $S(\omega)$.

2) $n(t)$ is a sample function from a wide sense stationary zero mean Gaussian process with spectral density $N(\omega)$.

3) $s(t)$ and $n(t)$ are statistically independent.

4) The observation time T is long compared with the correlation time (reciprocal bandwidth) of the signal and the noise, i.e., $WT/2\pi \gg 1$.

Under these assumptions, the log-likelihood is given by

$$\log f_Y(y;\underline{\theta}) = C - \sum_{\omega_\ell} \left\{ \frac{|\, Y(\omega_\ell)\,|^2}{|\, \sum_{k=1}^{N} \alpha_k e^{-j\omega_\ell \tau_k}\,|^2\, S(\omega_\ell) + N(\omega_\ell)} + \log\left[|\, \sum_{k=1}^{N} \alpha_k e^{-j\omega_\ell \tau_k}\,|^2\, S(\omega_\ell) + N(\omega_\ell) \right] \right\} \tag{25}$$

where C is a normalizing constant, and

$$Y(\omega_\ell) = \frac{1}{\sqrt{T}} \int_{T_i}^{T_f} y(t) e^{-j\omega_\ell t} dt \qquad \omega_\ell = \frac{2\pi}{T} \cdot \ell \tag{26}$$

The summation in (25) is carried over all ω_ℓ's in the signal frequency band. We note that the log-likelihood function depends only on the differences $(\tau_k - \tau_m)$ indicating that we can only estimate relative delays. We therefore assume, without any loss of generality, that $\tau_N = 0$. In that case, τ_k, $k = 0, 1, 2, \cdots, N-1$ are the delay parameters measured relative to the N^{th} signal path.

To obtain the ML estimate of the various unknown parameters, we must solve the following optimization problem:

$$\min_{\substack{\tau_1,\tau_2,\cdots,\tau_{N-1} \\ \alpha_1,\alpha_2,\cdots,\alpha_N}} \sum_{\omega_\ell} \left\{ \frac{|\, Y(\omega_\ell)\,|^2}{|\, \sum_{k=1}^{N} \alpha_k e^{-j\omega_\ell \tau_k}\,|^2\, S(\omega_\ell) + N(\omega_\ell)} + \log\left[|\, \sum_{k=1}^{N} \alpha_k e^{-j\omega_\ell \tau_k}\,|^2\, S(\omega_\ell) + N(\omega_\ell) \right] \right\} \tag{27}$$

This a complicated optimization problem. Using the EM method under the same "complete" data specifications as in the deterministic signal case (Eqs. (14)-(18)), we propose the following iterative algorithm to solve the optimization problem in (27):

E-step Compute:

$$\Psi^{(n)}(\omega_\ell) = \Lambda(\omega_\ell;\underline{\theta}^{(n)}) + \frac{\Lambda(\omega_\ell;\underline{\theta}^{(n)})\mathbf{1}\mathbf{1}^T\Lambda(\omega_\ell;\underline{\theta}^{(n)})}{\left(\mathbf{1}^T\Lambda(\omega_\ell;\underline{\theta}^{(n)})\mathbf{1}\right)^2}\left[\mid Y(\omega_\ell)\mid^2 - \mathbf{1}^T\Lambda(\omega_\ell;\underline{\theta}^{(n)})\mathbf{1}\right] \tag{28}$$

M-step

$$\max_{\substack{\tau_1,\tau_2,\cdots,\tau_{N-1}\\ \alpha_1,\alpha_2,\cdots,\alpha_N}} \sum_{\omega_\ell}\left[G(\omega_\ell)\underline{V}^\dagger(\omega_\ell)Q^{-1}\Psi^{(n)}(\omega_\ell)Q^{-1}\underline{V}(\omega_\ell)+\log G(\omega_\ell)\right] \Longrightarrow \begin{matrix}\tau_1^{(n+1)},\tau_2^{(n+1)},\cdots,\tau_{N-1}^{(n+1)}\\ \alpha_1^{(n+1)},\alpha_2^{(n+1)},\cdots,\alpha_N^{(n+1)}\end{matrix} \tag{29}$$

where

$$\Lambda(\omega;\underline{\theta}) = S(\omega)\underline{V}(\omega)\underline{V}^\dagger(\omega) + N(\omega)\cdot Q \tag{30}$$

$$\underline{V}^\dagger(\omega) = (\alpha_1 e^{\jmath\omega\tau_1}\alpha_2 e^{\jmath\omega\tau_2},\cdots,\alpha_N e^{\jmath\omega\tau_N}) \tag{31}$$

$$Q = diag(\beta_1,\beta_2,\cdots,\beta_N) \tag{32}$$

and

$$G(\omega) = \frac{S(\omega)/N^2(\omega)}{1+\frac{S(\omega)}{N(\omega)}\sum_{k=1}^N\frac{\alpha_k^2}{\beta_k}} \tag{33}$$

Detailed derivation of the algorithm is given in Appendix C.

5 Preliminary simulation results

The algorithms specified above have been tested using a simple simulation of the physical situation. We have concentrated more on testing the deterministic signal case. The ideas presented in this report has been also applied to solve a multi-target direction of arrival estimate, and the algorithm with some simulation results may be found in [12],[13].

5.1 Deterministic signals, example 1

Consider the following situation:

- The received signal is discrete and consists of 200 time samples.
- There are 5 paths. The relative time delays are 91, 95, 100, 104, and 109 samples.
- The signal is a rectangular pulse whose duration is 20 samples, i.e

$$s(t) = \begin{cases} 1. & \text{for } 0 \leq t < 20 \\ 0. & \text{otherwise} \end{cases}$$

- The amplitude of the paths are 1., 2., 1., 0.8, and 1.7 respectively.
- The noise is white and Gaussian with variance 0.1, (that is the SNR per path is between 9 to 13 db).

In Fig. 1(a) the received signal is plotted. In Fig. 1(b) the output of a matched filter (to the rectangular pulse) is plotted. We notice that the match filter essentially detects only one path.

The algorithm specified in Eqs. (19),(22),(23) has been applied, using a uniform choice for the β_k's. The result, using a random starting point, (delays 75,91,107,119,126, amplitudes uniformly 1), is given in table 1. To illustrate the performance of the algorithm, we have plotted in Fig. 2, the output of the M step matched filter for each path, after 1, 5, 10 and 30 iterations.

5.2 Deterministic signals, example 2

In this example we tried to check the robustness of the algorithm to the knowledge of the exact shape of transmitted pulse. In this example

- There are 3 paths, the time delays are 95, 100, 105.

- The signal is "raised sine" pulse i.e

$$s(t) = \begin{cases} \sin(\frac{\pi t}{20}) & \text{for } 0 \leq t < 20 \\ 0. & \text{otherwise} \end{cases}$$

- The amplitudes are 1., 2., and 1. respectively.

