AN ITERATIVE QUADRATIC TECHNIQUE FOR DESIGNING MINIMUM PHASE FILTERS

James L. Rasmussen and Delores M. Etter
Department of Electrical and Computer Engineering
University of Colorado
Boulder, CO 80309

Abstract

An adaptive convex filter (ACF) consists of a parallel bank of three fixed second-order filters. The output of each fixed filter is multiplied by a constrained weight. The fixed filters are chosen such that the ACF will always be minimum phase. Higher order filters are implemented by cascading ACF stages but this cascaded system does not have aquadratic error surface when updated using traditional algorithms. We present an algorithm that has a quadratic error surface for the ACF and converges to the global minimum.

1 INTRODUCTION

A convex filter (CF) [1-3] is composed of a bank of parallel fixed filters. Each fixed filter has a weight associated with it that determines the contribution of that filter to the output. If the weights vary with time, then the structure is an adaptive convex filter (ACF). In this paper, we use three fixed, second-order filters in the ACF as shown in Figure 1. Each weight functions as a weight or a gain with a value between zero and one. The sum of the weights is constrained to equal one. We call this a "convex filter" because the equivalent filter expressed in (1) can be written as two convex sums [6] of the fixed filters.

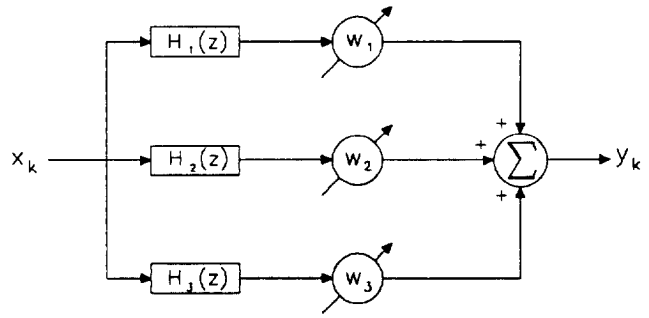

Figure 1 Adaptive Convex Filter Bank

The fixed filters in a ACF can be composed of FIR or IIR filters; in this paper we assume they are second-order FIR filters. (Without loss of generality, we assume that the general form of each of the FIR filters is $H_k(z) = 1 + a_{1,k}z^{-1} + a_{2,k}z^{-2}$).

The ACF is represented by an FIR filter of the same order as the fixed filters, with coefficients computed as expected values (or weighted averages) of the corresponding coefficients in the ACF. Therefore, we represent the ACF by the following average filter representation:

$$\begin{aligned} H_\mu(z) &= \sum_{k=1}^{3} H_k(z)\, w_k \\ &= 1 + (w_1 a_{1,1} + w_2 a_{1,2} + w_3 a_{1,3})\, z^{-1} + (w_1 a_{2,1} + w_2 a_{2,2} + w_3 a_{2,3})\, z^{-2} \\ &= 1 + a_{1\mu} z^{-1} + a_{2\mu} z^{-2} \end{aligned} \quad (1)$$

$$\text{where } a_{k\mu} = \sum_{k=1}^{3} (w_k a_{n,k})$$

and w_k represents the weight associated with $H_k(z)$.

We now define the matrix, **H**, whose columns are the coefficients of the fixed filters of the ACF.

$$\boldsymbol{H} = \begin{bmatrix} 1 & 1 & 1 \\ a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \end{bmatrix} = \begin{bmatrix} \boldsymbol{h}_1 & \boldsymbol{h}_2 & \boldsymbol{h}_3 \end{bmatrix}.$$

Note that $\mathbf{h}_k$ is also the impulse response of the kth filter. We also represent the weights and the impulse response of $H_\mu(z)$ as vectors and then the average filter can be conveniently represented in matrix notation as $\mathbf{h}_\mu = \mathbf{Hw}$.

