KALMAN PARAMETRIC AND NONPARAMETRIC SYSTEM IDENTIFICATION APPLIED TO ECHO CANCELLATION

Sasan Ardalan

CENTER FOR COMMUNICATIONS AND SIGNAL PROCESSING

North Carolina State University
Raleigh, NC 27650

CCSP-TR-83/4

May 13, 1983

ABSTRACT

The Kalman estmation algorithm for system identification is derived. Nonparametric estimation of the system impulse response and parametric estimation of the system poles and zeroes are treated. The Fast Kalman estimation algorithm for both parametric and nonparametric estimation is introduced. Simulation results applying parametric and nonparametric Fast Kalman estimation to the identification of a 10th order bandpass digital IIR filter are presented. Comparisons are made with the gradient search algorithm. It is shown that the Kalman algorithms outperform the gradient search algorithm in terms of convergence rate. The Parametric Fast Kalman algorithm converges very rapidly with superior estimate accuracy compared to the nonparametric method. The application of the estimation methods to adaptive echo cancellation is also studied.

TABLE OF CONTENTS

PARAMETRIC AND NONPARAMETRIC FAST KALMAN SYSTEM IDENTIFICATION

1 Introduction

One important area of modern signal processing research is that of systems identification. Given a known or derivable system input $x(t)$ and an observable system output $y(t)$, we desire to identify the impulse response, $h(t)$, of the unknown system. Two common methods of approaching this problem are denoted as the nonparametric method and the parametric method. The more general of the approaches is the nonparametric method in which no constraints, except for filter realizability, are placed upon the resulting solution. The "solution" is typically the N-point discrete approximation to the unknown system impulse response. Parametric methods, on the other hand, provide a model of the unknown system in terms of certain parameters which define the system completely. One popular method concerns that of describing the unknown system in terms of a rational pole-zero model. Typically, the number of parameters describing the model is much less than N, the effective number of points in the impulse response. It will be seen that this consequence allows us to use much shorter filter algorithms to estimate the pole-zero parameters, and achieve a substantial decrease in convergence time for the model identification algorithms.

Since echo cancellation can be formulated as a systems identification problem, the results of this report can be directly applied to the echo cancellation problem.

2. Theoretical Development of Nonparametric Solution

Consider the problem of estimating an unknown system's impulse response as displayed in Figure 2.1. Assume that due to system dynamics or prior filtering that the system h(t) is bandlimited in response. Based upon the Nyquist criteria, we can therefore sample the input and output of the system at the Nyquist rate with no loss of information. Thus, we may represent the continuous system identification problem by its discrete counterpart shown in Figure 2.2. For a causal, linear, time-invariant system we may write

$$y(n) = \sum_{k=0}^{\infty} h(k)\, x(n-k) \tag{2.1}$$

since $h(k)=0$ for $k<0$. If we assume that $h(k) \approx 0$ for $k>N$ and form the vectors $\underline{h}$ and $\underline{x}(n)$ by

$$\underline{h} = \left[h(0)\ h(1)\ \dots\ h(N-1)\right]^T \tag{2.2a}$$

$$\underline{x}(n) = \left[x(n)\ x(n-1)\ \dots\ x(n-N+1)\right]^T \tag{2.2b}$$

we may write (2.1) as

$$y(n) = \underline{x}^T(n)\underline{h} \tag{2.3}$$

Now consider the system's identification problem of estimating the impulse response vector, $\underline{h}$, from the input sequence $\underline{x}(n)$ and the observed output sequence $z(n)$,

$$z(n) = y(n) + v(n) \tag{2.4}$$

where $v(n)$ is an additive noise component which contaminates the time output measurement. (For this initial development we assume $\underline{h}$ to be constant.)

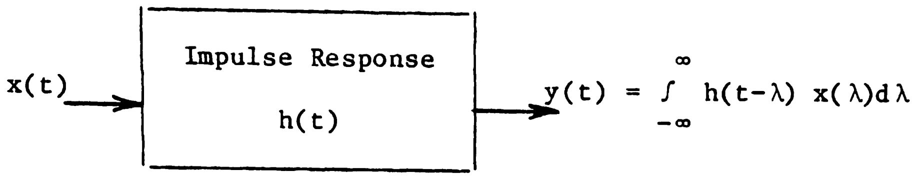

Figure 1. Bandlimited Continuous System

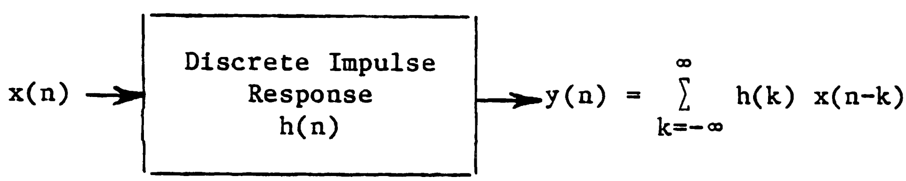

Figure 2. Equivalent Discrete System Representation

Suppose our system's identification algorithm (as yet undefined) produces at time k an estimate of $\underline{h}$, denoted by $\underline{b}(k)$. Then our estimate of the output becomes $\hat{y}(k)$,

$$\hat{y}(k) = \underline{x}^T(k)\ \underline{b}(k) \tag{2.5}$$

and our estimation error is given by

$$e(k) = y(k) - \hat{y}(k) \tag{2.6}$$

For the identification of the system we will seek to minimize the ensemble mean square estimation error,

$$\zeta(k) = E\{e^2(k)\} \tag{2.7}$$

where $E\{\cdot\}$ denotes the expectation operator. In other words, we want to obtain an estimate $\underline{b}(k)$ of $\underline{h}$ such that

$$\frac{\partial \zeta(k)}{\partial \underline{b}(k)} = \underline{0} \quad . \tag{2.8}$$

This method is a standard technique in optimization and the details are carried out in Appendix A. From Appendix A, it is seen that the solution for $\underline{b}(k)$ obtained from solving the N simulataneous equations in (2.8) is given by the well-known matrix equation

$$\underline{b}_o(k) = A^{-1}\underline{c} \tag{2.9}$$

where $\underline{b}_o(k)$ is the optimal system estimation at the kth time sample, and

$$A = E\ \{\underline{x}(k)\ \underline{x}^T(k)\}\ , \tag{2.10}$$

and

$$\underline{c} = E\{y(k)\ \underline{x}(k)\} \quad . \tag{2.11}$$

The NxN matrix A is sometimes denoted as the data autocorrelation matrix and $\underline{c}$ is called the cross-correlation vector.

It is a simple matter to show that the error, $e_o(k)$, resulting from using the optimal filter is orthogonal to the data vector $\underline{x}(k)$. Consider the vector $\underline{L}$,

$$\underline{L} = E\ \{\underline{x}(k)\ e_o(k)\}\ , \tag{2.12}$$

where

$$e_o(k) = y(k) - \underline{x}^T(k)\underline{b}_o \quad . \tag{2.13}$$

Substituting (2.13) into (2.12) and expanding,

$$\underline{L} = E\{y(k)\ \underline{x}(k) - \underline{x}(k)\ \underline{x}^T(k)\underline{b}_o\}$$

$$= E\{y(k)\ \underline{x}(k)\} - E\{\underline{x}(k)\ \underline{x}^T(k)\}\ \underline{b}_o \tag{2.14}$$

But substituting from (2.9) -(2.11), we get

$$L = \underline{c} - A(A^{-1}\underline{c}) = 0 \tag{2.15}$$

This is a statement of the orthogonality condition for minimum mean square error (MMSE) filters, namely, that at all time iterations k, the MMSE filter is that filter which produces an error e(k) which is orthogonal to the data used in forming the estimate y(k).