- The noise variance is 0.1

The received signal is plotted in Fig. 3(a). A matched filter output (using the true, raised sine, signal) is plotted in Fig. 3(b).

We have applied the algorithm above, however, in order to check the robustness of the algorithm to the precise knowledge of the waveform, we assumed that the transmitted pulse is triangular i.e

$$s_k(t) = \begin{cases} \frac{1}{10}t & \text{for } 0 \leq t \leq 10 \\ \frac{1}{10}(20 - t) & \text{for } 10 < t \leq 20 \end{cases}$$

The results, starting with delays 75., 119. and 126 and unit amplitudes are given in table 2. The output of the M step matched filter after 1, 5, 10, and 30 iterations is plotted in Fig. 4.

5.3 Stochastic signals

Consider the following situation:

- $s(t)$ is a sample function of a discrete, stationary Gaussian process whose power spectrum is uniformly 1, between $-1/4 \leq \omega \leq 1/4$, and is zero elsewhere.

- A record of 4096 points from this process was generated, i.e the number of degrees of freedom $(2WT)$, is 2048.

- The observed signal contains the direct path and an additional path, whose relative delay is 4 samples and its amplitude is 0.7.

- The noise is white and Gaussian, with spectral level 0.5, i.e the signal to noise ratio is 0 db.

In Fig. 5, the periodogram (in logarithmic scale) of the received signal and its autocorrelation are plotted. The autocorrelation function is also zoomed, to show the interesting (near the right delay) region.

The algorithm described in Eqs. (28),(29) has been applied, using $\beta_1 = \beta_2 = 1/2$. The initial guess was 2.5 for the delay, and 0.8 for the amplitude. The results, i.e the estimate of the relative delay and amplitude of the echo, is given in table 3. In Fig. 6 the output of the M-step, (that is a cross-correlation between the direct path and the echo contributions) is plotted, zoomed around the interesting region, after 1, 2, 5, 10, 20 and 30 iterations.

A The EM Algorithm

Let $H(\underline{X}) = \underline{Y}$ where $H(\cdot)$ is a non-invertible (many-to-one) transformation. Express densities

$$f_{\underline{X}}(\underline{x};\underline{\theta}) = f_{\underline{Y}}\left(H(\underline{x});\underline{\theta}\right) \cdot f_{\underline{X}/\underline{Y}=\underline{y}}(\underline{x};\underline{\theta}) \tag{A.1}$$

Taking the logarithm on both sides of (A.1), we obtain

$$\log f_{\underline{Y}}\left(H(\underline{x});\underline{\theta}\right) = \log f_{\underline{X}}(\underline{x};\underline{\theta}) - \log f_{\underline{X}/\underline{Y}=\underline{y}}(\underline{x};\underline{\theta}) \tag{A.2}$$

Denote $\underline{y} = H(\underline{x})$ and taking conditional expectations given $\underline{Y} = \underline{y}$ for a parameter value $\underline{\theta}'$, we obtain

$$\log f_{\underline{Y}}(\underline{y};\underline{\theta}) = E\left\{\log f_{\underline{X}}(\underline{x};\underline{\theta})/\underline{Y} = \underline{y};\underline{\theta}'\right\} -$$

$$-E\left\{\log f_{\underline{X}/\underline{Y}=\underline{y}}(\underline{X};\underline{\theta})/\underline{Y} = \underline{y};\underline{\theta}'\right\} = Q(\underline{\theta},\underline{\theta}') - H(\underline{\theta},\underline{\theta}') \tag{A.3}$$

Invoking the Jensen's inequality, we have that $H(\underline{\theta},\underline{\theta}') \leq H(\underline{\theta}',\underline{\theta}')$. Hence

$$Q(\underline{\theta},\underline{\theta}') > Q(\underline{\theta}',\underline{\theta}') \Longrightarrow \log f_{\underline{Y}}(\underline{y};\underline{\theta}) > \log f_{\underline{Y}}(\underline{y};\underline{\theta}') \tag{A.4}$$

Eq. (A.4) forms the basis to the EM algorithm. The algorithm starts with an initial guess $\underline{\theta}^{(0)}$, and let $\underline{\theta}^{(n+1)}$ be defined inductively by

$$\max_{\underline{\theta}} Q(\underline{\theta},\underline{\theta}^{(n)}) \Longrightarrow \underline{\theta}^{(n+1)} \tag{A.5}$$

We note that since $\underline{\theta}^{(n+1)}$ is the value of $\underline{\theta}$ that maximizes $Q(\underline{\theta},\underline{\theta}^{(n+1)})$, then according to (A.4), each iteration of the algorithm increases the value of $\log f_{\underline{Y}}(\underline{y};\underline{\theta})$.

B Derivation of the EM Algorithm for Deterministic Signals

Since the various $x_k(t)$'s are statistically independent Gaussian random processes, it can easily be verified that the log-likelihood of the "complete" data is

$$\log f_{\underline{X}}(\underline{x};\underline{\theta}) = \sum_{k=1}^{N} \log f_{X_k}(x_k;\underline{\theta}) = C - \sum_{k=1}^{N} \frac{1}{2N_k} \int_{T_i}^{T_f} \Big[x_k(t) - \alpha_k s(t-\tau_k)\Big]^2 dt \qquad (B.1)$$

where C is a normalizing constant and $N_k = \beta_k \cdot N_0$ is the spectral level of $n_k(t)$. It immediately follows that

$$E\Big\{log f_{\underline{X}}(\underline{x};\underline{\theta})/Y = y;\underline{\theta}^{(n)}\Big\} = C' - \sum_{k=1}^{N} \frac{1}{2N_k} \int_{T_i}^{T_f} \Big[x_k^{(n)}(t) - \alpha_k s(t-\tau_k)\Big]^2 dt \qquad (B.2)$$

where

$$x_k^{(n)}(t) = E\Big\{x_k(t)/\sum_{k=1}^{N} x_k(t) = y(t);\underline{\theta}^{(n)}\Big\}, \qquad (B.3)$$

and C' contains all the terms that are independent of $\underline{\theta}$. Using Eq. (7) to carry out the conditional expectation required in (B.3), we obtain Eq. (19). Since the pair (τ_k, α_k) enters the right side of (B.2) only through the k term in the sum, the joint maximization of (B.4) with respect to the various τ_k's and α_k's decouples into the N separate maximizations as suggested by Eq. (20).