2 ADAPTIVE CONVEX FILTER (ACF)

Any number of stages shown in Figure 1 can be cascaded to construct a higher order ACF as illustrated in Figure 2. Each stage has the same properties as discussed above. If we allow all weights in all stages to vary, then the error is a function of all stages. The error surface computed in this way is not quadratic and becomes more complex as the number of ACF stages in the structure increases. However, if we let only one stage vary and fix the weights of all other stages, then the error surface is quadratic. We examine the non-quadratic case first.

The output, y_k, of an n stage ACF in Figure 2 can be

1058-6393/95 $4.00 © 1995 IEEE

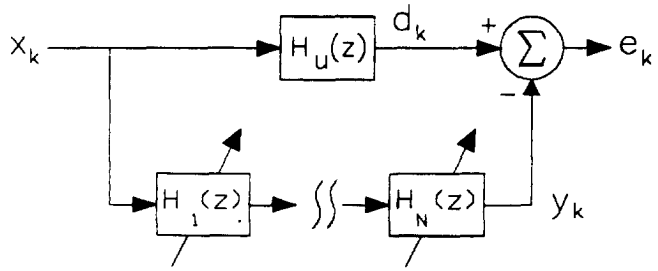

Figure 2. General 2N order ACF used in the system identification configuration.

expressed as

$$y_k = \mathbf{x}_k^T \mathbf{A}_1 \mathbf{A}_2 \cdots \mathbf{A}_n. \tag{2}$$

where $\mathbf{x}_k \in \mathbb{R}^{(2n+1)\times 1}$ ($\mathbf{x}_k$ is a 2n+1 by 1 vector) and the $\mathbf{A}_i$ are the convolution matrices containing the coefficients of $\mathbf{h}_i$. Therefore, the convolution matrices, $\mathbf{A}_i$, are dependent on the weights of the i^{th} stage of the ACF.

$$\mathbf{A}_i^T = \begin{bmatrix} 1 & a_{1i} & a_{2i} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & a_{1i} & a_{2i} \end{bmatrix} \in \mathbb{R}^{(2(n-i)+1)\times(2(n-i)+3)}$$

The error surface is the expected value of the squared error and can now be expressed as

$$\begin{aligned} J &= E[e_k^2] = E[d_k^2] + E[y_k^2] - 2E[d_k y_k] \\ J &= \mathbf{h}_u^T \mathbf{R}_{uu} \mathbf{h}_u + \mathbf{w}_N^T \mathbf{H}_N^T \mathbf{A}_{N-1} \cdots \mathbf{A}_1^T \mathbf{R}_{aa} \mathbf{A}_1 \cdots \mathbf{A}_{N-1} \mathbf{H}_N \mathbf{w}_N \\ &\quad - 2\mathbf{h}_u^T \mathbf{R}_{ua} \mathbf{A}_1 \cdots \mathbf{A}_{N-1} \mathbf{H}_N \mathbf{w}_N \end{aligned} \tag{3}$$

where $\mathbf{R}_{aa}$ and $\mathbf{R}_{uu}$ are the input correlation matrices of the input vectors, $\mathbf{x}_a$ and $\mathbf{x}_u$, of $H_a(z)$ and $H_u(z)$ respectively. $H_a(z)$ is the equivalent adaptive filter $H_1(z)H_2(z)\cdots H_N(z)$. We have distinguished the input vectors explicitly since the order of the unknown filter and the order of the equivalent ACF may be different, thus the length of the input vectors will not necessarily be the same. With some simplifications we can reduce (3) to the more familiar form for the error surface used in Widrow and Stearns [4]. Note that $\mathbf{A}_N$ is a 3x1 vector and is identical to $\mathbf{h}_N$.