To illustrate the meaning of (2.9), let the impulse response of the system be finite. Thus from (2.3),

$$y(k) = \underline{x}^T(k)\cdot\underline{h} \tag{2.16}$$

Substituting into (2.11) and realizing that $\underline{h}$ is independent of $\underline{x}(k)$, we get

$$\underline{c} = E[\underline{x}(k)\ \underline{x}^T(k)]\cdot\underline{h} \tag{2.17}$$

and from (2.10),

$$\underline{c} = A\ \underline{h} \tag{2.18}$$

Finally from (2.9) we derive

$$\underline{b}_o = A^{-1}\cdot A\cdot\underline{h} = \underline{h} \tag{2.19}$$

Returning to the definition of $\zeta(k)$ from (2.7), we may obtain

$$\zeta(k) = E\{[\underline{x}^T(k)\ \underline{b}_o - \underline{x}^T(k)\underline{b}(k) + e_o(k)]^T$$

$$\cdot[\underline{x}^T(k)\ \underline{b}_o - \underline{x}^T(k)\ \underline{b}(k) + e_o(k)]\} \tag{2.20}$$

where we have used the definition of $e_o(k)$ from (2.13). Denoting

$$\zeta_o = E\{e_o^2(k)\} \tag{2.21}$$

and using the relationship that $e_o(k)$ and $\underline{x}(k)$ are uncorrelated, as previously shown, we can write

$$\zeta(k) = E\{[\underline{b}_o - \underline{b}(k)]^T \underline{x}(k) \underline{x}^T(k) [\underline{b}_o - \underline{b}(k)]\}+\zeta_o . \tag{2.22}$$

Let us define a coefficient error covariance matrix at time k by the matrix P(k),

$$P(k) = E\{[\underline{b}_o - \underline{b}(k)]^T [\underline{b}_o - \underline{b}(k)]\} . \tag{2.23}$$

Then by using the following matrix identity,

$$\underline{x}^T \underline{y} \underline{y}^T \underline{x} = Tr(XY) , \tag{2.24}$$

where $X = \underline{xx}^T$, $y = \underline{yy}^T$, and Tr(Q) is the trace of the matrix Q, we can write (2.22) as

$$\zeta(k) = Tr [P(k)A] + \zeta_o \tag{2.25}$$

The above equation assumes the coefficient error is independent of $\underline{x}(k)$. If this is not the case, denote $P'(k) = [\underline{b}_o-\underline{b}(k)][\underline{b}_o-\underline{b}(k)]^T$ and $A' = \underline{x}(k) \underline{x}^T(k)$. Then

$$\zeta(k) = Tr [E\{P'(k)A'\}] + \zeta_o . \tag{2.26}$$

3. Kalman Estimation

In the previous development we denoted the estimated impulse response vector at time k as $\underline{b}(k)$. At each time k+1 we get a new measurement z(k+1) and based on this new information and the prior estimate $\underline{b}(k)$ we seek an updated estimate $\underline{b}(k+1)$. Equivalently we may use past information and the kth measurement to estimate $\underline{b}(k)$. Thus,

$$\underline{b}(k) = K'(k)\ \underline{b}(k-1) + \underline{K}(k)\ z(k) \tag{3.1}$$

where $K'(k)$ is an NxN gain matrix and $\underline{K}(k)$ is an Nx1 gain vector. Using the measurement model of (2.4) we may write

$$\underline{b}(k) = K'(k)\ \underline{b}(k-1) + \underline{K}(k)\underline{x}^T(k)\ \underline{h} + \underline{K}(k)\ v(k) \tag{3.2}$$

Define the error, $\underline{e}_b(k)$, in the estimated impulse response vector as

$$\underline{e}_b(k) = \underline{b}(k) - \underline{b}_o\ . \tag{3.3}$$

Substituting $\underline{b}(k)$ defined in (3.2) into (3.3) produces

$$\underline{e}_b(k) = K'(k)\ \underline{b}(k-1) + [\underline{K}(k)\ \underline{x}^T(k) - I]\ \underline{b}_o + \underline{K}(k)\ v(k) \tag{3.4}$$

But from (3.3),

$$\underline{b}(k-1) = \underline{b}_o + \underline{e}_b(k-1) \tag{3.5}$$

Substituting this into (3.4) then produces

$$\underline{e}_b(k) = K'(k)\ \underline{e}_b(k-1) + [K'(k) + \underline{K}(k)\ \underline{x}^T(k) - I]\underline{b}_o + \underline{K}(k)\ v(k) \tag{3.6}$$

Taking expectations of both sides of (3.6) and recognizing that $E[v(k)]=0$, one obtains

$$E\{\underline{e}_b(k)\} = K'(k)\ E\{\underline{e}_b(k)\} + \underline{b}_o\ E\{K'(k) + \underline{K}(k)\ \underline{x}^T(k) - I\} \tag{3.7}$$

In general, $K'(k)$ is not equal to the null matrix (i.e., matrix with all zero elements). Thus from (3.7) a sufficient condition for there to be no bias in the impulse response estimate (i.e. $E\{\underline{\varepsilon}_b(k)\}\to 0$) is that

$$K'(k)+\underline{K}(k)\underline{x}^T(k) - I = 0 \tag{3.8}$$

From (3.8) we may solve for the gain matrix $K'(k)$,

$$K'(k) = I - \underline{K}(k)\ \underline{x}^T(k) \tag{3.9}$$

Then substituting (3.9) into (3.1) and rearranging produces

$$\underline{b}(k) = \underline{b}(k-1) + \underline{K}(k)\ [z(k) - \underline{x}^T(k)\ \underline{b}(k-1)] \tag{3.10}$$

Equation (3.10) states basically that the new estimate is equal to the old estimate plus a correction term which is proportional to the error between the new measurement and a predicted measurement. It can be seen in (3.10) that the only unknown quantity remaining is the gain vector $\underline{K}(k)$. We will now derive this gain vector such that the mean square estimation error $\zeta(k)$ is minimized. We will essentially solve the set of N simultaneous equations given by

$$\frac{\partial\zeta(k)}{\partial\underline{K}(k)} = \underline{0} \tag{3.11}$$

We have seen from (2.26) that $\zeta(k)$ is a function of $P'(k)$, the error covariance iance matrix. Thus, we need to relate $P'(k)$ to $\underline{K}(k)$. From (2.26),

$$P'(k) = [\underline{b}_o - \underline{b}(k)][\underline{b}_o-\underline{b}(k)]^T\ . \tag{3.12}$$

Substituting (3.9) into (3.1), adding $\underline{b}_o$ to both sides and simplifying produces

$$[\underline{b}_o-\underline{b}(k)] = [I-\underline{K}(k)\underline{x}^T(k)][\underline{b}_o-\underline{b}(k-1)]\underline{K}(k)\ v(k) \tag{3.13}$$

From (3.3) and (3.12) we see immediately that

$$P'(k) = \underline{\varepsilon}_b(k)\ \underline{\varepsilon}_b^T(k) \tag{3.14}$$

and, using (3.13),

$$P'(k) = \{[I-\underline{K}(k)\ \underline{x}^T(k)]\underline{\varepsilon}_b(k-1) - \underline{K}(k)\ v(k)\}$$
$$\cdot\{[I-\underline{K}(k)\underline{x}^T(k)]\underline{\varepsilon}_b(k-1) - \underline{K}(k)\ v(k)\}^T \tag{3.15}$$

Let us denote

$$r(k) = E\{v^2(k)\}\quad , \tag{3.16}$$

to be the (possibly time-varying) variance of the measurement noise. We may make use of the property that the measurement errors are uncorrelated with $\underline{x}(k)$ and $\underline{\varepsilon}_b(k)$, producing

$$E\{\underline{\varepsilon}_b(k-1)\ v(k)\} = 0\quad . \tag{3.17}$$

Then from (2.23),

$$P(k) = E\ \{P'(k)\}$$
$$= E\ \{[I-\underline{K}(k)\ \underline{x}^T(k)]\ \underline{\varepsilon}_b(k-1)\ \underline{\varepsilon}_b^T\ (k-1)\ [I-\underline{K}(k)\underline{x}^T(k)]^T\}$$
$$+\ \underline{K}(k)\ r(k)\ \underline{K}^T(k) \tag{3.18}$$

or

$$P(k) = [I-\underline{K}(k)\underline{x}^T(k)]\ P(k-1)[I-\underline{K}(k)\ \underline{x}^T(k)]^T$$
$$+\underline{K}(k)\ r(k)\ \underline{K}^T(k) \tag{3.19}$$

If we neglect any correlation between the estimate error and input vector, then we can write (2.27) as

$$\zeta(k) = \mathrm{Tr}[P(k)\ A] + \zeta_o \tag{3.20}$$

We may make use of the identity

$$\frac{\partial}{\partial \underline{A}}\ [\mathrm{Tr}(\underline{A}\underline{B}\underline{A}^T)] = 2\underline{A}\underline{B} \tag{3.21}$$

when B is symmetric. Thus, substituting for P(k) and using (3.21) and standard vector differention properties, we get

$$\frac{\partial\zeta(k)}{\partial\underline{K}(k)} = -\ 2[I-\underline{K}(k)\ \underline{x}^T(k)]\ P(k-1)\ \underline{x}(k)A + 2\underline{K}(k)r(k)A \tag{3.22}$$

Setting the quantity in (3.22) equal to zero and solving for the resulting $\underline{K}(k)$ gives us the desired result:

$$\underline{K}(k) = P(k-1)\ \underline{x}(k)\ [\underline{x}^T(k)\ P(k-1)\ \underline{x}(k) + r(k)]^{-1} \tag{3.23}$$

This vector $\underline{K}(k)$ is now in the desired form and is called the Kalman gain vector.

Another quantity needed is an update equation for the estimation error covariance matrix, P(k). This may be obtained by substituting (3.23) into (3.19) and simplifying. The procedure is lengthy, although straightforward, and yields

$$P(k) = P(k-1) - P(k-1)\ \underline{x}(k)\ [\underline{x}^T(k)\ P(k-1)\ \underline{x}(k) + r(k)]^{-1}\underline{x}^T(k)\ P(k-1) \tag{3.24}$$

or by substituting (3.23),

$$P(k) = P(k-1) - \underline{K}(k)\ \underline{x}^T(k)\ P(k-1)\ . \tag{3.25}$$

Thus, we have a recursive update for the P(k) matrix. It is possible to obtain a more compact recursion formula if we write P(k) in terms of its inverse through the use of the <u>matrix inversion lemma</u> [4].

