MINIMUM PHASE FIR FILTER DESIGN FROM LINEAR PHASE SYSTEMS USING ROOT MOMENTS

Tania Stathaki, Anthony Constantinides, Georgios Stathakis

Signal Processing and Digital Systems Section, Imperial College, UK

ABSTRACT

In this contribution we propose a method for a minimum phase Finite Impulse Response (FIR) filter design from a given linear phase FIR function with the same amplitude response. We are concentrating on very high degree polynomials for which factorisation procedures for root extraction are unreliable. The approach taken involves the use Cauchy Residue Theorem applied to the logarithmic derivative of the transfer function. This leads into a set of parameters derivable directly from the polynomial coefficients which facilitate the factorisation problem. The concept is developed in a way that leads naturally to the celebrated Newton Identities. In addition to solving the above problem, the results of the proposed design scheme are very encouraging as far as robustness and computational complexity are concerned.

1. INTRODUCTION

The design of Finite Impulse Response (FIR) digital filters has attracted considerable attention. An influential representative of the methods is based on the Remez exchange algorithm. However, most procedures assume a linear phase response with the consequence that the resulting filters do not have the lowest group delay. Direct design with prespecified phase response is possible [3]. In this paper we address the following problem:

"Given a linear phase FIR digital filter transfer function to determine an FIR digital filter which has identical amplitude response but is of minimum phase"

At first glance this may appear to be a trivial problem. Indeed a naive approach would be to factorise the given FIR transfer function and replace each of the zeros outside the unit circle with its reciprocal. This, in principle at least, would leave the overall amplitude response unaltered and would make the resulting transfer function minimum phase. However factorisation is a process fraught with difficulties in that it is a "non well-posed" ill-conditioned computational problem.

A new approach for polynomial factorisation without root finding is employed in this paper. The fundamental concepts rely on the root moments of polynomials which have been first formulated by Isaac Newton and lead to the relationships now known as the Newton Identities.

2. PRELIMINARIES

We consider a linear phase FIR digital filter transfer function having the following form

$$H(z)=z^{n}+h_1 z^{n-1}+h_2 z^{n-2}+\cdots+h_n=\prod_{i=1}^{n}(z-r_i) \qquad (1)$$

where r_i are the roots of the polynomial $H(z)$.

We employ the following notation:

- $r_j = r_{jin}$ if the root r_j is inside the unit circle.
- $r_j = r_{jout}$ if the root r_j is outside the unit circle.
- $r_j = r_{jo}$ if the root r_j is on the unit circle.

Thus we can write

$$H(z)=\left[\prod_j (z-r_{jin})\right]\left[\prod_j (z-r_{jout})\right]\left[\prod_j (z-r_{jo})\right] \quad \text{or}$$

$$H(z)=H_{\min}(z)H_{\max}(z)H_0(z)$$

where $H_{\min}(z)$ is the minimum phase part of $H(z)$ and $H_{\max}(z)$ is the maximum phase part of $H(z)$. The factor $H_0(z)$ contains all roots that are on the unit circle.

Some useful general points need to be made.

- Linear phase FIR digital filter transfer functions have non-minimum phase. From this follows that a real given FIR transfer function has zeros at location z_0, $1/z_0$, z_0^* and $1/z_0^*$ for $|z_0|\neq 1$.
- The group delay of an n th order linear phase FIR transfer function is $\tau(\theta)=n/2$. A typical FIR digital filter transfer function may be of length 200 or more for a range of applications with stringent specifications as in telecommunications. For such filters the group delay may be undesirable particularly when it approaches about 200ms, when bidirectional human-to-human communication is not viable.
- Often in many applications the phase response is either unimportant or irrelevant. For example in some speech processing areas it is not significant. This form of freedom in the design of the filters is not normally taken into consideration by existing FIR filter design methods.
- For linear phase FIR filters, the amplitude response is a linear function of the design parameters, that is the impulse response coefficients. For minimum phase FIR filters the amplitude response is a nonlinear function of these coefficients. At any rate the design of such filters from amplitude specifications would inevitably lead to a stage of factorisation in order to select the appropriate zeros and hence problems with imprecision would arise.

3. PROPOSED DESIGN ALGORITHM

We aim in this paper to derive the required nonlinear phase FIR filter transfer functions from corresponding linear phase functions which are assumed to be designable by such standard means as the Remez exchange algorithm.

Let the linear phase FIR filter transfer function be $H(z)=H_{\min}(z)H_{\max}(z)H_0(z)$ as already indicated.

