A POLYNOMIAL FACTORISATION METHOD FOR LOW GROUP DELAY FIR EQUIRIPPLE FILTER DESIGN FROM LINEAR PHASE SYSTEMS

Ioannis Fotinopoulos and Tania Stathaki

Communications and Signal Processing Research Group
Imperial College of Science, Technology and Medicine
United Kingdom

ABSTRACT

In this contribution we propose a method for reduced-group-delay Finite Impulse Response (FIR) equiripple filter design from a given linear phase FIR filter. Our restriction is to keep the amplitude response identical. We are concentrating on very high degree polynomials (transfer functions) for which factorisation procedures for root extraction are unreliable. The initial linear phase transfer functions are obtained by standard design algorithms and particularly in this study by the Remez algorithm. The approach taken involves the use of the Cauchy Residue Theorem applied to the contour integral of the logarithmic derivative of the transfer function around appropriate curves. This leads into a set of parameters related directly to the polynomial coefficients, which facilitate the factorisation 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 [1-2]. 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. In this paper we address the following problem: "Given a linear phase FIR filter transfer function, design a new filter with the same amplitude response and lower group delay without distorting significantly the phase linearity properties"

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 give minimum group delay but highly non-linear phase response. To try to optimise this trade off we could replace a selection of zeros outside the unit circle with their reciprocal. Even then we would need to factorise the given FIR transfer function for root extraction. However, factorisation is a process fraught with difficulties in that it is a “non well-posed”, ill-conditioned computational problem [3-5].

In this paper a new approach for polynomial factorisation without root finding is employed. The fundamental concepts rely on the root moments, a non-linear transformation of polynomial coefficients, which have been first formulated by Isaac Newton and lead to the relationships known as the Newton Identities [6].

We consider a linear phase FIR digital filter transfer function

$$f(x) = x^n + p_1 x^{n-1} + p_2 x^{n-2} + \cdots + p_n = \prod_{i=1}^{n} (z - r_i) \qquad (1)$$

Linear phase, real FIR digital filter transfer functions are symmetric or anti-symmetric and have non-minimum phase. It can be easily shown that if a transfer function of the above type has a zero at the location r_i, it will also have the zeros $1/r_i$, r_i^* and $1/r_i^*$ for $|r_i| \neq 1$. We can also write:

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

where $H_{\min}(z)$, $H_{\max}(z)$ are the minimum and maximum phase parts of $H(z)$ respectively. The factor $H_{\mathrm{o}}(z)$ contains all roots that are on the unit circle.

Some useful general points need to be made.

- The group delay of an n^{th} order linear phase real FIR transfer function is $n/2$. For filters of length 200 or more the group delay may be undesirable particularly when it approaches large values, such that bi-directional human-to-human communication is not viable.
- The phase response of a signal need to be kept as linear as possible for several applications. In speech for example non linearity in phase should be kept to a minimum to maintain the quality of the perceived signal.

At any rate, the design of minimum group delay and almost linear phase FIR filters from linear phase FIR filters with specific amplitude requirements would inevitably lead to a stage of factorization in order to select the appropriate zeros, and hence problems with imprecisions would arise.

2. PROPOSED DESIGN ALGORITHM

At this section we aim to derive the required nonlinear phase FIR filter transfer functions from corresponding linear phase functions.

Let the linear phase real FIR filter transfer function be

$$H(z) = H_{\min}(z) H_{\max}(z) H_{\mathrm{o}}(z).$$

As already indicated to obtain a lower group delay and an almost linear phase version of a given transfer function we need to replace a selection of zeros outside the unit circle with their reciprocal. Let $H_{\min-sel}(z)$ correspond to the selected zeros inside the unit circle. Make each of its zeros of multiplicity 2. Then we find $H_{\max-not-sel}(z)$, that corresponds to the non-selected zeros outside the unit circle, $H_{\min-not-sel}(z)$ that correspond to the non-selected zeros inside the unit circle, and $H_{\mathrm{o}}(z)$, and finally construct the transfer function as

$$T(z)=[H_{\min-sel}(z)]^2 H_{\mathrm{o}}(z) H_{\max-not-sel}(z) H_{\min-not-sel}(z)$$

so that

$$|T(e^{\mathrm{j}\theta})|=|H(e^{\mathrm{j}\theta})|.$$

At first glance, the implementation of the above seems to require a root finding procedure. However, as already pointed out, root finding procedures are known to be inaccurate and unreliable for large order polynomials. 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 [5-6].