The matrix inversion lemma is a well-known technique in numerical methods and states that if

$$P^{-1}(n+1) = P^{-1}(n) + \underline{H}^T(n+1)\ R^{-1}(n+1)\ \underline{H}(n+1),$$

then

$$P(n+1) = P(n) - \underline{P}(n)\ \underline{H}^T(n+1)\left[\underline{H}(n+1)\ P(n)\ \underline{H}^T(n+1) + R(n+1)\right]^{-1} \underline{H}(n+1)\underline{P}(n).$$

Applying the lemma to (3.24) we obtain

$$P^{-1}(k) = P^{-1}(k-1) + \underline{x}(k)\ r^{-1}(k)\ \underline{x}^T(k) \tag{3.26}$$

where $\underline{x}(k)$ is associated with $\underline{H}(n)$ and $R(n)$ is associated with $r(k)$, a scalar. Note that (3.26) may be rewritten to reflect the initial state $P^{-1}(0)$ as

$$P^{-1}(k) = r^{-1}(k)\ \left[\sum_{i=1}^{k} \underline{x}(i)\ \underline{x}^T(i)\right] + P^{-1}(0)\ . \tag{3.27}$$

To display the impact of this result, write (3.23) as

$$\underline{K}(k) = P(k)\ P^{-1}(k)\ P(k-1)\ \underline{x}(k)\ \left[\underline{x}^T(k)\ P(k-1)\ \underline{x}(k)\ r(k)\right]^{-1}. \tag{3.28}$$

Substituting (3.26) into (3.28), and simplying, it is straightforward to show that

$$\underline{K}(k) = P(k)\ \underline{x}(k)\ r^{-1}(k)\ . \tag{3.29}$$

Substituting (3.27) into (3.29) for $P(k)$ we have

$$\underline{K}(k) = \left[\sum_{i=1}^{k} \underline{x}(i)\ \underline{x}^T(i)\right]^{-1}\underline{x}(k). \tag{3.30}$$

An important property of the Kalman estimation method is that the estimated state $\underline{b}(k)$ is orthogonal to the estimated state error,

$$E\ \{\underline{b}(k)\ \underline{\tilde{b}}^T(k)\} = 0 \tag{3.31}$$

This derivation is included in Appendix B.

Analysis of Speed of Convergence [1]

In (2.25) we have shown that the mean square error at time k is given by

$$\zeta(k) = \mathrm{Tr}\ [P(k)A] + \zeta_0 \tag{3.32}$$

Also, we have shown in (3.27) that

$$P(k) = r(k)\ \left[\sum_{i=1}^{k} \underline{x}(i)\ \underline{x}(i)\right]^{-1} \tag{3.33}$$

where we have neglected P(0). Since $r(k) = E\ [v^2(k)]$ then $r(k) = \zeta_0$. Now, we may approximate the matrix A, defined in (2.10), by

$$A \approx \frac{1}{k} \sum_{i=1}^{k} \underline{x}(i)\ \underline{x}^T(i) \tag{3.34}$$

as $k \to \infty$. Substituting (3.34) and (3.33) into (3.32) we get

$$\zeta(k) \approx \mathrm{Tr}\ \{\zeta_0 [\sum_{i=1}^{k} \underline{x}(i)\underline{x}^T(i)]^{-1} (\tfrac{1}{k})\ [\sum_{i=1}^{k} \underline{x}(i)\cdot\underline{x}^T(i)]\} + \zeta_0 \tag{3.35}$$

$$= \mathrm{Tr}\ [\frac{\zeta_0 I}{k}] + \zeta_0 \tag{3.36}$$

Thus, we have the result for the approximation to the speed of convergence,

$$\zeta(k) \approx \zeta_0\ (1 + \frac{N}{k}) \tag{3.37}$$

The mean square estimation error of the Kalman filter therefore decreases inversely to the number of iterations k. Further, the initial error (k=1) is proportional to the filter length N. The optimum error ζ_0 is approached rapidly as k>N.

KALMAN ESTIMATION SUMMARY

System model:

$$z(k) = y(k)+v(k)$$

where $v(k)$ is additive stationary noise and

$$E\{v^2(k)\} = r(k)$$

$$E\{v(k)\} = 0$$

State Estimation Update:

$$\underline{b}(k) = \underline{b}(k-1) + \underline{K}(k)[z(k)-\hat{y}(k)]$$

$$\hat{y}(k) = \underline{x}^T(k)\underline{b}(k-1)$$

$$\underline{K}(k) = \text{Kalman Gain vector}$$

Kalman Gain Update:

$$\underline{K}(k) = P(k-1)\underline{x}(k)[\underline{x}^T(k)P(k-1)\underline{x}(k)+r(k)]^{-1}$$

$$\underline{K}(k) = P(k)\underline{x}(k)r^{-1}(k)$$

$$\underline{K}(k) = [\sum_{i=1}^{k} \underline{x}(i)\underline{x}^T(i)]^{-1}\underline{x}(k)$$

Error Covariance Update:

$$P(k) = P(k-1)-P(k-1)\underline{x}(k)[\underline{x}^T(k)P(k-1)\underline{x}(k)+r(k)]^{-1}\underline{x}^T(k)P(k-1)$$

$$P(k) = P(k-1)-\underline{K}(k)\underline{x}^T(k)P(k-1)$$

$$P(k) = r^{-1}(k)[\sum_{i=1}^{k} \underline{x}(i)\underline{x}^T(i)]+P(0)$$

Properties:

(i) The mean square estimation error propagates according to

$$\zeta(k) = Tr[P(k)A]+\zeta_0$$

where

$$A = E\{\underline{x}(k)\underline{x}^T(k)\}$$

(ii) The error above converges to the optimum error, ζ_0, according

$$\zeta(k) = \zeta_0(1+N/k)$$

(iii) The expectation of the state estimation error is given by

$$|\underline{e}_b(k)|^2 = Tr[P(k)]$$

(iv) The optimum (minimum) mean square error filter has the property that the optimum state estimate and the error are orthogonal,

$$E\{\underline{b}(k)\underline{e}_b^T(k)\} = 0$$

Finally, the block diagram in figure 3 illustrates the computation flow in the nonparametric Kalman estimation algorithm. Figure 4 then illustrates the system's identification approach to the same problem.

