On Causal Linear Phase IIR Digital Filters

MARK A. CLEMENTS, MEMBER, IEEE, AND JERROLD W. PEASE

Abstract—This paper addresses the issue of the existence of causal linear phase IIR digital filters. The conditions for their existence are derived, and a set of examples satisfying these conditions is given. Specifically, the sequences must be integer samples of a continuous-time function, band-limited to \((-\pi, \pi)\), which is symmetric and has zero crossings at the negative integers. Also, the point of symmetry can be neither an integer nor a half integer. A number of interesting properties of the related discrete- and continuous-time functions are developed.

I. INTRODUCTION

One of the highly desirable properties that finite impulse response (FIR) digital filters possess is that they can be made to have linear phase. The essence of this idea is that a filter of a certain desired magnitude characteristic can be designed which is causal and deviates from the desired phase characteristic by an amount which is linear with respect to frequency. Common examples include window filter designs with symmetric windows and Parks-McClellan filter designs. For an FIR filter, the linear phase component corresponds to a delay imposed on the ideal filter, and this delay is exact. The point of symmetry of the filter. For odd-length filters, the point of symmetry must be an integer, and for even-length filters the point of symmetry is a half integer.

Consider the case of continuous-time real filters which are frequency selective with a phase of zero or \(\pi\) for all frequencies. In other words, the analog frequency response \(H_a(\Omega)\) is of the form

\[
H_a(\Omega) = \text{Re} \left[ H_a(\Omega) \right].
\]  

(1)

The linear phase, or time-delayed, filter will have a frequency response

\[
\hat{H}_a(\Omega) = H_a(\Omega) e^{-jld}
\]  

(2)

where \(d\) is the delay. Therefore, we will call \(\hat{H}_a(\Omega)\) linear phase if it can be expressed as (1) and (2). Clearly, this definition does not include step changes in the phase other than by integer multiples of \(\pi\). It is easy to show that if \(H_a(\Omega)\) can be expressed as in (1) and (2), then \(\hat{h}_a(t)\) must be symmetric around \(d\); i.e.,

\[
\hat{h}_a(d + t) = \hat{h}_a(d - t).
\]

(3)

From this result, it is clear that the following hold.

- If a continuous-time linear phase filter is causal, it must be finite in duration, with the impulse response necessarily zero for \(t > 2d\).
- No one-sided, infinite impulse response (IIR), real linear phase filters exist in continuous time.

II. DISCRETE-TIME LINEAR PHASE FILTERS

For the discrete-time case, we will define the filters of interest to be real with desired phases of 0 or \(\pi\) for all frequencies:

\[
H(e^{j\omega}) = \text{Re} \left[ H(e^{j\omega}) \right].
\]

(4)

The linear phase filters have responses of the form

\[
\hat{H}(e^{j\omega}) = H(e^{j\omega}) e^{-jpd}
\]

(5)

where \(d\) is the delay. Since \(h(n)\) is real, \(H(e^{j\omega})\) is conjugate symmetric, and since \(H(e^{j\omega})\) is real, \(h(n)\) is even.

For \(d\) an integer, \(\hat{h}(n)\) can be expressed as \(h(n - d)\), which must be symmetric around \(d\):

\[
\hat{h}(d - n) = \hat{h}(d + n).
\]

(6)

For \(d\) a half integer, \(\hat{h}(n)\) must also be symmetric about \(d\), implying

\[
\hat{h}(d + 1/2 + n) = \hat{h}(d - 1/2 - n).
\]

(7)

We therefore see that in the discrete-time domain, if the delay is either an integer or a half integer, a symmetry must exist within the samples. A statement can therefore be made which is analogous to that made for continuous time.

- All real, causal, linear phase, discrete-time filters with an integer or half integer delay \(d\) must be finite in extent with unit sample responses of 0 for \(n > 2d\).

If \(d\) is neither an integer nor a half integer, no symmetry conditions as in (6) or (7) exist since \((d + r)\) and \((d - r)\) cannot simultaneously be integers for any \(r\). A slightly different symmetry condition exists, however.