3. ROOT MOMENTS

In relation to polynomials typically given as in (1), Newton defined a set of parameters given by [5-6]:

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

where r_i is the i^{th} root of (1). The parameters S_m are known as the root moments of the polynomial $H(z)$.

The following fundamental relationships known as *Newton Identities* hold [7].

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

$$S_m + h_1 S_{m-1} + h_2 S_{m-2} + \cdots + mh_m = 0 \tag{3}$$

The Newton Identities yield the root moments of the entire signal. 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 $z=\rho(\theta)e^{\mathrm{j}\phi(\theta)}$ 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:

$$I_\Gamma(m) = S_m^\Gamma = \frac{1}{2\pi\mathrm{j}} \oint_\Gamma \frac{H'(z)}{H(z)} z^m dz \tag{4}$$

If the contour of integration is the unit circle $\Gamma:|z|=1$ then the resulting root moments from the above, correspond to those of the minimum phase component of $H(z)$ [7]. In this case we have the special form of (4)

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

(5) In [9] we derive analytically equations (4) and (5) as well as describe the implementation of (5) using the Fast Fourier Transform.

4. A NOVEL DESIGN METHOD

In [7] and [8], we applied the Cauchy residue theorem using different contours to invert all maximum phase zeros inside the unit circle and keep the zeros that are located on the unit circle as they are. We proved that although the group delay gets minimum, this results to a distortion in phase that gets highly non-linear.

In this paper we try to find the relationship between reduce in group delay and phase linearity. We propose a class of methods that attempt to invert some of the zeros outside the unit circle and leave the rest of them unaltered. For each type of equiripple FIR filters (low-pass, high-pass, band-pass and band-stop) it is possible to find a range of frequencies that contain the zeros of the pass band that we plan to invert. In this paper we will concentrate on low-pass filters without loss of generality. Given the analysis presented for low-pass filters, the analysis for any other type of filter is straightforward.

4.1 Selection of zeros for inversion with respect to their modulus

To recall, we aim to invert some of the zeros that lie outside the unit circle. It is known that a zero that lies on the unit circle produces an infinite group delay for a specific frequency, while the contribution to the group delay, of a zero with very large modulus is insignificant. In general, as the modulus of a zero gets close to 1, its contribution to the group delay increases.

Taking into consideration the above an initial thought would be to select the zeros for inversion with respect to their modulus. We can form circles of different radii, all with the same centre, i.e., the origin of the z-plane. Then we could invert the zeros that are located within particular rings.

This seems initially an interesting approach.

This approach would perform well if the zeros were located in various positions with respect to their modulus. However, this is not the case with polynomials designed using the Remez algorithm. Most of the zeros of these polynomials within the pass-band are located very close to two different circles. One of them with radius slightly less that unity and the other with radius slightly greater that unity. Very few zeros may located far from the unit circle, but as already seen, their contribution to the group delay is very insignificant.

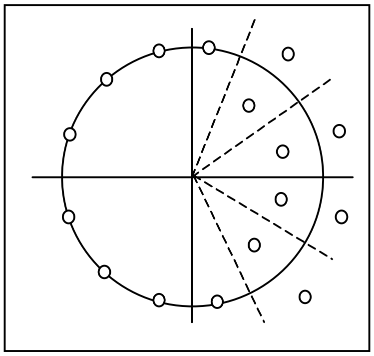
Hence, it is obvious that the choice of zeros for inversion with respect to their modulus is not realisable in our case.

We then proceed to seek for zeros for inversion with respect to their phase.