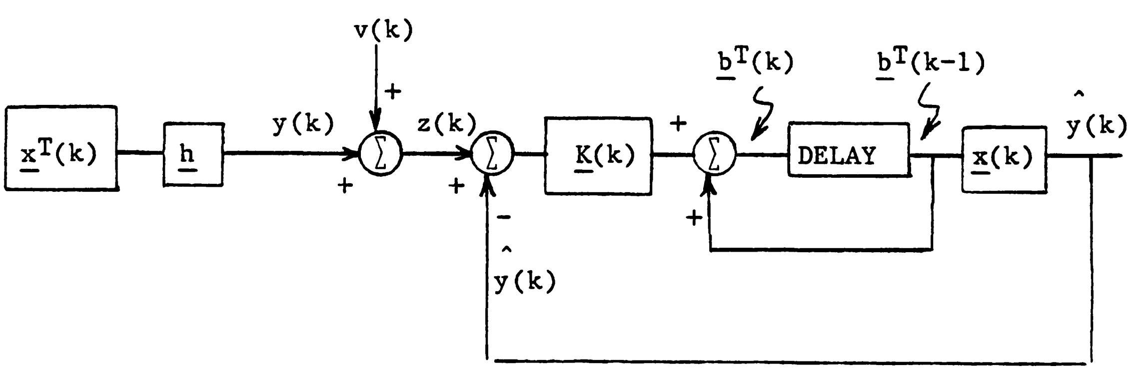

Figure 3. Block Diagram Illustrating Computation Sequence

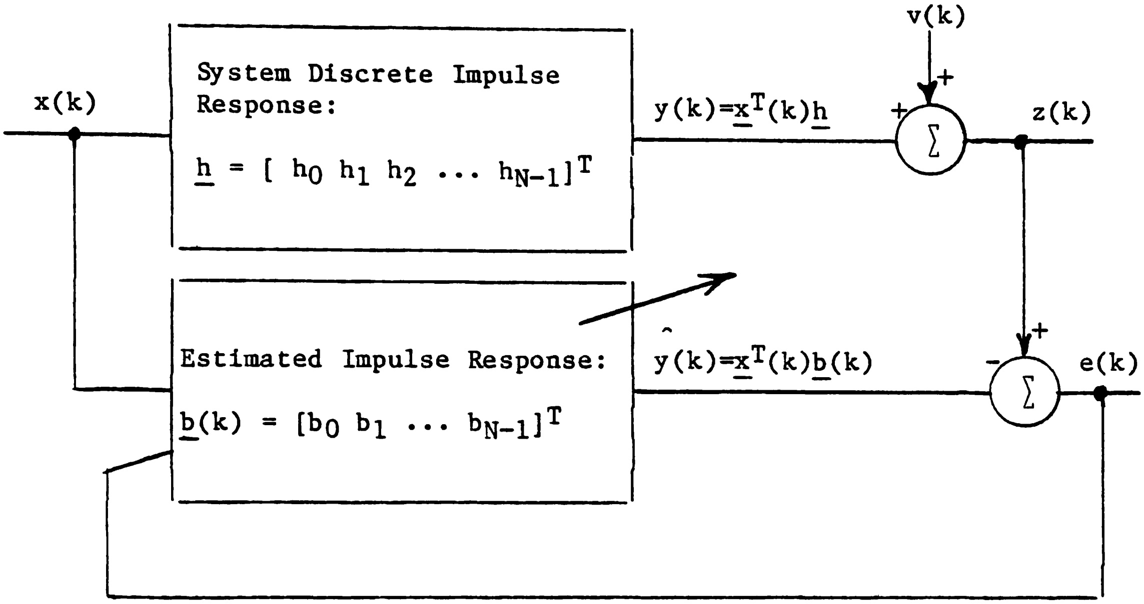

Figure 4. Block Diagram Illustrating System Identification Approach

4. Parameter Estimation Via Pole/Zero Modeling

In section 3, we modeled the system S to be identified by its impulse response. An equivalent model is to consider the system's transfer function H(z). From linear systems theory, we have

$$H(z) = \sum_{n=\infty}^{\infty} h(n) z^{-n} \tag{4.1}$$

where h(n) is the discrete impulse response. In general, we can define H(z) by the rational form

$$H(z) = \frac{c_0 + \sum_{i=1}^{M-1} c_i z^{-i}}{1 + \sum_{i=1}^{N} a_i z^{-i}} \tag{4.2}$$

Or, equivalently,

$$Y(z) + \sum_{i=1}^{N} a_i Y(z) z^{-i} = \sum_{i=0}^{M-1} c_i X(z) z^{-i} \tag{4.3}$$

since H(z) = Y(z)/X(z). Taking the inverse z-transform of (4.3) we get

$$y(k) = \sum_{i=0}^{M-1} c_i x(k-i) - \sum_{i=1}^{N} a_i y(k-i) \tag{4.4}$$

If we define the vector

$$\underline{b} = [c_0\ c_1\ \ldots\ c_{M-1}\ a_1\ a_2\ \ldots\ a_N]^T \tag{4.5}$$

and the signal vector as

$$\underline{x}(k) = \left| \begin{array}{l} x(k) \\ x(k-1) \\ \ldots \\ x(k-N+1) \\ -y(k-1) \\ -y(k-2) \\ \ldots \\ -y(k-N) \end{array} \right| \tag{4.6}$$

we can write (4.4) as

$$y(k) = \underline{x}^T(k)\underline{b} \quad . \tag{4.7}$$

Let the system measurement be

$$z(k) = \underline{x}^T(k)\underline{b} + v(k) \quad . \tag{4.8}$$

Then based upon our development of the Kalman estimation algorithm we can write the state estimation update equation as

$$\underline{b}(k) = \underline{b}(k-1) + \underline{K}(k) \; [z(k) - \underline{x}^T(k)\underline{b}(k-1)] \tag{4.9}$$

where

$$\underline{K}(k) = [\sum_{i=1}^{k} \underline{x}(i)\underline{x}^T(i)]^{-1} \; \underline{x}(k) \quad . \tag{4.10}$$

In (4.10), $\underline{x}(k)$ is given by (4.6) with z(k-i) replacing y(k-1), and

$$\hat{\underline{b}}(k) = [\hat{c}_0(k) \; \hat{c}_1(k) \; \ldots \hat{c}_{M-1}(k) \; \hat{a}_1(k) \; \ldots \hat{a}_N(k)]^T . \tag{4.11}$$

The above method estimates the parameter vector $\underline{b}$ and is in the class of parametric identification methods. The estimation of the impulse response falls in the catagory of nonparametric methods.

5. The Fast Kalman Algorithm

The Kalman gain for the system identification problem was shown to be

$$\underline{\underline{K}}(n) = [\sum_{j=0}^{n} \underline{x}(j)\underline{x}^T(j)]^{-1}\underline{x}(n) \tag{5.1}$$

The vector $\underline{x}(n+1)$ is obtained from $\underline{x}(n)$ by shifting in the new element $x(n+1)$ and shifting out $x(n-N)$. Using the shifting property, Ljung, Morf, and Falconer [2] developed a fast estimation scheme to calculate $\underline{K}(n)$. Per time update, their method requires 10N+4 multiplications compared with $3N^2+3N$ multiplications that are required in the conventional Kalman method. Furthermore, since the evaluation of an NxN matrix is avoided considerable saving in storage is achieved for large N. Two algorithms are presented below for the parametric and nonparametric estimation methods. The proofs to the algorithms can be found in [2].

Algorithm 1: Nonparametric Estimation

Let $\{\xi(n)\}$ be a sequence of p-vectors such that $\xi(j) = 0$ for $j<0$, and let

$$\underline{x}(n) = [\xi(n) \; \xi(n-1) \; \ldots \xi(n-N+1)]^T$$

then the quantity

$$\underline{k}(n) = [\sum_{j=0}^{n} \underline{x}(j)\underline{x}(j)^T+\delta I]^{-1}\underline{x}(n)$$

can be determined recursively as

$$\varepsilon_0(n) = \xi(n) + A^T(n-1)\underline{x}(n) \tag{5.2}$$

$$A(n) = A(n-1) - \underline{k}(n)\,\varepsilon_0{}^T(n) \tag{5.3}$$

$$\varepsilon(n) = \xi(n) + A^T(n)\underline{x}(n) \tag{5.4}$$

$$\Sigma(n) = \Sigma(n-1) + \varepsilon(n)\,\varepsilon_0^T(n) \tag{5.5}$$

$$\underline{k}'(n) = \begin{vmatrix} \Sigma^{-1}(n)\,\varepsilon(n) \\ \underline{k}(n) + A(n)\,\Sigma^{-1}(n)\,\varepsilon(n) \end{vmatrix} \tag{5.6}$$

Partition k'(n) as

$$\underline{k}(n) = \begin{vmatrix} \underline{m}(n) \\ \mu(n) \end{vmatrix} \begin{matrix} (Np)x1 \\ px1 \end{matrix}$$

Let

$$\eta_0(n) = \xi(n-N) + D^T(n-1)\underline{x}(n+1) \tag{5.7}$$

$$D(n) = [D(n-1) - \underline{m}(n)\,\eta_0^T(n)][I - \mu(n)\,\eta_0^T(n)]^{-1} \tag{5.8}$$

$$\underline{k}(n+1) = \underline{m}(n) - D(n)\,\mu(n) \tag{5.9}$$

The initial conditions can be taken as

$$\underline{k}(1) = 0,\ A(0) = 0,\ \Sigma(0) = \delta I,\ D(0) = 0. \tag{5.10}$$

In the above equations $\xi(n)$, $\varepsilon_0(n)$, $\varepsilon(n)$, $\mu(n)$, and $\eta_0(n)$ are px1 vectors. In the non parametric estimation problem applied to echo cancellation, p =1 and $\xi(n) = x(n)$.

Application to Parametric Estimation

In parametric estimation, the vector $\underline{x}^*(n)$ is obtained by shifting in the scalars x(n+1) and y(n) and shifting out x(n-N+1), y(n-N) where x(n) can be the input to a system and y(n) the output.

$$\underline{x}^*(n) = [x(n)\ x(n-1)\ \ldots\ x(n-N+1)\ y(n-1)\ y(n-2)\ \ldots\ y(n-N)]^T$$

Let

$$\underline{x}(n) = [x(n)\ y(n-1)\ x(n-1)\ y(n-2)\ \ldots\ x(n-N+1)\ y(n-N)]^T$$

then

$$\underline{x}^*(n) = \Psi\underline{x}(n) \tag{5.11}$$

where Ψ is a permutation matrix.

In this case $\xi(n) = [x(n)\ y(n-1)]^T$. Also,

$$\underline{k}^*(n) = [\sum_{j=0}^{n} \underline{x}^*(j)\underline{x}^*(j)]^{-1}\underline{x}^*(n) = [\Psi \sum_{j=0}^{n} \underline{x}(j)\underline{x}^T(j)\Psi^T]^{-1}\ \Psi\underline{x}(n)$$

$$\underline{k}^*(n) = \Psi^{-T}[\sum_{j=0}^{n} \underline{x}(j)\underline{x}^T(j)]^{-1}\underline{x}(n) = \Psi^{-T}\underline{k}(n)\ . \tag{5.12}$$

Consequently, if $\underline{x}^*(n)$ is obtained as a linear map of $\underline{x}(n)$, where $\underline{x}(n)$ has the shifting property described in the algorithm, then the 'gain matrix' $\underline{k}(n)$ associated with $\underline{x}^*(n)$ can be obtained by the algorithm through equation (5.12). This leads to a general algorithm presented below.