Theorem: A real, discrete-time signal \(x(n)\) has linear phase if and only if

\[
x_d(t) = \sum_{n = -\infty}^{\infty} x(n) \frac{\sin \pi(t - n)}{\pi(t - n)}
\]

is symmetric about some point \(d\).
Proof:
Sufficiency: If \( x_\nu(d + t) = x_\nu(d - t) \), then, by (3),
\[
X_c(\Omega) = \begin{cases} R(\Omega) e^{-jkd} & \text{over } (-\pi, \pi) \\ 0 & \text{otherwise} \end{cases} \tag{8}
\]
where \( R(\Omega) \) is real. Since \( x(n) \) represents \( x_\nu(t) \) sampled at its Nyquist rate of one sample per unit time, \( X(e^{j\Omega}) = X_\nu(\omega) \) over \((-\pi, \pi)\). \( \tag{9} \)

Direct substitution gives
\[
X(e^{j\omega}) = R(e^{j\omega}) e^{-jkd}, \tag{10}
\]
which has linear phase.

Necessity: For \( x(n) \) to have linear phase, it must be expressible as in (10). The sampling theorem ensures that (8) holds in the reverse direction. Therefore, (7) would follow from (9). Since only symmetric analog filters have linear phase, \( x_\nu(t) \) must therefore be symmetric.

III. CAUSAL IIR LINEAR PHASE DISCRETE-TIME FILTERS

For causal linear phase filters with integer or half integer delays, the responses must be FIR. It is also well known that no stable causal filter with poles in its z-transform can have linear phase (excepting, of course, poles at \( z = 0 \) or \( z = \infty \)). One causal discrete-time signal which would appear to have linear phase is
\[
x(n) = (-a)^n u(n) \tag{11}
\]
in the limit as \( a \to 1 \) where \( u(n) \) is the unit step function. Here, \( X(e^{j\omega}) = 1/(1 + ae^{-j\omega}) \), which as \( a \to 1 \) is equal to \((2 \cos \omega/2)^{-1}e^{-j\omega/2}\). However, \( x(n) \) is neither square nor absolutely summable, and \( x_\nu(t) \), the continuous-time signal constructed, is neither absolutely nor square integrable. The absolute and square integrals of \( X(e^{j\omega}) \) over \((-\pi, \pi)\) also do not exist. In addition, \( x_\nu(t) \) is unbounded, a fact which can be readily seen from
\[
x_\nu(t) = \sum_{k=0}^{\infty} (-1)^k \sin \pi(t - k). \tag{12}
\]
The sidelobes of the individual terms constructively interfere to cause \( x_\nu(t) \) to be unbounded at least once per interval of length 1. Similar arguments can also be raised regarding other signals which have poles on the unit circle.

From the discussion in the preceding section, it is clear that if any continuous-time signal, band-limited to \((-\pi, \pi)\) and symmetric about time \( d \), is sampled with period 1, the resulting discrete-time signal will have linear phase. Imagine such a continuous-time signal which crosses through zero at times \((-1, -2, -3, \ldots)\) and \((1 + r, 2 + r, 3 + r, \ldots)\) where \( r \) is greater than \(-1\). The point of symmetry would be \( r/2 \), giving a causal, discrete-time signal with phase of \(-\omega r/2\). Naturally, the question arises as to whether or not such a signal exists, and if so, is it bounded, square integrable, and/or absolutely integrable? The first part of this question can be reduced to the following.

Does a continuous-time signal \( x_\nu(t) \) band-limited to \((-\pi, \pi)\) exist subject to the constraints
\[
x_\nu(t) = 0 \quad t = -1, -2, \ldots \tag{12}
\]
\[
x_\nu(t) = 0 \quad t = r + 1, r + 2, \ldots \tag{13}
\]
\[
x_\nu(t) = c \quad t = 0 \tag{14}
\]
where \( c \) is an arbitrary constant we will set to 1 for the remainder of our discussion? For \( x_\nu(t) \) to be symmetric and pass through 0 at the negative integers, the above constraints must hold. It may be possible that a nonsymmetric function will satisfy the constraints. If such a function were found, however, it would be a simple exercise to construct \( y_\nu(t) = x_\nu(t) + x_\nu(r - t) \), which would then satisfy all the criteria.