$$\begin{aligned} J &= \mathbf{h}_u^T \mathbf{R}_{uu} \mathbf{h}_u + \mathbf{w}_N^T \tilde{\mathbf{R}}_N \mathbf{w}_N - 2\,\mathbf{w}_N^T \tilde{\mathbf{p}}_N \\ \text{where } \tilde{\mathbf{R}}_N &= \mathbf{H}_N^T \mathbf{A}_{N-1}^T \cdots \mathbf{A}_1^T \mathbf{R}_{aa} \mathbf{A}_1 \cdots \mathbf{A}_{N-1} \mathbf{H}_N ; \quad \tilde{\mathbf{R}}_N \in \mathbb{R}^{3\times 3} \\ \text{and } \tilde{\mathbf{p}}_N &= \mathbf{H}_N^T \mathbf{A}_{N-1}^T \cdots \mathbf{A}_1^T \mathbf{R}_{au} \mathbf{h}_u ; \quad \tilde{\mathbf{p}}_N \in \mathbb{R}^{3\times 1} \end{aligned} \tag{4}$$

$\tilde{\mathbf{R}}_N$ is the modified input correlation matrix resulting from passing white gaussian noise through all the stages. Similarly, $\tilde{\mathbf{p}}_N$ is the modified cross correlation vector. The product of these stages is responsible for the non-quadratic nature of the error surface.

3 QUADRATIC ACF (QACF)

A disadvantage of the error surface for the structure shown in Figure 2 is that it is not a quadratic function of the weights. This results in the possibility of simple algorithms not converging to the global minimum. However, if only one stage is treated as a variable and all the other stages are fixed, then the error surface would be a function of just one stage and would be quadratic.

This quadratic ACF structure is shown in Figure 3. The following expressions for the error and error surface are produced.

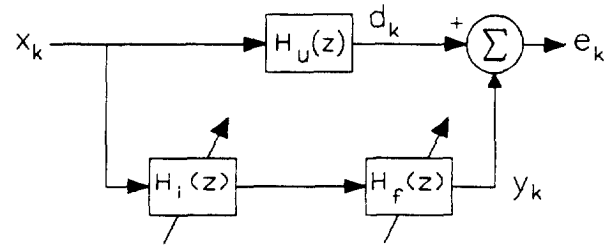

Figure 3 Quadratic structure for obtaining higher order quadratic ACFs.

$$\begin{aligned} J &= E[e_k^2] = E[d_k^2] + E[y_k^2] - 2E[d_k y_k] \\ J &= \mathbf{h}_u^T \mathbf{R}_{uu} \mathbf{h}_u + \mathbf{w}_i^T \mathbf{H}_i^T \mathbf{A}_f^T \mathbf{R}_{aa} \mathbf{A}_f \mathbf{H}_i \mathbf{w}_i - 2\mathbf{h}_u^T \mathbf{R}_{ua} \mathbf{A}_f \mathbf{H}_i \mathbf{w}_i \\ J &= \mathbf{h}_u^T \mathbf{R}_{uu} \mathbf{h}_u + \mathbf{w}_i^T \tilde{\mathbf{R}}_i \mathbf{w}_i - 2\,\mathbf{w}_i^T \tilde{\mathbf{p}}_i \end{aligned} \tag{5}$$

There are only two convolution matrices for this algorithm; $\mathbf{A}_f^T \in \mathbb{R}^{(2(n-1)+1)\times(2(n-1)+3}$ and $\mathbf{A}_i^T \in \mathbb{R}^{1\times 3}$ as in (3.2). Recall that $\mathbf{A}_i$ is actually $\mathbf{h}_i$, the coefficient vector of the i^{th} stage of the ACF.

The theoretical solution for $\mathbf{w}_i$ can now easily be obtained by setting the gradient equal to zero and solving [5]. The gradient for (5) is

$$\frac{\partial J}{\partial \mathbf{w}_i} = 2\mathbf{H}_i^T \mathbf{A}_f^T \mathbf{R}_{aa} \mathbf{A}_f \mathbf{H}_i \mathbf{w}_i - 2\mathbf{H}_i^T \mathbf{A}_f^T \mathbf{R}_{ua} \mathbf{h}_u \tag{6}$$

and thus the ideal solution for the weights, $\mathbf{w}_i$, of the adaptive filter stage is

$$\begin{aligned} \mathbf{w}_i^* &= (\mathbf{A}_f^T \mathbf{R}_{aa} \mathbf{A}_f)^{-1} \mathbf{A}_f^T \mathbf{R}_{ua} \mathbf{h}_u \\ &= \mathbf{H}^{-1} (\mathbf{A}_f^T \mathbf{A}_f)^{-1} \mathbf{A}_f^T \mathbf{h}_u ; \quad \text{if } \mathbf{x}_k \text{ is UVWGN} \end{aligned} \tag{7}$$