C Derivation of the EM Algorithm for Stochastic Gaussian Signals

Let the "complete" data $\underline{X}$ be generated by Fourier-analyzing the various $\underline{x}_k(t)$, that is,

$$\Bigg(\underbrace{X_1^*(\omega_1), \cdots, X_K^*(\omega_1)}_{\underline{X}(\omega_1)^\dagger}, \underbrace{X_1^*(\omega_2), \cdots, X_K^*(\omega_2)}_{\underline{X}(\omega_2)^\dagger}, \cdots \Bigg)^\dagger \tag{C.1}$$

where

$$X_k(\omega_l) = \frac{1}{\sqrt{T}} \int_{T_i}^{T_f} x_k(t) e^{-\jmath\omega_l t} dt \tag{C.2}$$

For observation time $T = T_f - T_i$ long compared with the correlation time of the signal and the noises, i.e. $WT/2\pi \gg 1$, the various $\underline{X}(\omega_\ell)$'s are asymptotically uncorrelated zero-mean Gaussian random vectors with the covariance matrix $E\Big\{\underline{X}(\omega_\ell)\underline{X}^\dagger(\omega_\ell)\Big\} = \Lambda(\omega_\ell;\underline{\theta})$, where $\Lambda(\omega;\underline{\theta})$ is defined in (30). It follows that

$$\log f_{\underline{X}}(\underline{X};\underline{\theta}) = \sum_\ell \log f_{\underline{X}(\omega_\ell)}\Big(\underline{X}(\omega_\ell);\underline{\theta}\Big) = -\sum_\ell \Big\{ \log\det\Big[\pi\Lambda(\omega_\ell;\underline{\theta})\Big] + X^*(\omega_\ell)\Lambda^{-1}(\omega_\ell;\underline{\theta})X(\omega_\ell) \Big\} \tag{C.3}$$

Now,

$$\begin{aligned}
\det\Lambda(\omega;\underline{\theta}) &= \det\Big[S(\omega)\underline{V}(\omega)\underline{V}^\dagger(\omega) + N(\omega)\cdot Q\Big] \\
&= \det[N(\omega)\cdot Q]\det\Big[I + \frac{S(\omega)}{N(\omega)}Q^{-1}\underline{V}(\omega)\underline{V}^\dagger(\omega)\Big] \\
&= \det[N(\omega)\cdot Q]\cdot\Big[1 + \frac{S(\omega)}{N(\omega)}\underline{V}^\dagger(\omega)Q^{-1}\underline{V}(\omega)\Big] \\
&= \Big[\Pi_{k=1}^N \beta_k N(\omega)\Big]\frac{S(\omega)/N^2(\omega)}{G(\omega)}
\end{aligned} \tag{C.4}$$

$$\begin{aligned}
\underline{X}^\dagger(\omega)\Lambda^{-1}(\omega;\underline{\theta})\underline{X}(\omega) &= \underline{X}^\dagger(\omega)\Big[S(\omega)\underline{V}(\omega)\underline{V}^\dagger(\omega) + N(\omega)\cdot Q\Big]^{-1}\underline{X}(\omega) \\
&= \frac{1}{N(\omega)}\underline{X}^\dagger(\omega)\Big[Q^{-1} - \frac{S(\omega)/N(\omega)}{1+\frac{S(\omega)}{N(\omega)}\underline{V}^\dagger(\omega)Q^{-1}\underline{V}(\omega)}Q^{-1}\underline{V}(\omega)\underline{V}^\dagger(\omega)Q^{-1}\Big]\underline{X}(\omega)
\end{aligned}$$

$$= \frac{1}{N(\omega)} \underline{X}^{\dagger}(\omega) Q^{-1} \underline{X}(\omega) - G(\omega) \underline{V}^{\dagger}(\omega) Q^{-1} \underline{X}(\omega) \underline{X}^{\dagger}(\omega) Q^{-1} \underline{V}(\omega) \qquad (C.5)$$

where $G(\omega)$ is defined in (33). Substituting (C.4) and (C.5) into (C.3), we obtain

$$\log f_{\underline{X}}(\underline{X}; \underline{\theta}) = C + \sum_{\omega_\ell} \Big[G(\omega_\ell) \underline{V}^{\dagger}(\omega_\ell) Q^{-1} \underline{X}(\omega_\ell) \underline{X}^{\dagger}(\omega_\ell) Q^{-1} \underline{V}(\omega_\ell) + \log\ G(\omega_\ell) \Big] \qquad (C.6)$$

where C contains all the terms that are independent of $\underline{\theta}$. It follows that

$$E\Big\{ log f_{\underline{X}}(\underline{X}; \underline{\theta}) / Y; \underline{\theta}^{(n)} \Big\} = C + \sum_{\omega_\ell} \Big[G(\omega_\ell) \underline{V}^{\dagger}(\omega_\ell) Q^{-1} \Psi^{(n)}(\omega_\ell) Q^{-1} \underline{V}(\omega_\ell) + \log\ G(\omega_\ell) \Big] \quad (C.7)$$

where $\Psi^{(n)}(\omega_\ell) = E\Big\{ \underline{X}(\omega_\ell) \underline{X}^{\dagger}(\omega_\ell); \underline{\theta}^{(n)} \Big\}$. Observing that $Y(\omega_\ell) = \mathbf{1}^T \underline{X}(\omega_\ell)$ and using Eq. (8) to compute the conditional expectation, $\Psi^{(n)}(\omega_\ell)$ is given by Eq. (28). Thus, the E-step of the algorithm consists of the computation of $\Psi^{(n)}(\omega_\ell)$ for all ω_ℓ in the signal frequency band, and the M-step is as defined by Eq. (29).

Acknowledgments

This study has been supported under Contract No. N00014-80-C-0381.

The authors wish to thank Dr. A.H. Nuttall and Dr. J.P. Ianniello of NUSC New London for many fruitful discussions, and to Ms. Cindy Leonard for her excellent secretarial assistance.

References

[1] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Annales of the Royal Statistical Society*, 1–38, 1977.

[2] L. E. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occuring in the statistical analysis of probabilistic functions of markov chains. *The Annals of MathematicalStatistics*, 41:164–171, 1970.

[3] H. O. Hartley and R. R. Hocking. The analysis of incomplete data. *Biometrics*, 27:783–808, 1971.

[4] T. Orchard and M. A. Woodsbury. A missing information principle: theory and applications. In *Proceedings of 6th Berkley Symposium on math. stat. and prob.*, pages 697–715, 1972.

[5] J. R. Grindon. *Estimation theory for a class of nonlinear random vector channels.* PhD thesis, Washington University, St. Louis, MO., 1970.

[6] J. S. Lim and A. V. Oppenheim. All pole modeling of degraded speech. *IEEE Trans. Acoustics, Speech, and Signal Processing*, ASSP-26:197–210, 1978.

[7] B. R. Musicus. *An iterative technique for maximum likelihood estimation with noisy data.* Master's thesis, Massachusetts Institute of Technology, Feb. 1979.

[8] B. R. Musicus. *Iterative algorithms for optimal signal reconstruction and parameter identification given noisy and incomplete data.* PhD thesis, Massachusetts Institute of Technology, September 1982.