Let also n_0 be the number of zeros of $H_0(z)$ and n_i the number of zeros of $H_{\min}(z)$.

On the unit circle with $z=e^{j\theta}$ we have $H_{\min}(e^{j\theta})=A(\theta)e^{j\phi(\theta)}$, $H_{\max}(e^{j\theta})=e^{jn_i\theta}A(\theta)e^{-j\phi(\theta)}$ and

$$H_0(e^{j\theta}) = \prod_{r=1}^{n_0/2}(z^2 - 2\cos\theta_r z + 1) = e^{jn_0\theta/2}B(\theta)$$

Hence $H(e^{j\theta}) = e^{j(n_i+n_0/2)\theta}\left[A(\theta)\right]^2 B(\theta)$

The group delay as a fraction of the sampling period is $\tau(\theta) = (n_i + \frac{n_0}{2}) = \frac{n}{2}$. Moreover, $|H_{\min}(e^{j\theta})| = |H_{\max}(e^{j\theta})|$

Thus, in principle, to obtain a minimum phase version of the given transfer function we can follow the steps below.

Step 1

Either determine $H_{\max}(z)$ and reflect its zeros into the unit circle, or determine $H_{\min}(z)$ and make each of its zeros of multiplicity 2.

Step 2

Find $H_0(z)$

Step 3

Construct the transfer function as $T(z) = \left[H_{\min}(z)\right]^2 H_0(z)$

Then we shall have $|T(e^{j\theta})| = |H(e^{j\theta})|$.

Both Step 1 and Step 2 imply at first glance that a root finding procedure may be required. However, as already pointed out, root finding procedures are known to be inaccurate and unreliable for large order polynomials. Factorisation without root finding forms also the basis of the procedure developed in [1],[3],[4]. In [1],[4] use is made of the real cepstral parameters as in [5], where the cepstral aliasing problem is recognised and careful procedures are recommended to reduce its effects. In [3] they approach the factorisation problem from the Lagrange interpolation point of view. In the above procedures it is assumed that the zeros of the transfer function on the unit circle are a priori known. We make no such assumption in our present paper.

An alternative and direct polynomial construction procedure without having to go through root estimation procedures is possible through the Root Moments of a given polynomial [7-8], or the differential cepstrum [6].

4. ROOT MOMENTS

In relation polynomials typically given as in equ.(1) Newton defined a set of parameters given by

$$S_m = r_1{}^m + r_2{}^m + \cdots + r_n{}^m = \sum_{i=1}^{n} r_i{}^m \qquad (2)$$

where r_i is the ith root of (1). The roots of (1) are not needed explicitly to compute S_m in that these parameters can be determined directly from the coefficients h_i. The parameters S_m are known as the root moments of the polynomial $H(z)$. They are related to many signal processing operations, dominant amongst which is the differential cepstrum. However it would be limiting to think of them purely in this sense since a wider perspective enables us to provide answers to many digital signal processing problems that have been, hitherto, unattainable [7].

4.1 Iterative Estimation of Root Moments

By writing the polynomial (1) as a product of factors we can write $H'(z) = \sum_{i=1}^{n} \frac{H(z)}{z - r_i}$ and given that $H(r_i) = 0$ we have

$$H'(z) = nz^{n-1} + (S_1 + nh_1)z^{n-2} + (S_2 + h_1 S_1 + nh_2)z^{n-3} + \cdots$$

$$+ (S_m + h_1 S_{m-1} + h_2 S_{m-2} + \cdots + nh_m)z^{n-m-1} + \cdots$$

By direct differentiation of equation (1) we have

$$H'(z) = nz^{n-1} + (n-1)h_1 z^{n-2} + (n-2)h_2 z^{n-3} + \cdots + (n-m)h_m z^{n-m-1} + \cdots \qquad (3)$$

Hence by equating the last two expressions we obtain the following fundamental relationships known as *Newton Identities*

$S_1 + nh_1 = (n-1)h_1$ or $S_1 + h_1 = 0$

$S_2 + h_1 S_1 + nh_2 = (n-2)h_2$ or $S_2 + h_1 S_1 + 2h_2 = 0$

and generally

$S_m + h_1 S_{m-1} + h_2 S_{m-2} + \cdots + nh_m = (n-m)h_m$ or

$$S_m + h_1 S_{m-1} + h_2 S_{m-2} + \cdots + mh_m = 0 \qquad (4)$$

When the signal treated by this means is infinitely long, the above equation is repeatedly used to calculate successive values of the root moments. If the signal is of finite duration then for $m > n$ $S_m + h_1 S_{m-1} + h_2 S_{m-2} + \cdots + h_n S_{m-n} = 0$.