Implementation of the LMS algorithm is relatively easy since we are in effect just filtering the input data with a big filter, $\mathbf{h}_f$, and then only adapting a small filter, $\mathbf{h}_i$. The update to the weights at each iteration is easily found to be

$$\begin{aligned} \frac{\partial e}{\partial \mathbf{w}_i} &= -\mathbf{H}_i^T \mathbf{A}_f^T \mathbf{x}_k \\ \text{and } \mathbf{w}_i(k+1) &= \mathbf{w}_i(k) + 2\mu e_k \mathbf{H}^T \mathbf{A}_f^T \mathbf{x}_k \end{aligned} \tag{8}$$

The QACF retains the quadratic nature of the error surface at the expense of updating only one ACF at a time while fixing all others. The consequences of this are discussed in the next section.

3.1 Convergence of QACF

If the QACF can be shown to converge in a reasonable number of cycles, then we would have a filter that is not only guaranteed to converge to the global minimum because of the quadratic error surface, but would also be guaranteed to be stable since the roots of each ACF are guaranteed to remain inside the unit circle at each iteration. In addition, we have the added benefit of adapting only a second order filter during each cycle. This decreases any misadjustment in the adaptive process. Similarly, the order of the fixed filter has been decreased from the ordinary ACF structure.

Assume that we have an Nth order QACF and an Nth order minimum phase unknown system. Therefore, all roots of the unknown will be inside the unit circle. Theoretically, the QACF should be able to model the unknown exactly and thus give zero error. If the ACF stages are started with arbitrary roots, then the ACF can be represented by an equivalent initial filter, $H_{eq,i}(z) = \mathbf{h}_{eq,i}^T \psi(z)$. The initial error between the output of the unknown and the output of the initial ACF is

$$\begin{aligned} J &= h_u^T h_u + h_{eq,i}^T h_{eq,i} - 2 h_u^T h_u \\ &= (h_u - h_{eq,i})^T (h_u - h_{eq,i}) \end{aligned} \tag{9}$$

Since the error surface is quadratic, the adaptive process will drive the error to its smallest value possible plus some misadjustment. Substitution of the ideal weights given in (7) into (5) gives the minimum error.

$$\begin{aligned} J_{min} &= h_u^T h_u - h_u^T A_f (A_f^T A_f)^{-1} A_f^T h_u \\ &= h_u^T (I - P) h_u \end{aligned} \tag{10}$$

where **P** and (**I** - **P**) are both projection matrices.

After stage one converges, the actual error is equal the error given in (10) plus some misadjustment. If we now look at the system where the weights of stage two are variable and the weights of all other stages are fixed, there will obviously be a different error surface. However, the final $\mathbf{h}_{eq,1}$ is equal to the beginning $\mathbf{h}_{eq,2}$. Therefore, in accordance with (9), the ending point on the error surface for iteration one is the same as the starting point for iteration two. Since the error surface for iteration two is also quadratic, the error can only get smaller as stage two is adapted. The only way for the QACF roots not to change is if it is already at the ideal solution. Performing the same process on every stage of the ACF in turn produces the minimum error at each iteration. As we continue this process, each ACF changes to produce the minimum error unless its roots already exactly match two of the roots of the unknown system. Since at each iteration, the error can get no worse; it must proceed to the minimum error possible under the initial assumptions.