[9] C. F. J. Wu. On the convergence properties of the EM algorithm. *The Annals of Statistics*, 11:95–103, 1983.

[10] B. D. O. Anderson and J. B. Moore. *Optimal Filtering.* Prentice Hall, Englewood Cliffs, NJ, 1979.

[11] H. L. Van Trees. *Detection Estimation and Modulation theory (Part I).* Wiley, New York, 1968.

[12] M. Feder and E. Weinstein. Optimal multiple source location estimation via the EM algorithm. In *Proceedings of the 1985 International Conference on Acoustics Speech and Signal Processing*, pages 1762–1765, 1985.

[13] M. Feder and E. Weinstein. Multipath and multiple source array processing via the EM algorithm. In *Proceedings of the 1986 International Conference on Acoustics Speech and Signal Processing*, 1986.

Iteration No.	Path 1		Path 2		Path 3		Path 4		Path 5	
	Delay	Amp.	Delay	Amp.	Delay	Amp.	Delay	Amp.	Delay	Amp.
1	75.123	0.80	94.655	1.46	106.325	1.45	109.035	0.89	126.024	0.71
2	75.375	0.164	95.009	3.59	106.962	1.64	109.785	0.95	100.37	0.08
5	104.087	0.03	95.902	2.01	108.983	2.04	90.124	1.18	100.989	1.31
10	104.846	0.63	95.908	1.69	108.995	1.67	91.955	1.44	100.946	1.01
30	104.847	0.64	95.909	1.69	108.995	1.66	91.956	1.43	100.946	1.008

Table 1: Deterministic signals, Example 1

Iteration No.	Path 1		Path 2		Path 3	
	Delay	Amplitude	Delay	Amplitude	Delay	Amplitude
1	110.284	0.039	100.072	4.069	90.601	0.64
2	109.529	0.707	100.229	3.823	90.777	0.764
5	108.727	0.776	100.156	3.68	91.882	0.81
10	107.83	0.89	100.106	3.43	92.795	0.90
30	106.043	1.318	100.093	2.448	94.914	1.404

Table 2: Deterministic signals, Example 2

Iteration No.	Delay	Amplitude
1	2.738	0.789
2	2.980	0.807
5	3.6776	0.706
10	4.0835	0.698
30	4.2033	0.6918

Table 3: Stochastic signals, Example 1

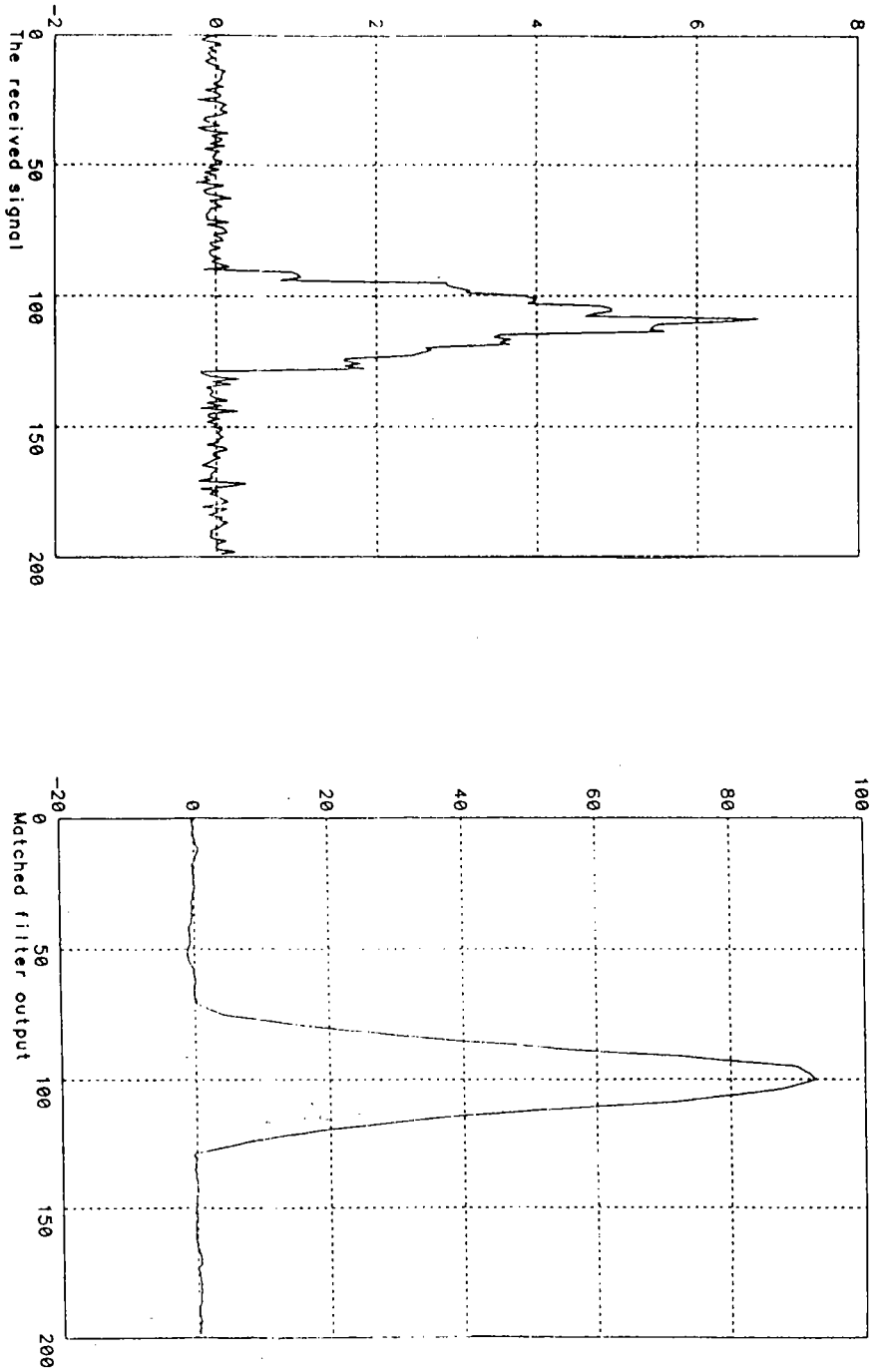


Figure 1: Det. signals, example 1 - The received signal and a conventional matched filter