The same relationship as above can be used to calculate S_m for $m < 0$ by inserting successively values for m equal to $n-1$, $n-2$, $n-3, \cdots$ etc. It should be noted that S_m for either positive or negative values of m are evaluated recursively from the coefficients of equation (1) alone.

The above relationships also follow from the definition of the differential cepstrum and are essentially included in [6]. However in [6] n is assumed to be finite a priori known. This is only a minor point as the iteration in equ.(4) do not require n to be finite and, hence, they can be applied to infinite duration signals. It is sufficient at this juncture to observe that both finite duration signals and infinite duration signals of exponential entire function type interpretation can be treated in the same way [7]. To facilitate the exposition, the parameters in (2) are referred to as the root moments. This terminology emphasises the deviation from the differential cepstrum.

4.2 Implications and Interpretation

Essentially one can interpret the set of equations (4) as a transformation of the coefficients $\{h_r\}$ to the parameter set $\{S_m\}$ of the same cardinality. The transformations are one-to-one and hence we can have the following existence corollaries.

Corollary 1 Given a set of coefficients $\{h_r\}$ of the n th degree polynomial in equation (1) which has roots $\{r_i\}$ $i = 1, \cdots, n$, there exists a set of parameters $\{S_m\}$ $m = 1, \cdots, n$, $S_0 = n$, given by equ.(2).

Corollary 2 Conversely given a set of root moments $\{S_m\}$ there exists a set of coefficients $\{h_r\}$ $r = 1, \cdots, n$, for a polynomial as in equ.(1) determinable recursively through equ.(4). The proofs are self evident from the above analysis.

4.3 Root Moments of Products of Signals

In our main problem we need the following result. Assume that the root moments of the polynomial $f_1(z)$ are $S_m^{f_1(z)}$ and the root moments of the polynomial $f_2(z)$ are $S_m^{f_2(z)}$. Then the root moments of the product $f(z) = f_1(z) f_2(z)$ are $S_m^{f(z)} = S_m^{f_1(z)} + S_m^{f_2(z)}$.

4.4 Non Iterative Estimation of Root Moments

The Newton Identities yield the root moments of the entire signal, encompassing not only those roots that lie within the unit circle but also those that are outside the unit circle. However, it is often the case that a specific factor of a given polynomial $H(z)$ is required, such as the minimum phase factor and in this case its root moments can be determined in a different manner.

Let a closed contour Γ defined as $z=\rho(\theta)e^{j\theta}$ contain the roots of the required factor of $H(z)$. Then it follows from the Cauchy residue theorem that the root moments of this factor are given by

$$I_\Gamma(m)=S_m^\Gamma=\frac{1}{2\pi j}\oint_\Gamma \frac{H'(z)}{H(z)}z^m dz \tag{5}$$

This is evident from the fact

$$S_m^\Gamma=\frac{1}{2\pi j}\oint_\Gamma \sum_i \frac{1}{(z-r_i)}z^m dz$$

and the contribution to the integration are those due to those r_i's that lie within Γ. (It is assumed that we have no zeros on Γ.)

In practice the contour integration will have to be effected directly from the coefficients of $H(z)$ and this can be done quite conveniently through the use of the DFT as it is shown below.

Equation (5) becomes for $z=\rho(\theta)e^{j\theta}$

$$S_m^\Gamma=\frac{1}{2\pi j}\int_{-\pi}^{\pi} g(\theta)e^{j(m+1)\theta}d\theta \tag{6}$$

where

$$g(\theta)=\frac{H'(\rho(\theta)e^{j\theta})}{H(\rho(\theta)e^{j\theta})}\left(\rho(\theta)\frac{d\rho(\theta)}{d\theta}+j\rho(\theta)\right)\rho^m(\theta) \tag{7}$$

Discretisation of equation (6) suitable for DFT use requires values $\theta_k=\frac{2\pi}{N}k,\ k=0,1,\cdots,N-1$ for an N-point transform. Therefore, we have the inverse DFT

$$S_m^\Gamma\approx\frac{1}{jN}\sum_{k=0}^{N-1} g(\theta_k)e^{j(m+1)\theta_k} \tag{8}$$

If the contour of integration is the unit circle C: $|z|=1$ then the resulting root moments from the above, correspond to those of the minimum phase component of $H(z)$. In this case we have the special form of (8)

$$S_m^{f_{min}(z)}\approx\frac{1}{N}\sum_{k=0}^{N-1}\frac{H'(\theta_k)}{H(\theta_k)}e^{j(m+1)\theta_k} \tag{9}$$

For either equ.(7) or for the special form of equ.(8) the computation of $g(\theta_k)$ can be done through the use of the DFT also.