A first impression would suggest that there is such a function for \( r > 0 \). In the case \( r = 0 \), we have exactly one solution, that being \( x_\nu(t) = (\sin \pi t)/(\pi t) \). At \( r = 1 \), there is an infinity of solutions, all of the form
\[
x_\nu(t) = \frac{\sin \pi t}{\pi t} + x(1) \frac{\sin \pi(t - 1)}{\pi(t - 1)} \tag{15}
\]
where \( x(1) \) is arbitrary. The linear phase filter would then be \( x_\nu(t) \) and \( x_\nu(1 - t) \). For \( 0 < r < 1 \), there should be at least one solution, and for \( r > 1 \), there should be many.

Consider now the function
\[
s_r(t) = \frac{\Gamma(1 + r)}{\Gamma(1 + t) \Gamma(1 + r - t)} \tag{16}
\]
where \( \Gamma(t) \) is the gamma function. We hold that, under certain restrictions regarding the values of \( r \), \( s_r(t) \) satisfies all the constraints in (12)-(14). Fig. 1 depicts this function for typical values of \( r \).

Proof: First, since the function \((\Gamma(t))^{-1}\) is zero at all negative integers and zero, \( s_r(t) \) is zero at \((-1, -2, -3, \ldots)\) and \((1 + r, 2 + r, 3 + r, \ldots)\), and it is one at \( t = 0 \) by simple substitution.

Second, its band-limitedness can be shown by reference to two formulas found in a table of definite integrals [2]:
\[
\int_{-\pi}^{\pi} \frac{\sin \Omega dt}{\Gamma(p + t) \Gamma(q - t)} = \begin{cases} \frac{(2 \cos \Omega/2)^{p+q-2} \sin (\Omega(q - p)/2)}{\Gamma(p + q - 1)} & \text{if } |\Omega| < \pi \\ 0 & \text{if } |\Omega| > \pi \end{cases} \tag{17}
\]
$$s_r(t) = 0, \quad t = 0, -1, -2, \cdots$$  \hspace{1cm} (21)
$$s_r(t) = 0, \quad t = r + 1, r + 2, \cdots$$  \hspace{1cm} (22)

which can be added to $s_r(t)$. One such function is

$$w_r(t) = \frac{\Gamma(r)}{\Gamma(t) \Gamma(r - t + 1)}$$  \hspace{1cm} (23)

where

$$W_r(\Omega) = \left(2 \cos \frac{\Omega}{2}\right)^{r-1} e^{-\frac{\Omega}{2(r+1)^2}} e^{-\pi < \Omega < \pi}.$$  \hspace{1cm} (24)

Examination of $w_r(t)$ in a manner similar to that used for $s_r(t)$ shows $w_r(t)$ to be bounded for $r > 0$, square integrable for $r > 1/2$, and absolutely integrable for $r > 1$. We can therefore say $s_r(t)$ is not the unique bounded solution for $r > 0$, is not the unique square integrable solution for $r > 1/2$, and is not the unique absolutely integrable solution for $r > 1$.

We can construct

$$v_r(t) = s_r(t) - \frac{s_r(r)}{w_r(r)} w_r(t).$$  \hspace{1cm} (25)

This is nothing more than

$$v_r(t) = \frac{\Gamma(r)}{\Gamma(1 + t) \Gamma(r - t)},$$  \hspace{1cm} (26)

which has zero crossings at $t = -1, -2, \cdots$ and at $t = r, r + 1, \cdots$ and is one at $t = 0$. In other words, $s_r(t)$ has had its values of $r$ reduced by 1, placing a further constraint. Our earlier analysis, however, shows, for $0 < r < 1$, that $L_1$ stability of $v_r(t)$ will be lost, and for $0 < r < 1/2$, the $L_2$ stability of $v_r(t)$ will be lost. However, ignoring stability issues, is $s_r(t)$ the unique function band-limited to $(-\pi, \pi)$ satisfying constraints (22) for $-1 < r < 0$? We have been unable to prove the answer to this question, but strongly suspect the answer to be yes. If $r$ is set to zero, the function $s_r(t)$ becomes

$$s_0(t) = \frac{1}{\Gamma(1 + t) \Gamma(1 - t)} = \sin \frac{\pi t}{\pi}.$$  \hspace{1cm} (27)

This is, of course, known to be the unique band-limited function satisfying the constraints. Even here, the function is not absolutely integrable. If $r$ is made less than zero, even stronger constraints are placed on the function. If a function did exist satisfying the constraints, it would therefore probably be unique.