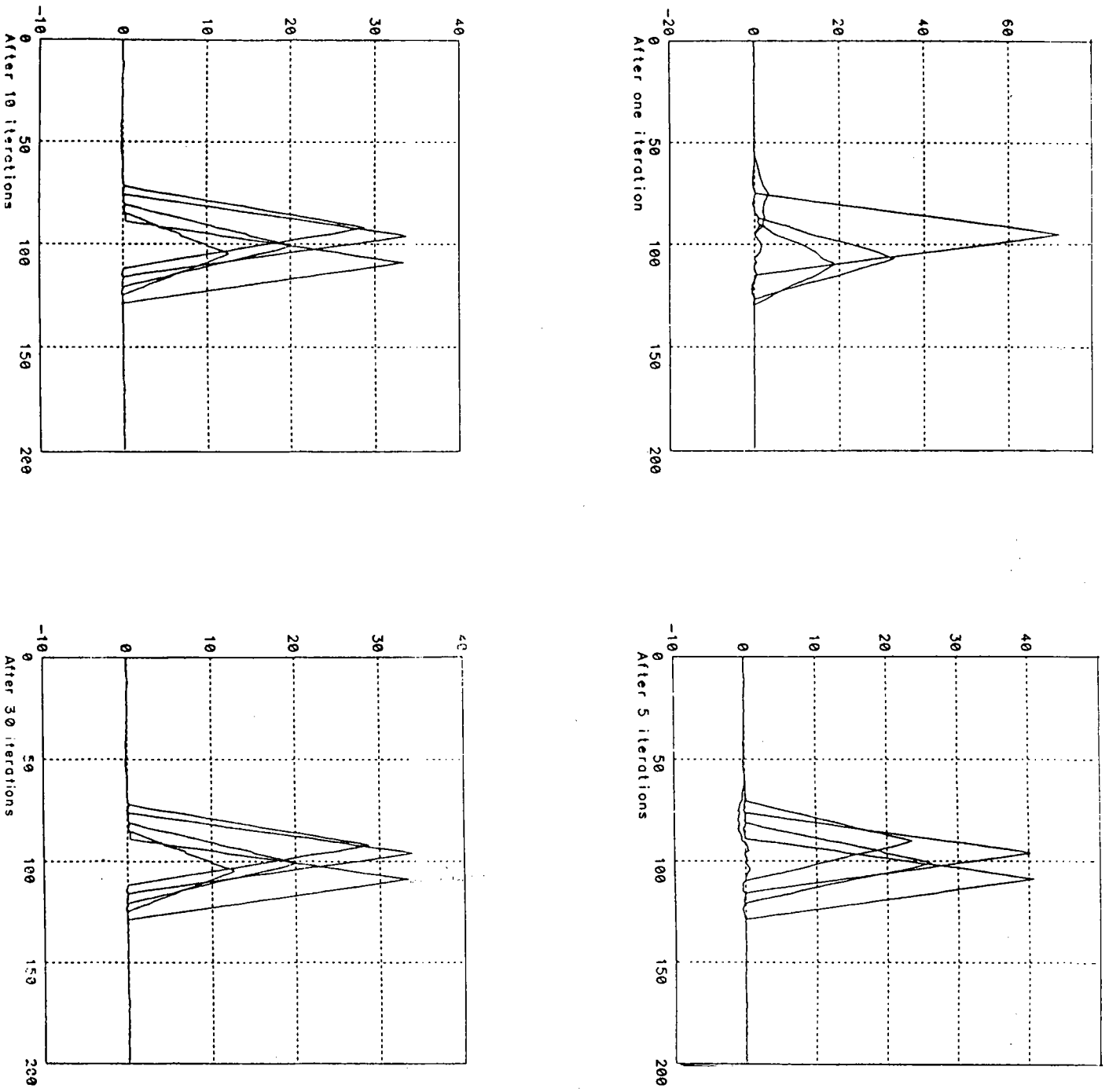


Figure 2: Det. signals, example 1 - The M step output of the proposed algorithm

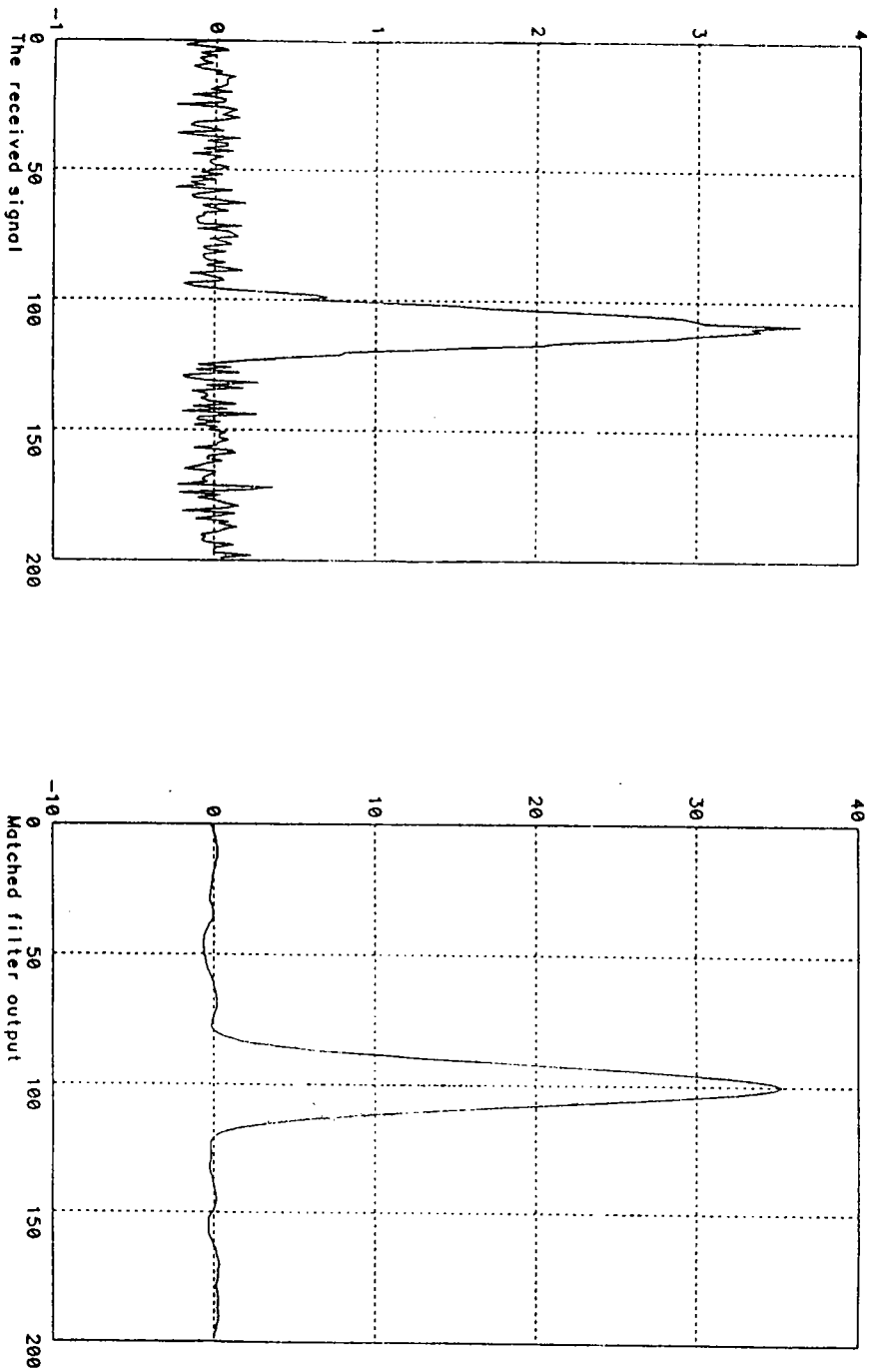


Figure 3: Det. signals, example 2 - the received signal and a conventional matched filter