3.2 QACF Simulations

Three unknowns (second-order, fourth-order and sixth-order) were simulated. Each was modeled with a fourth-order ACF. The coefficients of the unknowns used were

1. $\mathbf{h} = [1 \quad .4 \quad .2]^T$; roots at -.2±j.4
2. $\mathbf{h} = [1 \quad -.6 \quad .54 \quad .096 \quad .148]^T$; roots at -.2±j.4, .5±j.7
3. $\mathbf{h} = [1 \quad .6 \quad .27 \quad .474 \quad .5062 \quad .2208 \quad .0666]^T$; roots at -.2±j.4, .5±j.7, -.6±j.2

Theoretical solutions were done using (7) until the squared error was less than 10^{-8} from the theoretical minimum error. Simulations were run for 1000 iterations using a modified LMS algorithm with μ equal to .01. Each stage was started at -.5 ± j.5 or $\mathbf{h} = [1 \quad 1 \quad .5]^T$. $\mathbf{h}_i$ was obtained by averaging the last 500 iterations of each adaptation. This process was repeated until the squared error was less than 10^{-8} from the theoretical minimum error. The minimum theoretical error was zero for the first two cases and .0532 for the third case. Figures 4-9 show the root trajectories for the theoretical and LMS simulations for the three cases. The number of iterations required for convergence was smallest for the sixth-order unknown, 19 iterations; the fourth-order unknown required 23 iterations and the second-order unknown required 751 iterations.

Simulations show that extremely large number of iterations are required for convergence when the unknown is of order less than the QACF order. In this case, at least one of the roots of the QACF will be on the real axis and at least one will be zero. When the ACF roots are at or close to zero, the change in error is very small for a change in the root location. The further away from the real line the ACF roots are the larger the change in error for a similar change in the root locations. The delta that each weight is updated can be viewed like a quantization level that decreases as the error at each iteration decreases. This sensitivity is analogous to the root location sensitivity due to quantization levels discussed in [6]. This effect is shown in Figures 5 and 7.

Results for the 2nd order unknown are shown in Figures 4 and 5. Figure 6 shows the root trajectories for the theoretical solution. The roots of both ACF stages are started at -.5 ± j.5 and marked with an "o". The ideal root locations are indicated with an "x". The effect of the root sensitivity on the error is easily seen. Note the distance from the final ACF roots to the ideal root locations. The roots of the stage that converge to the unknown roots at -.2 ± j.4 are much closer to the ideal location than the roots converging to zero. However, both amount to equivalent value of error. Since the amount the root location changes is small, many iterations will be required to get the roots of the QACF which converge to zero to get as close as QACF roots which converge to unknown roots located

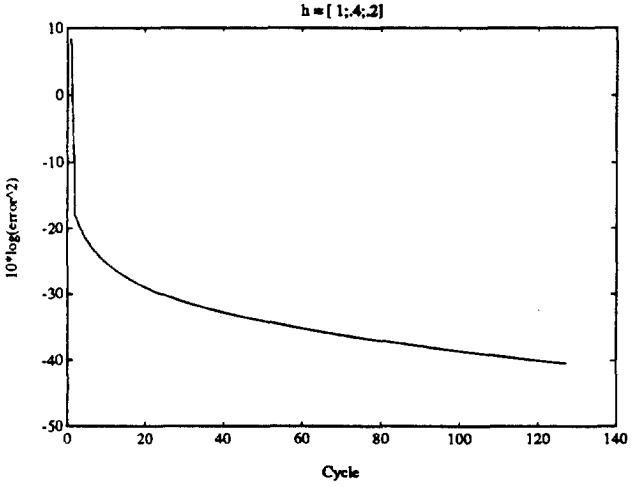

Figure 4 Log of squared error for 4th order ACF modeling a 2nd order unknown system **h**.