4.2 Selection of zeros for inversion with respect to their phase

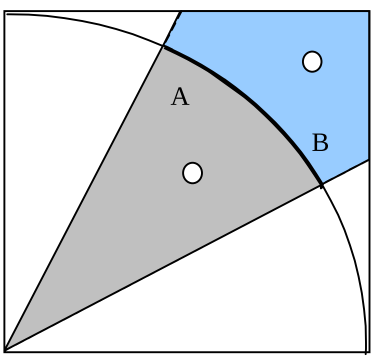
In this section we propose to select particular regions outside the unit circle whose zeros outside will be inverted. The shape of these regions should be chosen in a way such that the zeros that lie within may be of any modulus and of phase that is restricted within a specified range of frequencies.

This yields to a situation shown in the following figure.



The pass-band may be divided into regions that have the shape as shown in the figure below. More specifically, each of these is formed by two rays that start from the origin and go to infinity.

These regions may vary in width and may include a minimum of two zeros -one inside and one outside the unit circle- and a maximum of all the zeros that are located within the pass-band. We propose for some of these regions to invert the zeros outside the unit circle and for the rest to leave the zeros unaffected. To be able to implement the method, lets take one of the regions described earlier. This is now divided into two parts the part that includes the area inside the unit circle and the part that includes the area outside the unit circle In the figure these areas are named A and B respectively.



What is of great interest is the way to find out the S-values (root moments) that correspond to the zeros outside the unit circle of this selected region. This is a quite difficult task because these zeros can sometimes take large values and therefore it is not guaranteed that they will lie inside a closed curve. As a result, the Cauchy-Residue theorem cannot be applied directly to these zeros. What follows is a method that overpasses this problem.

We describe the main necessary steps of the method.

1. Using the Newton identities find the values $S^A(m)$ of the zero(s) located inside the unit circle and at the selected region. This will be done by a special contour integration. The contour is formed by two rays starting from the origin and ending on the unit circle and one arc. This arc lies on the unit circle and is of width selected such that the region (region A) will include the zeros whose reciprocals need to be inverted inside.

These values $S^A(m)$ yield to the reconstruction of a transfer function. The order of this transfer function can be found from the value $S^A(0)$.

2. At this transfer function z is replaced by z^{-1} and the new transfer function will represent the part of the selected region that lies outside the unit circle, namely region B in the figure above. In the MATLAB environment a transfer function is described by its coefficients. From the implementation point of view we first need to flip the vector that is formed by the coefficients of the transfer function that corresponds to region A. Then to divide by the first element since we are working with normalised polynomials. The resulting vector will contain the coefficients of a transfer function whose zeros are these zeros of the initial transfer function that lie within the region B.

We find the values $S^B(m)$ that correspond to the zeros outside the unit circle of the selected region using the Newton Identities.

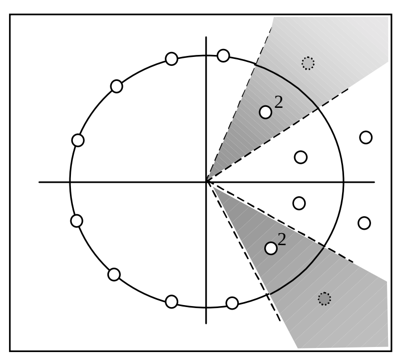
3. Using the Newton identities, we find the S-values of the pre-specified filter given by the Remez algorithm. These values $S(m)$ will correspond to all zeros inside, on and outside the unit circle including the zeros in the selected regions A and B of which the S-values were found in steps 1 and 2.

4. From the S-values of the mixed phase filter transfer function $S(m)$ we subtract the S-values corresponding to the zeros of region B, $S^B(m)$, and we add the S-values corresponding to the zeros of region A, $S^A(m)$. We then get

$\tilde{S}(m) = S(m) + S^A(m) - S^B(m)$

5. Using the values $\tilde{S}(m)$ and the Newton identities, a transfer function of a filter of the same order as the pre-specified filter is reconstructed. This filter will have bigger group delay than the minimum phase filter version of the initial transfer function but much better phase properties as far as phase linearity is concerned.

An example of a possible implementation of the above algorithm is graphically illustrated in the figure below. In that case the algorithm is applied for the two regions shown.