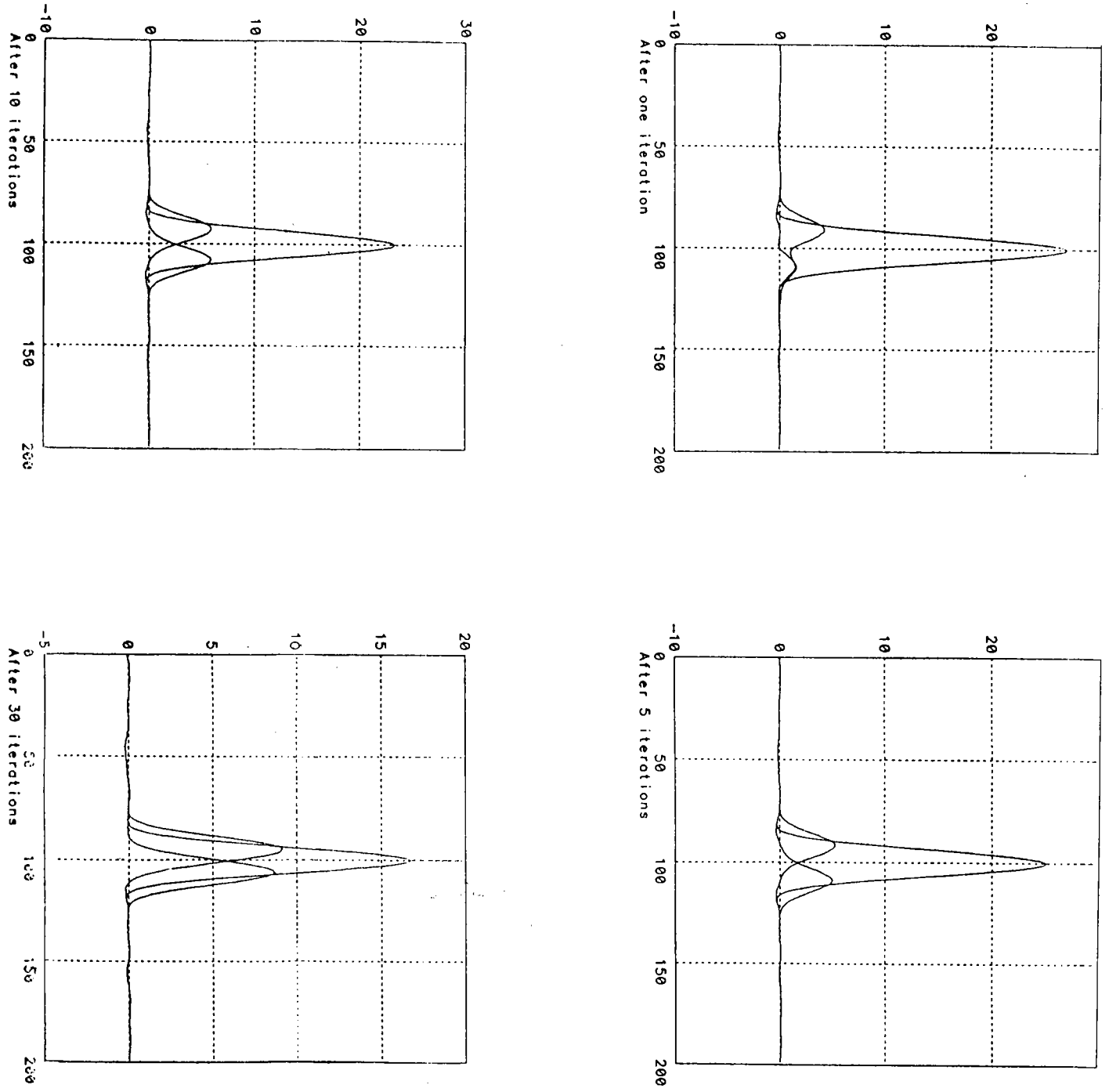


Figure 4: Det. signals, example 2 - The M step output of the proposed algorithm

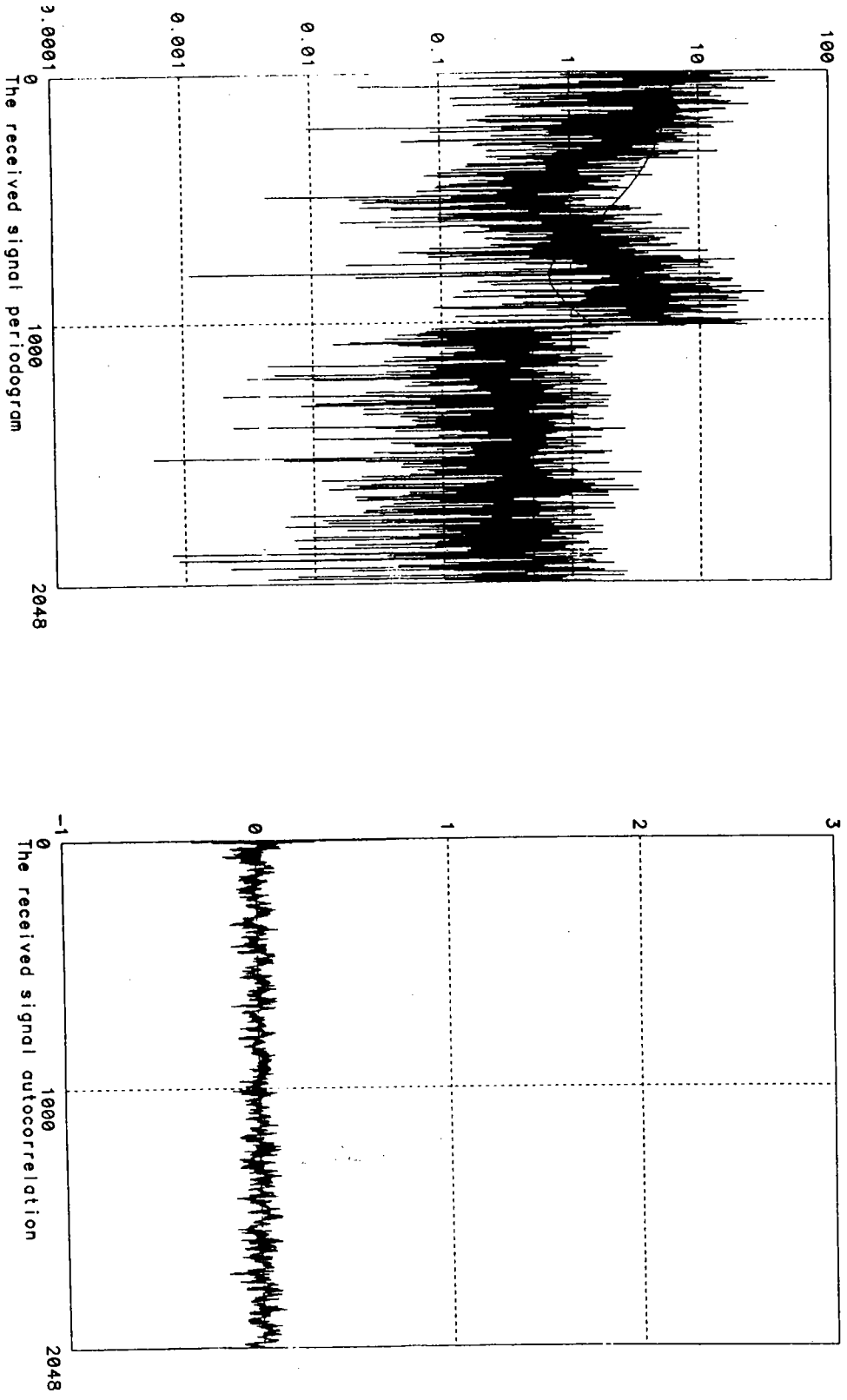


Figure 5: Det. signals - The autocorrelation and periodogram of the received signal