It is observed that on $z=\rho(\theta)e^{j\theta}$ we can write

$$H'(\rho(\theta)e^{j\theta})=e^{j(n-1)\theta}\sum_{i=0}^{n-1}(n-i)h_i\rho^{n-i-1}(\theta)e^{-ji\theta}$$

which for $\theta=\theta_k$ can be computed as

$$H'(\rho(\theta_k)e^{j\theta_k})=e^{j(n-1)\theta_k}\,\mathrm{DFT}\left\{(n-i)h_i\rho^{n-i-1}(\theta_k)\right\} \tag{10}$$

Similarly we have

$$H(\rho(\theta_k)e^{j\theta_k})=e^{jn\theta_k}\,\mathrm{DFT}\left\{h_i\rho^{n-i}(\theta_k)\right\} \tag{11}$$

and hence

$$g(\theta_k)=e^{-j\theta_k}\frac{\mathrm{DFT}\left\{(n-i)h_i\rho^{n-i-1}(\theta_k)\right\}}{\mathrm{DFT}\left\{h_i\rho^{n-i}(\theta_k)\right\}}\left(\rho(\theta)\frac{d\rho(\theta)}{d\theta}+j\rho(\theta)\right)\rho^m(\theta)$$

With N a power of 2 we can use the Fast Fourier Transform (FFT) algorithm.

5. THE ALGORITHM

The algorithm relies on the direct extraction of the appropriate factors from the FIR linear phase transfer function needed to implement $T(z)$ above.

Step 1:
Integrate around a circle centred at the origin and of radius less than unity. The radius of the contour is of crucial importance and it is examined separately below.
The integration with a careful choice of the contour radius gives $S_1(m)=S_{in}(m)$ which parameters correspond to the root moments of that part of the original FIR transfer function which has its zeros inside the unit circle.

Step 2:
Integrate around a circle centred at the origin and of radius greater than unity. Again the radius of the contour must be selected carefully, but a good selection in Step 1 yields a correspondingly good selection as the reciprocal of the radius. The integration produces the parameters $S_2(m)=S_{in}(m)+S_0(m)$ where $S_0(m)$ are the root moments of that factor of the original FIR digital filter transfer function which has its zeros on the unit circle.

Step 3:
The required transfer function has the root moments $S(m)=2S_{in}(m)+S_0(m)$, obtained as $S(m)=S_1(m)+S_2(m)$.

Step 4:
From **Step 3** and from the Newton Identities we form the required polynomial FIR transfer function of degree $S_1(0)+S_2(0)$.

5.1 Estimation of the Radii of the Integration

The radii of integration in the above algorithm must be chosen so as to enclose the appropriate zeros of the given FIR digital filter transfer function. Thus for $S_1(m)$ the radius of the integration contour r must be such that $1>r>\max(|r_{in}|)$, while for $S_2(m)$ the radius of the integration contour r must be chosen such that $1<r<\min(|r_{out}|)$.

For equiripple piecewise constant filters the required radii can be estimated as follows.
Let us remove the linear phase factor from the frequency response to yield only a real function. This function now we shift vertically half way between its maximum and minimum values. Since the initial transfer function is equiripple the result of these operations will be a real function of equiripple modulus almost everywhere. The ripple variation remains unchanged, namely a normalised response will vary between $1+\delta$ and $1-\delta$ almost everywhere except in the transition band. In Fig. 2 we indicate the zeros of the shifted transfer function.

A reasonable representation of this zero pattern almost everywhere, is given by

$$C(z) = (z^n - a^n)(z^n - \frac{1}{a^n}) \qquad (12)$$

The above transfer function is equiripple, linear phase and its zeros are located on two circles controlled by the parameter a. The amplitude characteristic is equiripple between the values

$C_{\max} = 2 + (a^n + \frac{1}{a^n})$ and $C_{\min} = 2 - (a^n + \frac{1}{a^n})$.