Note that there are various ways to implement the proposed algorithm with respect to the selection of different contours and this is an issue for further investigation.

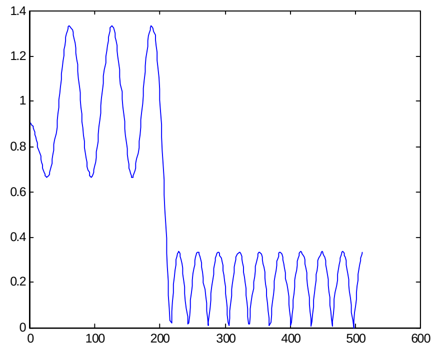
5. EXPERIMENTAL RESULTS

The proposed method was tested with different filters. It performs very well as far as both accuracy and computational complexity are concerned.

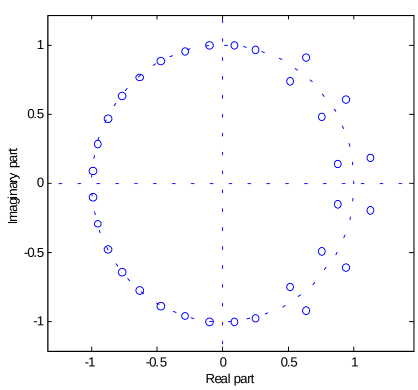
Some indicative results are shown in the figures below.

The given filter is a low pass filter of size 32.

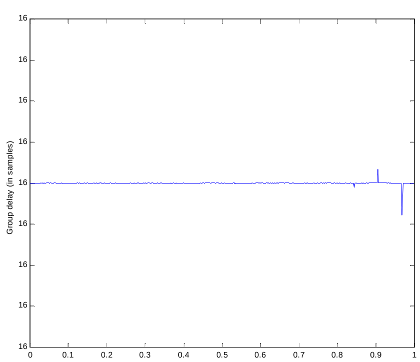
The bands are:		
	Pass band:	0-0.4
	Transition band:	0.4-0.41
	Stop band:	0.41-1



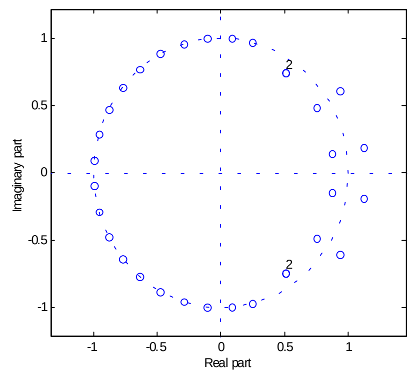
The amplitude response remains unaffected no matter what the selected region is



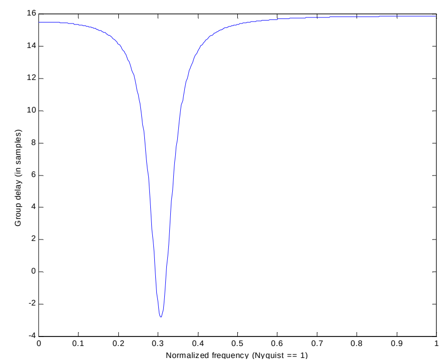
Mixed-phase filter zeros on the complex plane



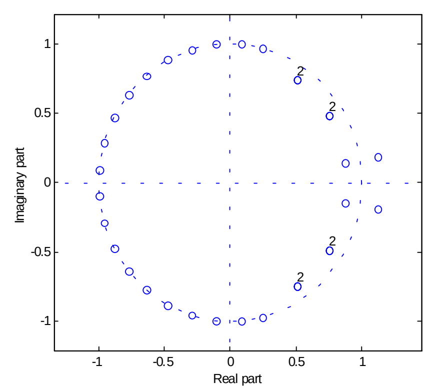
Group delay of mixed-phase filter



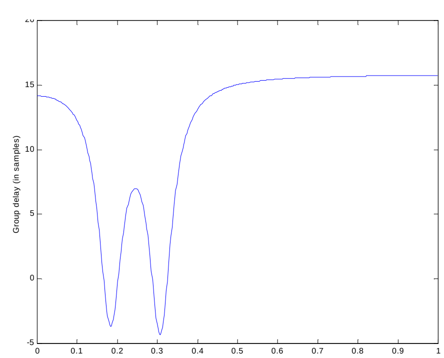
Experiment 1
The outside zeros of the passband are inverted