<u>Algorithm 2</u>: <u>Parametric Estimation</u>

Let $\{\underline{x}(n)\}$ be a sequence of N-vectors, such that its components are shifted in some fashion as n increases. Suppose that p new elements enter $\underline{x}(n+1)$; collect these in the vector $\xi(n)$. At the same time p elements leave $\underline{x}(n+1)$ (i.e. they belong to x(n) but not to x(n+1)). Collect these in the vector $\rho(n)$. To describe the way in which the elements enter and leave, let $\underline{x}'(n)$ be an extended (N+p) vector containing the elements of $\underline{x}(n)$ as well as those of $\xi(n)$. Assume that

$$\Psi_F'\underline{x}'(n) = \begin{vmatrix} \xi(n) \\ \underline{x}(n) \end{vmatrix} \tag{5.13}$$

$$\Psi_B'\underline{x}'(n) = \begin{vmatrix} \underline{x}(n+1) \\ \rho(n) \end{vmatrix} \tag{5.14}$$

where Ψ_F' and Ψ_B' are permutation matrices (or in fact any orthonormal matrices). Assume that $\xi(j) = 0,\ j<0$.

Then the quantity

$$\underline{k}(n) = [\sum_{j=1}^{n}\underline{x}(j)\underline{x}^T(j) + \delta I]^{-1}\underline{x}(n) \quad (5.15)$$

can be determined recursively as

$$\varepsilon_0(n) = \xi(n)+A^T(n-1)\underline{x}(n) \quad (5.16)$$

$$A(n) = A(n-1) - \underline{k}(n)\,\varepsilon_0^T(n) \quad (5.17)$$

$$\varepsilon(n) = \xi(n) + A^T(n)\underline{x}(n) \quad (5.18)$$

$$\Sigma(n) = \Sigma(n-1) + \varepsilon(n)\,\varepsilon_0^T(n) \quad (5.19)$$

$$\underline{k}'(n) = \Psi_F{}'^T \begin{vmatrix} \Sigma^{-1}(n)\,\varepsilon(n) \\ \underline{k}(n) + A(n)\Sigma^{-1}(n)\,\varepsilon(n) \end{vmatrix} . \quad (5.20)$$

Partition $\Psi_B{}'\underline{k}'(n)$ as

$$\Psi_B{}'\underline{k}'(n) = \begin{vmatrix} \underline{m}(n) \\ \mu(n) \end{vmatrix} \begin{matrix} N\text{x}1 \\ p\text{x}1 \end{matrix} \quad (5.21)$$

$$\eta_0(n) = \rho(n) + D^T(n-1)\underline{x}(n+1) \quad (5.22)$$

$$D(n) = [D(n-1) - \underline{m}(n)\,\eta_0^T(n)][I-\mu(n)\,\eta_0^T(n)]^{-1} \quad (5.23)$$

$$\underline{k}(n+1) = \underline{m}(n) - D(n)\,\mu(n) \quad (5.24)$$

The initial conditions may be taken as

$$\underline{k}(1) = \underline{0},\ A(0) = 0,\ \Sigma(0) = \delta I,\ D(0) = 0 \quad . \quad (5.25)$$

In the echo cancellation problem,

$$\xi(n) = [x(n)\ y(n-1)]^T$$

and

$$\rho(n) = [x(n-N+1)\ y(n-N)]^T$$

where x(n) is the speech input sample and y(n) is the echo sample.

In the parametric fast kalman estimation algorithm, p=2, and 20N+22 multiplications are required compared to $3N^2+3N$ for the conventional method.

6. System Identification Simulation Results

In this section the results of the simulation of parametric and nonparametric Fast Kalman algorithms to simple and complex system's identification will be presented.

6.1. Gradient Search Versus Kalman Estimation

In the first set of simulations, the system consisted of a 9-tap FIR digital filter. The filter coefficients are presented in Table 1. The input of the filter was a white gaussian signal with unity power. Figures 5 and 6 show the estimate error in dB's versus the iteration number for the conventional gradient search algorithm and the Kalman algorithm respectively. The superiority of the Kalman algorithm over the gradient search algorithm is immediately evident from the plots. The Kalman method converges to an estimate error 60 dB below the actual system output within 12 iterations. The gradient search method, on the other hand, achieves this error after 200 iterations. Furthermore, the Kalman algorithm basically converges within 10 taps, approximately the vector length N=9 as predicted theoretically in section 3. In chapter 4 we show that the gradient method converges at a constant rate for white guassian signals as demonstrated by the simulation result in figure 5. Table 1 also shows the estimated tap coefficients of the Kalman method after 100 iterations.

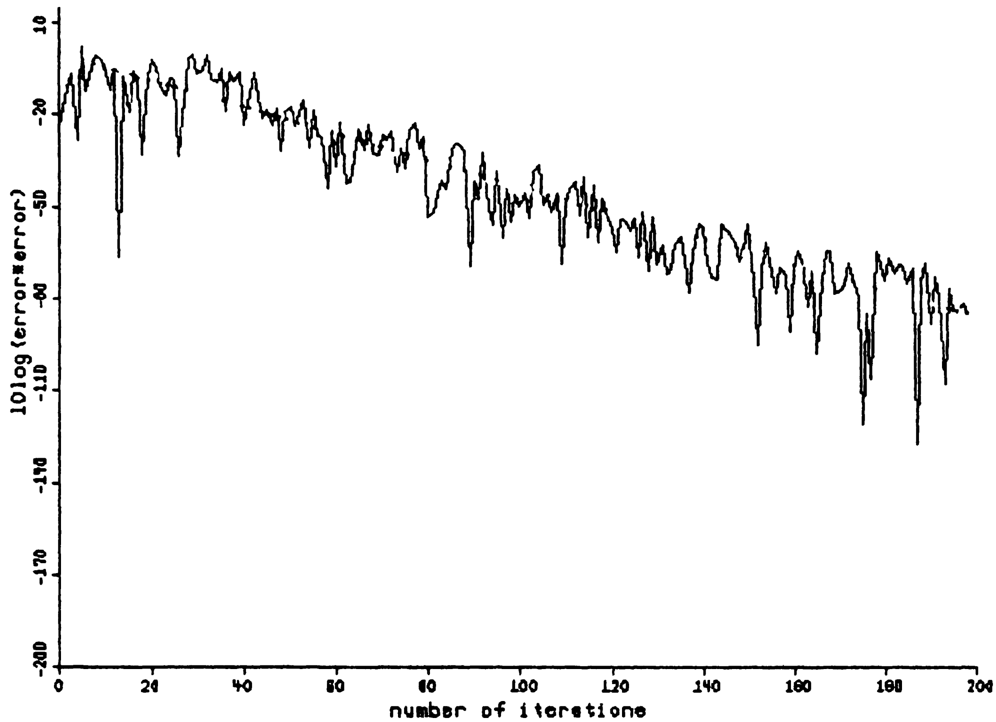

Figure 5. Gradient Search Algorithm, 9 Taps

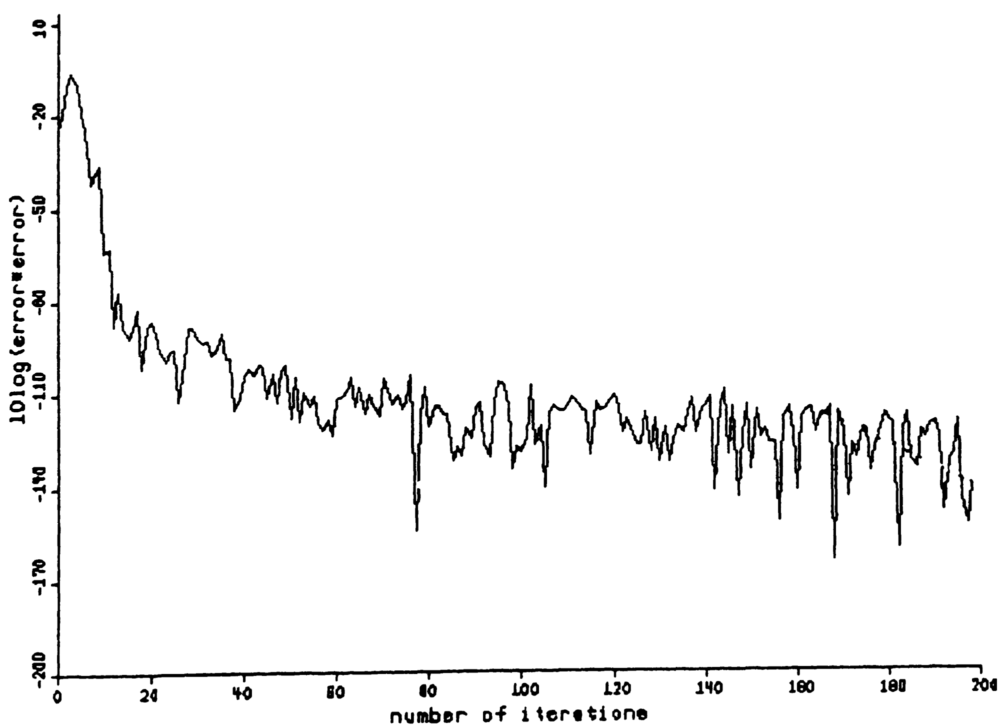

Figure 6. Kalman Algorithm, 9 Taps

Table 1. FIR Filter Coefficients and Estimated Coefficients By Kalman Algorithm

n	FIR COEFFICIENTS $h(n)$	ESTIMATED IMPULSE RESPONSE $\hat{h}(n)$
0	0.127324	0.1273236
1	-0.212207	-0.2122065
2	0.636620	0.6366192
3	1.000000	0.9999859
4	0.636620	0.6366192
5	-0.212207	-0.2122066
6	0.127324	0.1273237
7	0.	0.00000008
8	0.	0.00000016