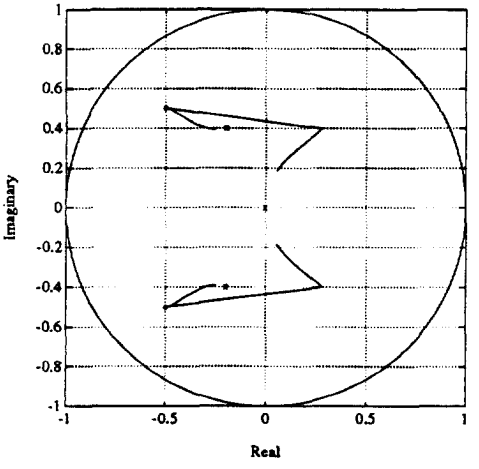

Figure 5 Root trajectories for 4th order ACF modeling a 2nd order unknown system **h**.

further away from the real axis. The convergence was stopped when the error was 10^{-5} for the 2nd order unknown to illustrate this effect. 127 cycles were required to achieve this error and 751 cycles were required to get 10^{-8} error. While not included, the error learning curve shows that the error is not decreasing significantly after 100 iterations of each adaptation. One thousand iterations were done to give a better average for the $\mathbf{h}_i$ of each stage.

It is expected that both ACF stages will converge to the 4th order unknown. These results are shown in Figures 6 and 7. Figure 6 shows the theoretical solution of the log of the squared error and Figure 7 shows the root trajectories for the LMS simulation. Note that for the same error, the roots locations of the ACF stages have converged exactly to the unknown roots. Only 23 cycles were required for the 4th order simulation compared to 751 cycles for the 2nd order simulation.

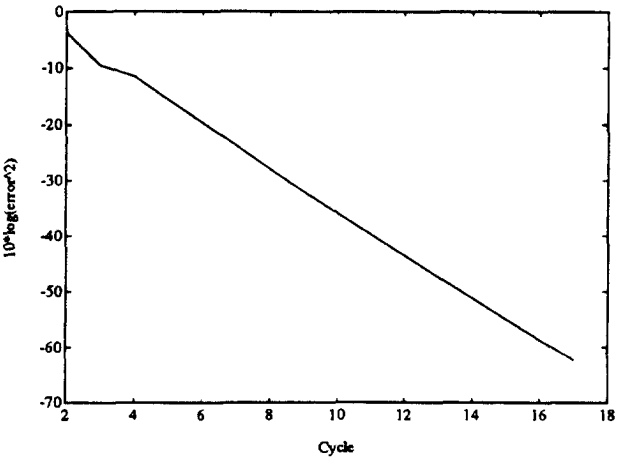

Figure 6 Log of squared error for 4th order QACF modeling a 4th order unknown system **h**.

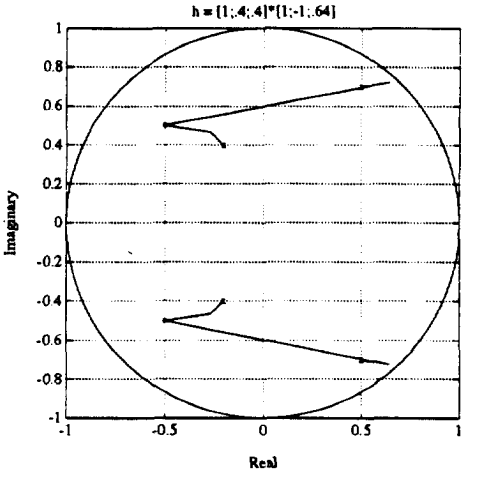

Figure 7 Root trajectories for 4th order QACF modeling a 4th order unknown system **h**.

The minimum error for the 6th order case was .0532. The theoretical root trajectories are shown in Figure 10. The simulation root trajectories are shown in Figure 11. The number of cycles required for convergence is smallest for the 6th order unknown.

The large difference in cycles required for convergence for unknowns of order less than the order of the QACF can be exploited to produce a more efficient algorithm. If the roots of an ACF stage are observed to be converging to zero, then that stage could essentially be deleted from the QACF. The process would be restarted and ACFs could be deleted until no further convergence to zero was detected.