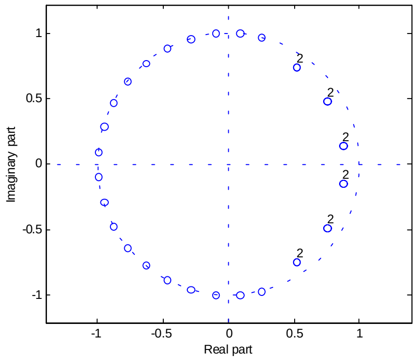
Experiment 1 : Group delay



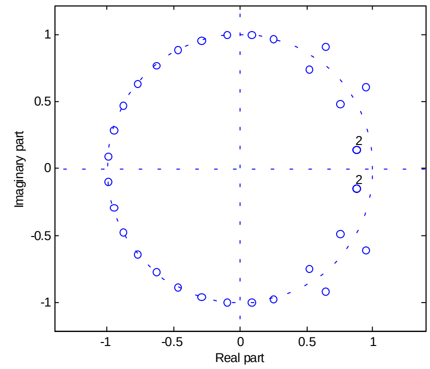
Experiment 2
The outside plus the middle zeros of the passband are inverted



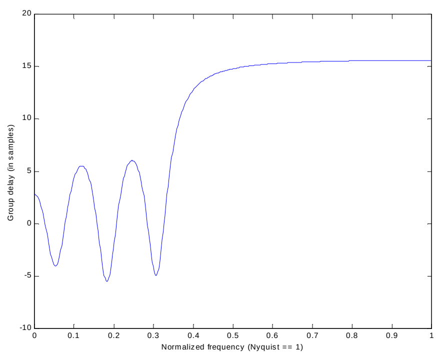
Experiment 2 : Group delay



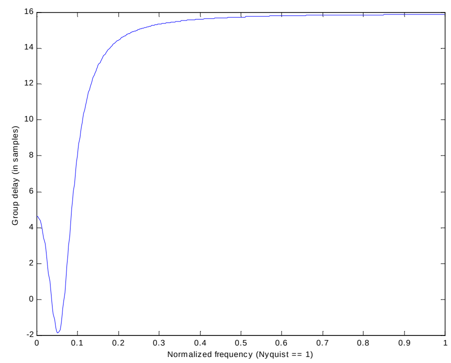
Experiment 3
All the zeros of the passband are inverted



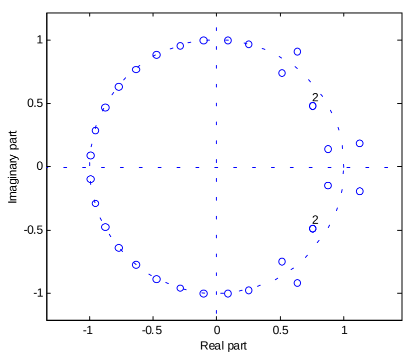
Experiment 5
The inner zeros of the passband are inverted