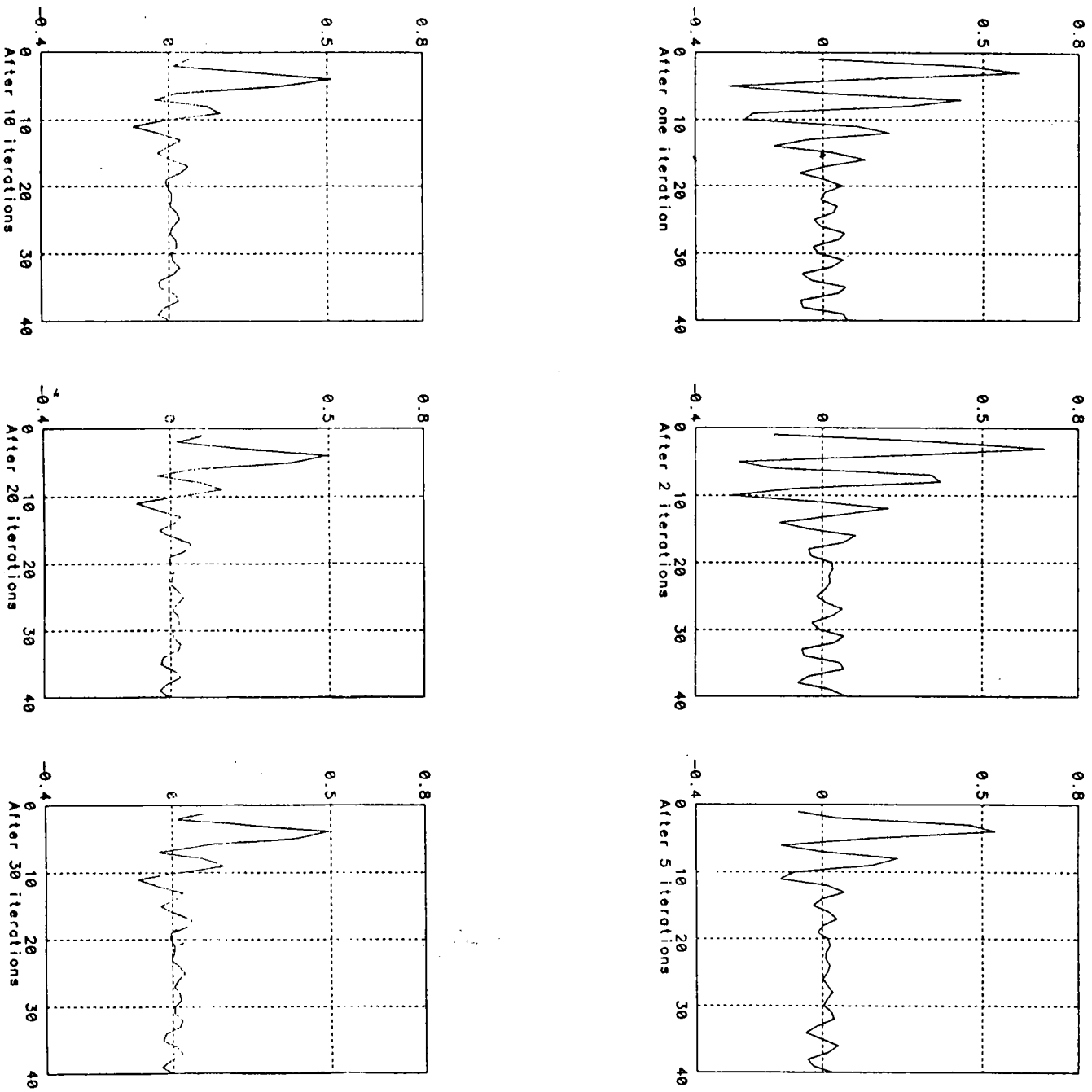


Figure 6: Stochastic signals - The M step output of the proposed algorithm

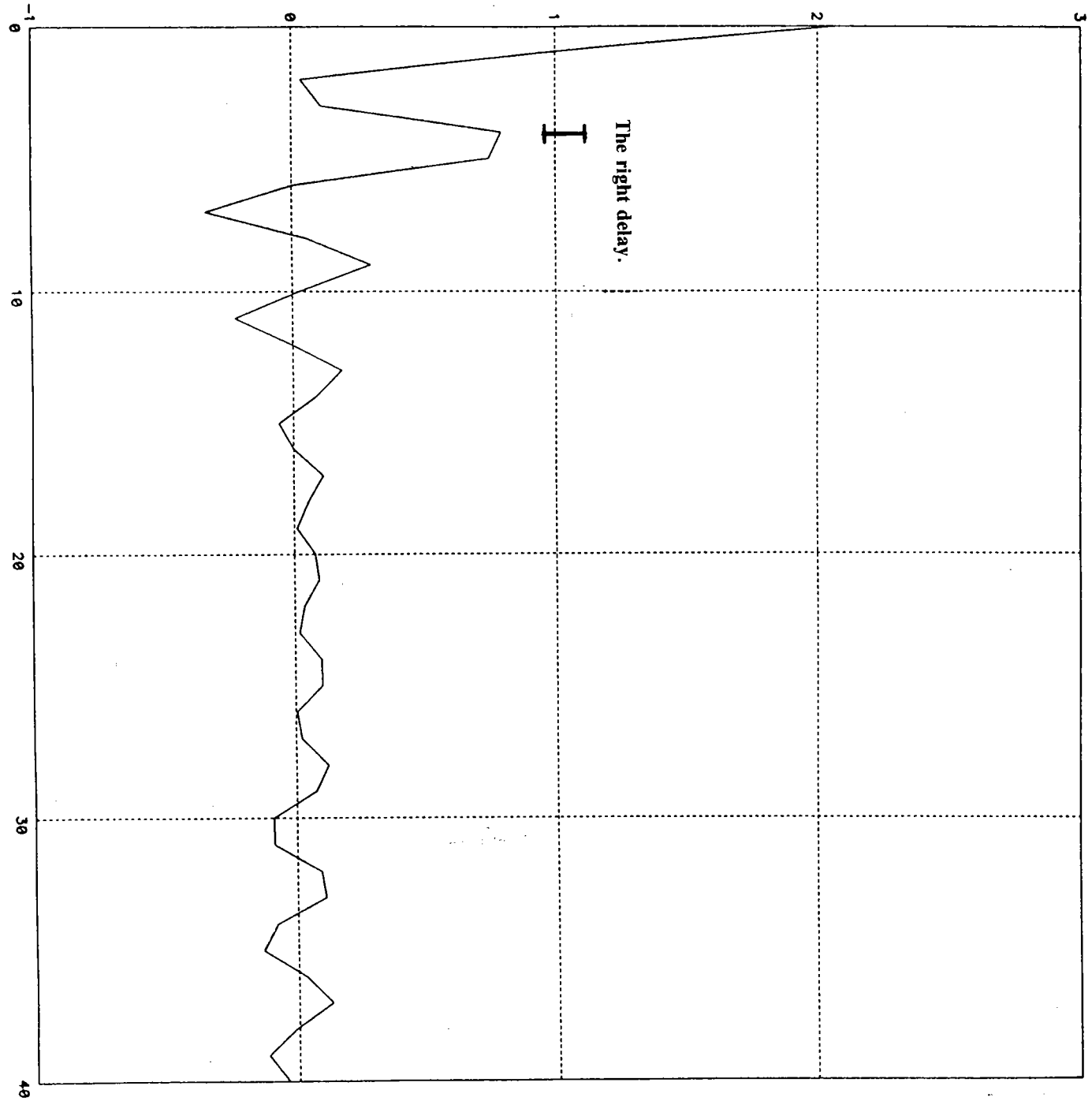


Figure 7: Zoomed Autocorrelation

DOCUMENT LIBRARY

November 21, 1986

Distribution List for Technical Report Exchange

Institute of Marine Sciences Library
University of Alaska
O'Neill Building
905 Koyukuk Ave., North
Fairbanks, AK

Attn: Stella Sanchez-Wade
Documents Section
Scripps Institution of Oceanography
Library, Mail Code C-075C
La Jolla, CA 92093

Hancock Library of Biology & Oceanography
Alan Hancock Laboratory
University of Southern California
University Park
Los Angeles, CA 90089-0371

Gifts & Exchanges
Library
Bedford Institute of Oceanography
P.O. Box 1006
Dartmouth, NS, B2Y 4A2, CANADA

Office of the International
Ice Patrol
c/o Coast Guard R & D Center
Avery Point
Groton, CT 06340

Library
Physical Oceanographic Laboratory
Nova University
8000 N. Ocean Drive
Dania, FL 33304

NOAA/EDIS Miami Library Center
4301 Rickenbacker Causeway
Miami, FL 33149

Library
Skidaway Institute of Oceanography
P.O. Box 13687
Savannah, GA 31416

Institute of Geophysics
University of Hawaii
Library Room 252
2525 Correa Road
Honolulu, HI 96822

Library
Chesapeake Bay Institute
4800 Atwell Road
Shady Side, MD 20876

MIT Libraries
Serial Journal Room 14E-210
Cambridge, MA 02139

Director, Ralph M. Parsons Laboratory
Room 48-311
MIT
Cambridge, MA 02139

Marine Resources Information Center
Bldg. E38-320
MIT
Cambridge, MA 02139

Library
Lamont-Doherty Geological Observatory
Colombia University
Palisades, NY 10964

Library
Serials Department
Oregon State University
Corvallis, OR 97331

Pell Marine Science Library
University of Rhode Island
Narragansett Bay Campus
Narragansett, RI 02882

Working Collection
Texas A&M University
Dept. of Oceanography