The ripple width of $C(z)$ can be found from its maximum and minimum values which yield the ripple variation of the shifted FIR filter $\delta = \frac{2}{(a^n + \frac{1}{a^n})}$. The quantity δ is an a priori known design parameter. Hence we can estimate the radius of the circle on which the zeros are expected to be located as $a = \left(\frac{1}{\delta} \pm \sqrt{(\frac{1}{\delta^2} - 1)}\right)^{\frac{1}{n}}$. For small ripple width the above can approximated to $a = \left(\frac{2}{\delta}\right)^{\frac{1}{n}}$. We show a specific minimum phase design result in Fig. 3 along with further results as indicated below.

5.2 Variations in the Procedures

It may be the case in a system application that the reduction in the group delay obtained by the above algorithms is more than the required amount. Then we can improve the non linearity in the phase response as follows.

- The root moments corresponding to the stop band transmission zeros remain the same as above.
- From the rest of the zeros we can select an appropriate number in conjugate form, for real transfer functions, in an arbitrary fashion.

Fig. 3 shows such an example, from which it is seen that the group delay here is more than the minimum but less than its maximum value. It is almost equiripple to a constant. Further work is necessary to explore the options open here.

6. EXPERIMENTAL RESULTS

Figure 1 shows the amplitude of the shifted frequency response while Figure 2 shows its zeros. It is seen that these zeros are similar of those of $C(z)$. Figure 3 shows the group delay responses for the minimum phase and for another intermediate solution for which the maximum phase zeros have been selected on alternate basis. It is evident that the proposed method operates as expected. However, there are many questions that need to be addressed particularly in relation to the choice of contour for the implementation of equ.(9). Moreover, there is the need for exploring further the implications of equ.(12).

7. REFERENCES

[1] Boite R. and H. Leich, "A New Procedure for the Design of High Order Minimum Phase FIR Digital or CCD FIlters", *Signal Processing*, vol. 3, pages 101-108, 1981.

[2] Herman O. and H.W. Schussler, "Design of Non Recursive Digital Filters With Minimum Phase", *Electronics Letters*, vol. 6, pages 329-330, 1970.

[3] Kamp Y. and Wellekens C.J., "Optimal Design of Minimum Phase FIR Filters", *IEEE Trans. on ASSP*, vol. 31, pages 922-926, 1983.

[4] Mian G.A. and A.P. Naider, "A Fast Procedure to Design Equiripple Minimum Phase FIR FIlters", *IEEE Trans. on Circuits and Systems*, vol. 29, no. 5, pages 327-331, 1982.

[5] Nikias C.L. and A. Petropulu, *Higher Order Spectra Analysis*, Prentice Hall, 1993.

[6] Polydoros A. and A.T. Fam, "The Differential Cepstrum Definition and Properties", *Proc. IEEE Int. Symposium on Circuits and Systems*, pages 77-80, April 1981.

[7] Stathaki T., "Cumulant Based and Algebraic Techniques for Signal Modelling", *PhD Thesis*, University of London, 1994.

[8] Todhunter I., *Theory of Equations*, McMillan & Co. London 1882.

Figure 1: Shifted frequency response

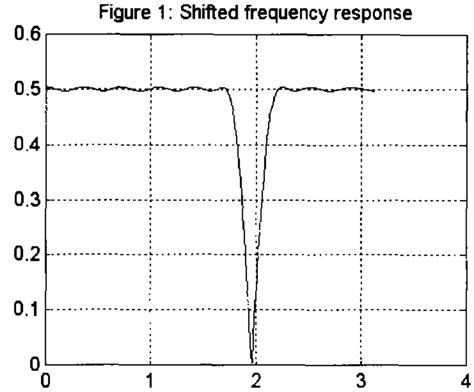

Figure 2: Zeros of shifted zero-phase response

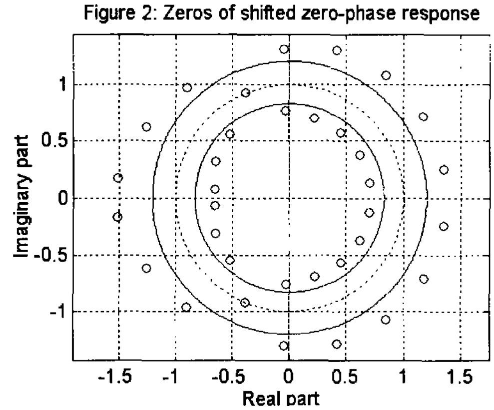

Figure 3: Group delay responses