4 CONCLUSION

Higher order adaptive convex filters can be obtained by cascading second order ACF stages together. The error

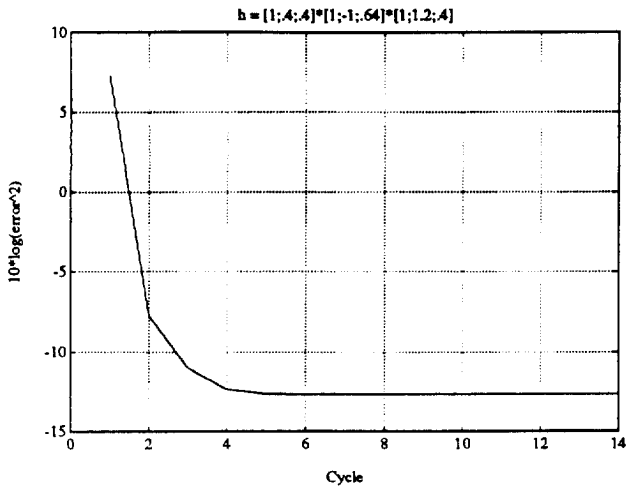

Figure 8 Log of squared error for 4th order QACF modeling a 6th order unknown system **h**.

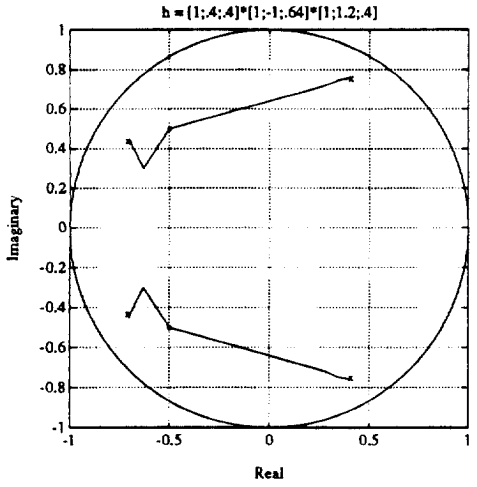

Figure 9 Root trajectories of 4th order QACF modeling a 6th order unknown system **h**.

surface of this filter is not quadratic but the solution is guaranteed to remain minimum phase during the adaptative process. A modified LMS algorithm that would have a quadratic error surface for higher order ACFs was examined. This method is called the quadratic ACF or QACF.

Since this structure has a quadratic error surface, the adaptive process will converge to the global minimum or theoretical solution of the error surface. While computationally intensive, the process appears to require relatively few iterations of the procedure for unknowns of higher order than the QACF order. Thus it appears that at least computationally it is not a disadvantage to model an unknown system with a lower order QACF. The misadjustment is smaller than if the entire ACF was being adapted simultaneously. A final advantage of this structure is that the roots of the QACF stages are guaranteed to be minimum phase at each iteration of the adaptive process. This advantage is an interesting possibility that is being studied further.

REFERENCES

[1] T. J. Canales and D. M. Etter, "Adaptive Convex Filters," 32nd Midwest Symposium on Circuits and Systems, 1989

[2] J. L. Rasmussen and D. M. Etter, "An Adaptive Technique for Designing Minimum Phase Models", Twenty-Fifth Annual Asilomar Conference on Signals, Systems, and Computers, 1991.

[3] J. L. Rasmussen and D. M. Etter, "Analysis of the Error Surface of the Adaptive Convex Filter", Twenty-Sixty Annual Asilomar Conference on Signals, Systems, and Computers, 1992.

[4] B. Widrow and S.D. Stearns, *Adaptive Signal Processing*, Englewood Cliffs, N.J. Prentice-Hall, 1985.

[5] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, Englewood Cliffs, J.J. Prentice-Hall, Section 6.8, 1989.

[6] D.G. Luenberger, *Introduction to Linear and Nonlinear Programming*, Reading, Massachusetts, Addison-Wesley, Chapter 10, 1973.