College Station, TX 77843

Library
Virginia Institute of Marine Science
Gloucester Point, VA 23062

Fisheries-Oceanography Library
151 Oceanography Teaching Bldg.
University of Washington
Seattle, WA 98195

Library
R.S.M.A.S.
University of Miami
4600 Rickenbacker Causeway
Miami, FL 33149

Maury Oceanographic Library
Naval Oceanographic Office
Bay St. Louis
NSTL, MS 39522-5001
ATTN: Code 4601

272-101

REPORT DOCUMENTATION PAGE

1. REPORT NO. WHOI-87-3

2.

3. Recipient's Accession No.

4. Title and Subtitle
Multipath Time-Delay Estimation via the EM Algorithm

5. Report Date
January 1987

6.

7. Author(s)
Meir Feder and Ehud Weinstein

8. Performing Organization Rept. No.
WHOI-87-3

9. Performing Organization Name and Address

Woods Hole Oceanographic Institution
Woods Hole, Massachusetts 02543

10. Project/Task/Work Unit No.

11. Contract(C) or Grant(G) No.
(C) N00014-80-C-0381
(G)

12. Sponsoring Organization Name and Address
Naval Underwater Systems Center
Code 32, Building 80
New London, CT 06320
Attn: Jack Iannelo

13. Type of Report & Period Covered
Technical

14.

15. Supplementary Notes
This report should be cited as: Woods Hole Oceanog. Inst. Tech. Rept., WHOI-87-3.

16. Abstract (Limit: 200 words)

We consider the application of the EM algorithm to the multipath time delay estimation problem. The algorithm is developed for the case of deterministic (known) signals, as well as for the case of wide-sense stationary Gaussian signals.

17. Document Analysis a. Descriptors

1. Time-delay estimation
2. EM algorithm
3. Optimal estimation

b. Identifiers/Open-Ended Terms

c. COSATI Field/Group

18. Availability Statement:
Approved for publication; distribution unlimited.

19. Security Class (This Report)
UNCLASSIFIED

20. Security Class (This Page)

21. No. of Pages
29

22. Price