6.2. Parametric Versus Nonparametric Estimation

The convergence of the parametric and the nonparametric Fast Klaman algorithm are compared in figure 7 for an infinite impulse response system. The system was a 32 kHz voiceband bandpass digital IIR filter. From figure 7, the steady state error of the parametric method was approximately 130 dB below the filter output compared to 30dB for the nonparametric method. This follows from the fact that for the non-parametric case the impulse response is truncated at 150 samples. In contrast the parametric method models the finite order filter exactly. Furthermore, the paramteric method converges within 20 iterations, i.e., M+N where N is the pole order (10) and M-1 is the zero order (9). The filter consisted of a 6th order lowpass elliptic filter in cascade with a 4th order Chebyshev highpass filter. The significant improvement in the convergence rate of the parametric method over the nonparametric method is quite clear from the simulation (figure 7). For a bandlimited system it is possible to reduce the sampling rate and, therefore, the number of significant discrete impulse response samples. Hence, the poewrformance of the parametric method can be improved. However, since sampling rate reduction does not change the order of a system as long as the Nyquist criterion is maintained, the parametric method basically converges within the same number of iterations (N+M).

In order to test this assumption the 1024-point impulse response of the 32 kHz digital filter was decimated to a 256-point 8 kHz impulse response. These samples were then used as coefficients of an FIR digital filter with a white guassian input signal. The parametric Fast Kalman algorithm was used with the same order (M=10, N=10) to estimate the transfer function numerator and denominator coefficients from which the poles and zeroes were derived. The algorithm again converged within 20 iterations. Table 2 shows the estimated and predicted polar coordinates of the decimated bandpass filter poles in the Z-plane. The correspondence between the estimated and predicted polar coordinates is very close.

Table 2. Estimated and Predicted Z-plane Poles, Decimated Bandpass IIR Filter Impulse Response

Estimated Pole Coordinates		Predicted Pole Coordinates f_s/f_s'	
$\hat{\rho}'$	$\hat{\theta}'$	$\rho'=\rho$	$\theta'=\theta*f_s/f_s'$
0.7220	153.96	0.7218	153.96
0.5381	143.05	0.5362	143.121
0.1968	72.76	0.1975	73.72
0.9626	7.37	0.9693	7.6
0.8307	6.28	0.8514	6.72

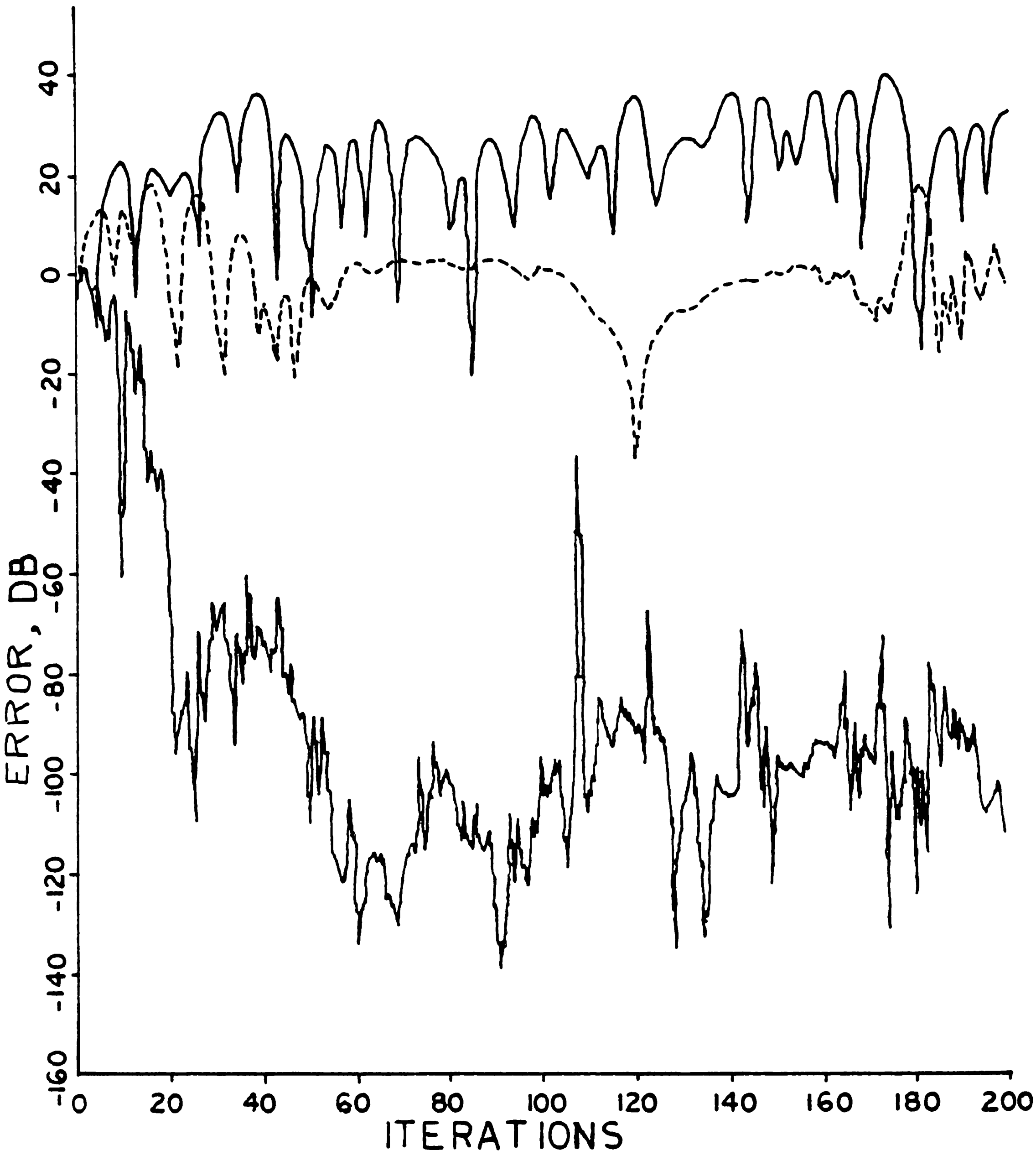

Figure 7. Comparison of Convergence Between Parametric and Nonparametric Fast Kalman Estimation. The Signal to be Estimated is the Top Curve.

6.3. Parametric Estimation Pole/Zero Distribution

The order of the poles, N, and the zeroes, M-1, is seldom known apriori for a given system. In this section the effects of order overspecification and underspecification will be studied for the parametric Fast Kalman algorithm. The system studied consists of the bandpass IIR digital filter discussed before with the Z-plane pole/zero distribution shown in figure 8. The poles and zeroes clustered around the real axis at the unity circle correspond to the highpass Chebyshev IIR section and the other poles and zeroes correspond to the lowpass elliptic IIR section.

The estimated pole/zero distribution for the parametric estimation method with overspecified pole and zero order (M=14, N=14) appears in figure 9. We observe that the extra poles were cancelled out by the extra zeroes. In the case of real poles the cancellation was exact. The performance of the algorithm was similar to the case where the order specification was exact (figure 7.).

Figure 10. shows the results obtained when the zero order is overspecified by 10 (M=20, N=10). In this case the extra zeroes were distributed in a circular fashion around the origin. The lowpass complex zeroes in the left hand plane were moved closer to the imaginary axis. The overall effect is, however, to obtain the same overall response. This observation is of significance in the application of parametric estimation to echo cancellation since the echo path has an unknown pure delay component which can be modelled as zeroes at the

Z-plane origin. The extra zeroes specified can account for the unknown delay component in the echo path.

In figure 11 the results of underspecifying the zero order is presented. In this case the algorithm made an approximate estimate of the lowpass poles but did not estimate the highpass poles and zeroes. This phenomena is reflected in the estimated impulse response shown in figure 11(b). The impulse response exhibits a purely lowpass characteristic. Nevertheless, the estimate error in this case was 50dB below the filter output.