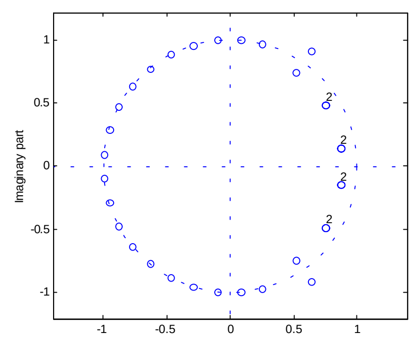
Experiment 3 : Group delay



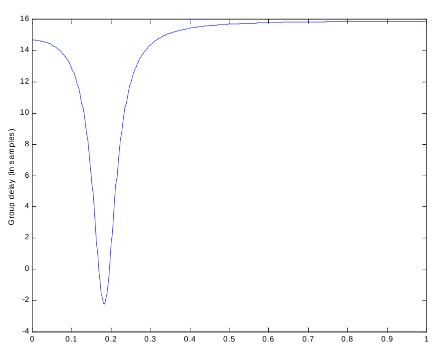
Experiment 5 : Group delay



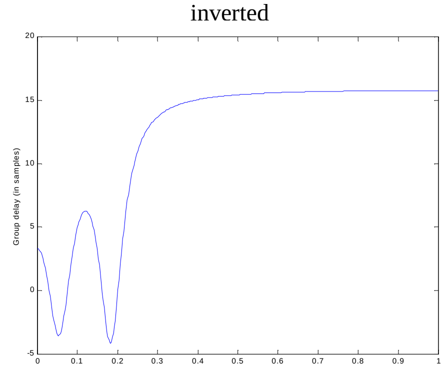
Experiment 4
The middle zeros of the passband are inverted



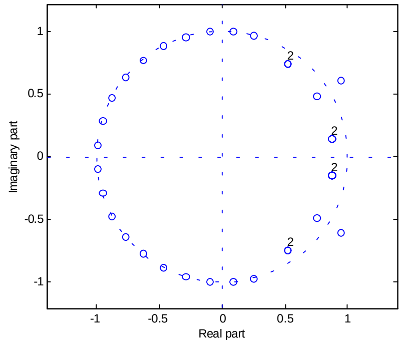
Experiment 6
The middle and inner zeros of the passband are inverted



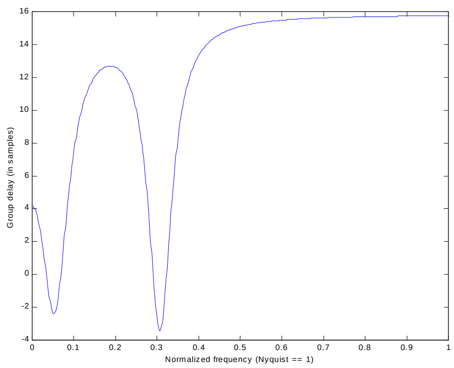
Experiment 4 : Group delay



Experiment 6 : Group delay



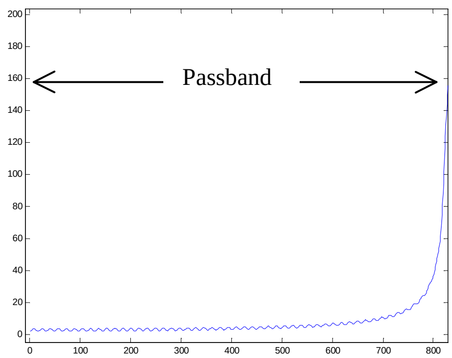
Experiment 7
The outside and inner zeros of the passband are inverted



Experiment 7 : Group delay

As the order of the filter gets higher the ripples in the group delay response get smaller. Therefore for a filter of high order and a good selection of zeros for inversion, the group delay could give very small ripples for the whole passband, thus approaching an almost flat group delay response for the passband region.

A graphical example of the above is shown below:



In this example the given filter is a low pass filter of order 512. The sampling frequency is 2048 and the bands are: Passband : 0-0.4, Stopband 0.41-1. It is clear that the ripples in the passband region (up to the 819th sample) are so small as to be ignored.

6. REFERENCES

[1] Kamp Y. and Wellekens C.J., "Optimal design of minimum phase FIR filters", *IEEE Trans. on ASSP*, Vol. 31, pages 922-926, 1983.

[2] 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.

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

[4] 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.

[5] Stathaki T., "Root moments: a digital signal processing perspective", *IEE Proc. of Vision Signal and Image Processing*, Vol 145, No. 4, August 1998.

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

[7] Tania Stathaki, Ioannis Fotinopoulos, Anthony Constantinides, “Root moments: a nonlinear signal transformation for minimum FIR filter design”, *Proc. of the NSIP 99*, 20-23 June, Antalya, Turkey.

[8] Ioannis Fotinopoulos, “Equiripple minimum phase FIR filter design from linear phase systems using a novel technique for polynomial factorisation”, *MEng Thesis*, University of London, 1999