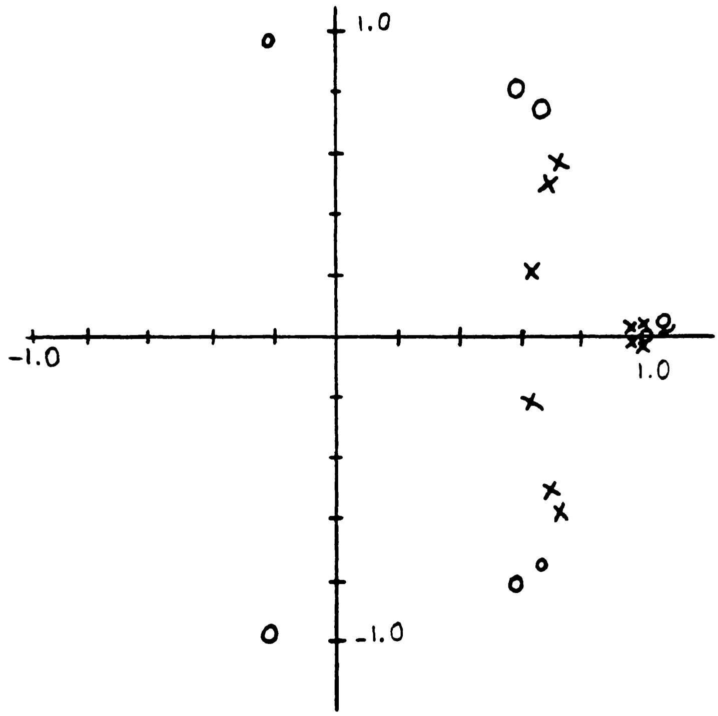

Figure 8. Pole/Zero Distribution IIR Digital Bandpass Filter (Voiceband).

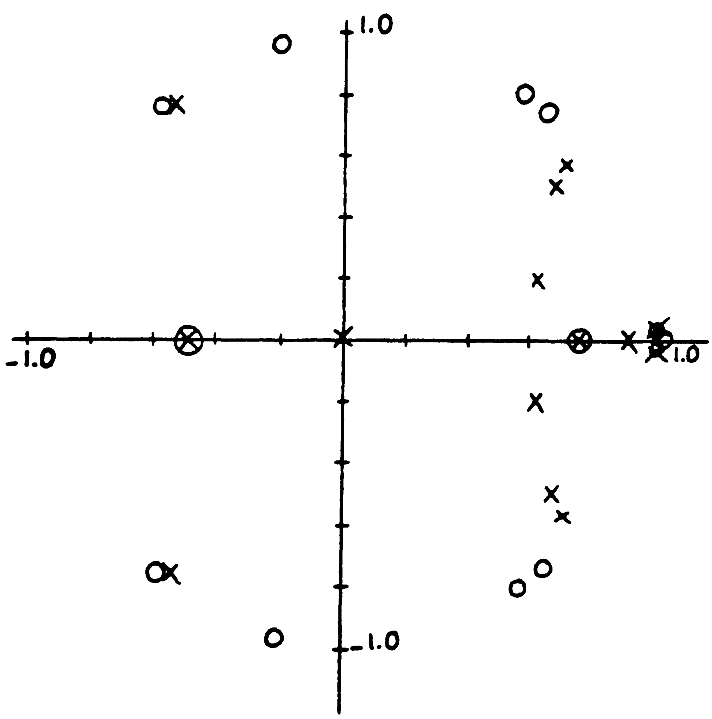

Figure 9. Pole/Zero Distribution Parametric Fast Kalman Estimation, M=14, N=14.

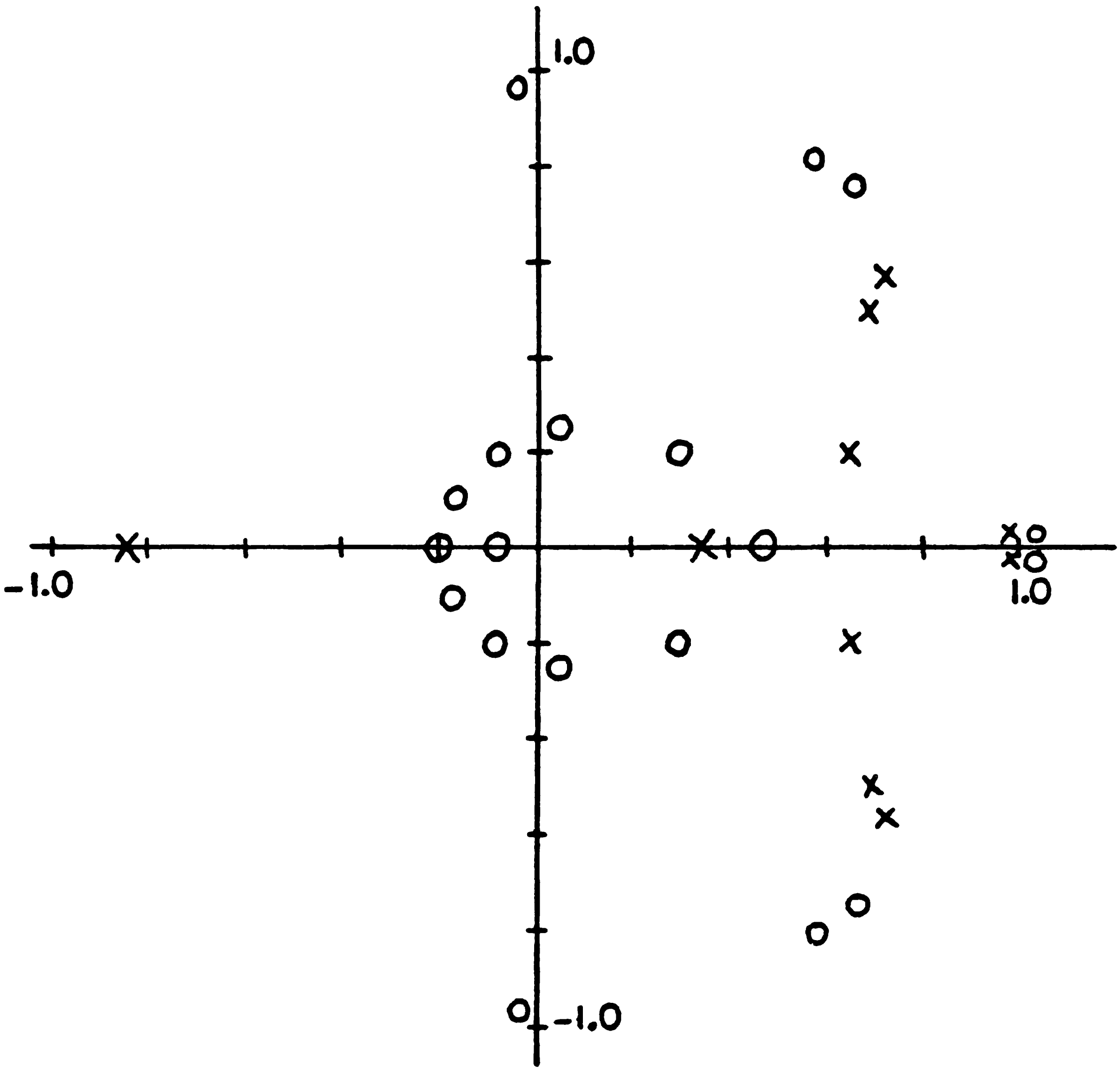

Figure 10. Pole/Zero Distribution Parametric Fast Kalman Estimation, M=20, N=10.

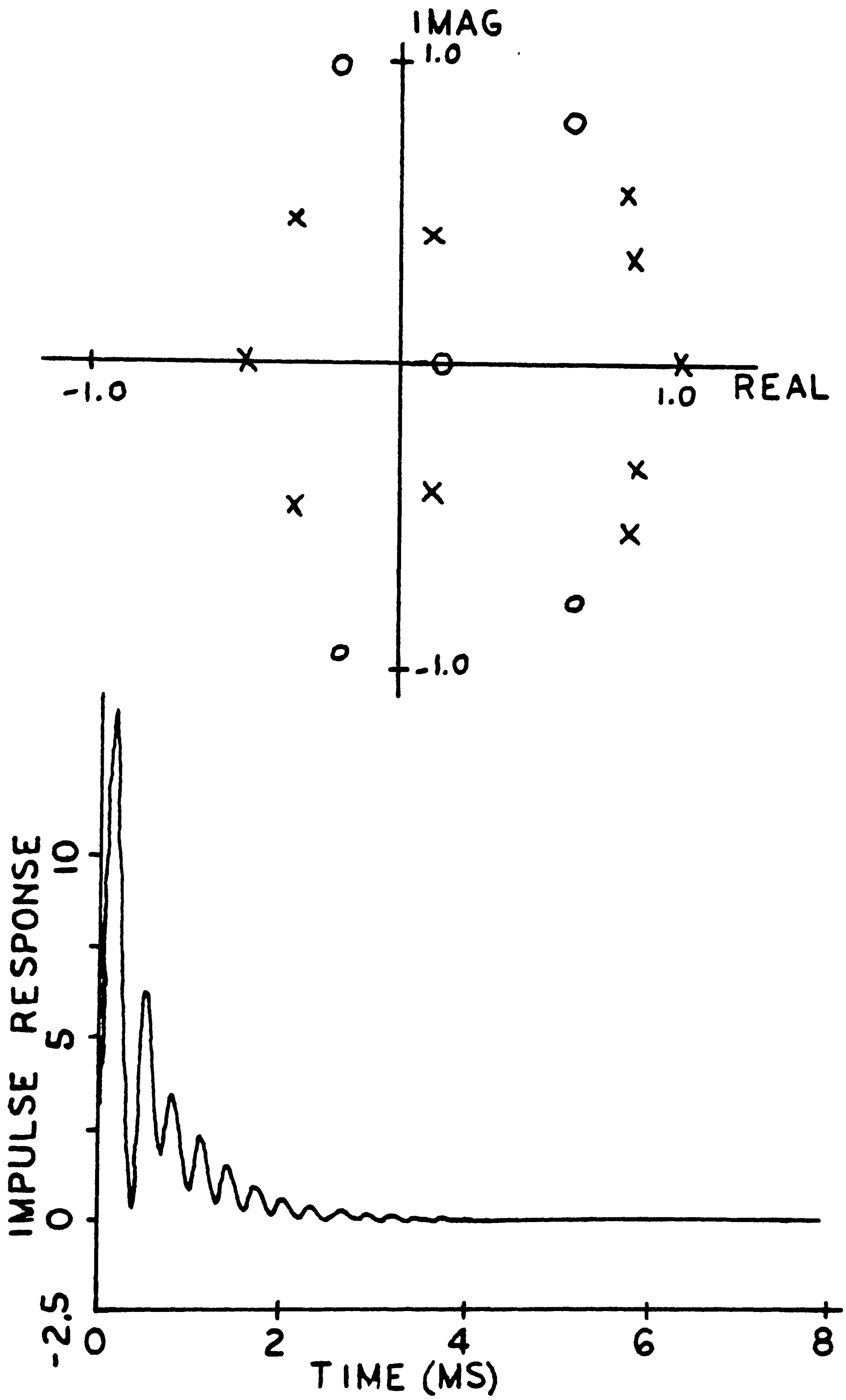

Figure 11. (a) Pole/Zero Distribution Parametric Fast Kalman Estimation, M=6, N=10.
(b) Impulse Response.

APPENDIX A

Derivation of the Optimum Estimate

We have,

$$\zeta(k) = E[e^2(k)] = E\{[y(k)-\underline{x}^T(k)\underline{b}(k)]^T[y(k)-\underline{x}^T(k)\underline{b}(k)]\} \quad . \qquad \text{(A.1)}$$

$$\zeta(k) = E\{y^2(k) - 2\underline{x}^T(k)\underline{b}(k)y(k) + [\underline{x}^T(k)\underline{b}(k)]^2 \} \quad . \qquad \text{(A.2)}$$

Taking the partial derivitave with respect to the vector $\underline{b}(k)$,

$$\frac{\partial\zeta(k)}{\partial\underline{b}(k)} = - \frac{\partial}{\partial\underline{b}(k)} E\{ 2\underline{x}^T(k)\underline{b}(k)y(k)\} + \frac{\partial}{\partial\underline{b}(k)} E\{[\underline{x}^T(k)\underline{b}(k)]^2\} \qquad \text{(A.3)}$$

Now,

$$[\underline{x}^T(k)\underline{b}(k)]^2 = [\underline{x}^T(k)\underline{b}(k)]^T[\underline{x}^T(k)\underline{b}(k)] = \underline{b}^T(k)\underline{x}(k)\underline{x}^T(k)\underline{b}(k).$$

Let $X(k) = \underline{x}(k)\underline{x}^T(k)$. Then,

$$E\{ \frac{\partial}{\partial\underline{b}(k)}[\underline{b}^T(k)X(k)\underline{b}(k)] \} = E\{ X(k)\underline{b}(k) + X^T(k)\underline{b}(k)\} \quad . \qquad \text{(A.4)}$$

Since $X(k) = X^T(k)$, (B.4) becomes,

$$2E\{ X(k)\underline{b}(k) \} \quad . \qquad \text{(A.5)}$$

Also,

$$2E\{ \frac{\partial}{\partial\underline{b}(k)} \underline{x}^T(k)\underline{b}(k)y(k) \} = 2E\{ \underline{x}(k)y(k) \} \quad . \qquad \text{(A.6)}$$

Substituting (A.5) and (A.6) into (A.3) and setting the derivitive to zero we obtain,

$$-2E\{ \underline{x}(k)y(k) \} + 2E\{ X(k) \}\underline{b}_o(k) = 0 \quad . \qquad \text{(A.7)}$$

Or,

$$\underline{b}_o(k) = A^{-1} \cdot \underline{c} \qquad \text{(A.8)}$$

where $A = E\{ \underline{x}(k)\underline{x}^T(k) \}$ and $\underline{c} = E\{ \underline{x}(k)y(k) \}$.

APPENDIX B

Orthogonality of the State Estimate and State Estimate Error [3]

We will show by induction that

$$E\{\ \underline{b}^T(k)\underline{\varepsilon}(k)\ \} = 0 \tag{B.1}$$

We start by calculating $\underline{b}(1)$,

$$\underline{b}(1) = \underline{b}(0) + \underline{K}(1)[\underline{x}^T(1)\underline{h} + v(1) - \underline{x}^T(1)\underline{b}(0)] \tag{B.2}$$

where $\underline{h}$ is the true state and we have substituted $z(1) = \underline{x}^T(1)\underline{h} + v(1)$. Subtracting $\underline{h}$ from both sides we obtain,

$$\underline{\varepsilon}(1) = \underline{b}(0)-\underline{h}+\underline{K}(1)\underline{x}^T(1)\underline{h}+\underline{K}(1)v(1)-\underline{K}(1)\underline{x}^T(1)\underline{b}(0)$$

$$\underline{\varepsilon}(1) = [I-\underline{K}(1)\underline{x}^T(1)]\underline{\varepsilon}(0) + \underline{K}(1)v(1) \tag{B.3}$$

Multiplying (B.2) by (B.3) we obtain

$$\begin{aligned}\underline{b}(1)\underline{\varepsilon}^T(1) &= \underline{b}(0)\underline{\varepsilon}^T(0)[I-\underline{K}(1)\underline{x}^T(1)]^T+\underline{b}(0)\underline{K}^T(0)\\ &-\underline{K}(1)\underline{x}^T(1)\underline{\varepsilon}(0)\underline{\varepsilon}^T(0)[I-\underline{K}(1)\underline{x}^T(1)]^T-\underline{K}(1)\underline{x}^T(1)\underline{\varepsilon}(0)\underline{K}^T(1)v(1)\\ &+\underline{K}(1)v(1)\underline{\varepsilon}^T(0)[I-\underline{K}(1)\underline{x}^T(1)]^T+\underline{K}(1)\underline{\underline{K}}^T(1)v(1)\end{aligned} \tag{B.4}$$

Simplifying and taking the expectation we obtain

$$E\{\underline{b}(1)\underline{\varepsilon}^T(1)\} = -\underline{K}(1)\underline{x}^T(1)P(0)[I-\underline{x}(1)\underline{K}^T(1)]+\underline{K}(1)\underline{K}^T(1)r(1) \tag{B.5}$$

where we have used the fact that

$$E\{v(1)\underline{\varepsilon}^T(0)\} = 0$$

and

$$E\{\underline{b}(0)\underline{\varepsilon}^T(0)\} = 0$$

since $v(1)$ is uncorrelated with $\underline{b}(0)$ and the initial vector $\underline{b}(0)$ is random. We can rewrite (B.5) as

$$E\{\underline{b}(1)\underline{\varepsilon}^T(1)\} = \underline{K}(1)\{\ -\underline{x}^T(1)P(0) + [\underline{x}^T(1)P(0)\underline{x}(1)+r(1)]\underline{K}^T(1)\ \} \tag{B.6}$$

However,

$$\underline{K}(1) = P(0)\underline{x}(1)[\underline{x}^T(1)P(0)\underline{x}(1) + r(1)]^{-1} . \qquad (B.7)$$

Substituting into (C.6) for $\underline{K}(1)$ in the brackets we obtain

$$E\{\underline{b}(1)\underline{\varepsilon}^T(1)\} = \underline{K}(1)[-\underline{x}^T(1)P(0) + \underline{x}^T(1)P(0)] = 0 .$$

We can then show that

$$E\{\underline{b}(2)\underline{\varepsilon}^T(2)\} = 0 .$$

Or by induction

$$E\{\underline{b}(k)\underline{\varepsilon}^T(k)\} = 0 .$$