**IV. PROPERTIES AND SPECIAL CASES**

1) It has already been observed that $s_1(t) = \sin \frac{\pi t}{\pi}$.  
2) If $-1 < r < 0$, (20) shows that $S_r(\Omega)$ has a linear phase corresponding to an advance in time of $-r/2$. The maximum advance achievable with a bounded impulse re-
response is $1/2$; for a finite energy response it is $1/4$. The same statements can be made regarding $s_r(n)$. A causal filter allowing this behavior is interesting, especially since the filter is high-pass. The inverse for $s_r(n)$ can be gotten by noting that a change of sign of the filter is high-pass. Therefore, the inverse for $s_r(n)$, $-1 < r < 0$, is

$$s_{-r}(n) = s_r^{-1}(n) = \frac{\Gamma(1 - r)}{\Gamma(1 + n) \Gamma(1 - r - n)},$$

(28)

which is causal and $L_1$ and $L_2$ stable. We find it interesting that a causal linear phase discrete-time filter can have a causal linear phase discrete-time inverse. A similar property was noted by Fettweis for maximally flat digital filters [6].

3) For $r$ of the form $1/q$, with $q$ a positive integer,

$$S_r^q(e^{j\omega}) = \left(2 \cos \frac{\omega}{2}\right) e^{-j\omega/2} - \pi < \omega < \pi,$$  

(29)

or

$$s_r(n) * s_r(n) * \cdots * s_r(n) = \left[\delta(n) + \delta(n - 1)\right].$$

(30)

4) For $r$ of the form $-1/q$, with $q$ a positive integer,

$$S_r^q(e^{j\omega}) = \left(2 \cos \frac{\omega}{2}\right)^{-1} e^{j\omega/2} - \pi < \omega < \pi,$$  

(31)

or

$$s_r(n) * s_r(n) * \cdots * s_r(n) = (-1)^n u(n).$$

(32)

This is an example, for $-1/2 < r < 0$, showing an $L_2$ stable system, when cascaded with itself is no longer $L_2$ stable.

5) More generally,

$$S_r(e^{j\omega}) S_0(e^{j\omega}) = S_{r+0}(e^{j\omega})$$

(33)

and

$$S_0^q(e^{j\omega}) = S_0^q(e^{j\omega}) = S_r^q(e^{j\omega})$$

(34)

where the region of convergence includes the unit circle for $r > 0$ and does not otherwise. We see that if $z = e^{j\omega}$, $S_r(z)$ reduces to $S_r(e^{j\omega})$ where

$$s_r(e^{j\omega}) = \left(2 \cos \frac{\omega}{2}\right)^{r} e^{-j\omega r}.$$  

(37)

For $r$ not an integer, $(2 \cos (\omega/2))^r$ will be multivalued. However, since $S_r(z)$ only has one branch cut, along the ray $\omega = \pi$, we can stay on the principal branch and hence can choose a real positive value for $(2 \cos (\omega/2))^r$. In contrast, however, consider a slightly different function:

$$H(z) = (1 + z^{-1})^r.$$  

(38)

giving

$$H(e^{j\omega}) = e^{-j\omega}(2 \cos \omega)^r.$$  

(39)

A first impression might convince one that $h(n)$ is also linear phase. This is not necessarily true, however. Consider the case of $r = 1/2$. As $\omega$ passes through $\pi/2$, a new branch of $H(z)$ must be entered, and $(2 \cos (\omega))^1/2$ cannot be real. Stated more generally, the function

$$H(z) = [H_0(z)]^r$$

(40)

with a region of convergence $|z| > |a|$ where $H_0(z)$ corresponds to a linear phase FIR filter with delay $d$ will not usually produce a linear phase filter. Although the Fourier transform

$$H(e^{j\omega}) = [H_0(e^{j\omega})]^r = [R_0(e^{j\omega})]^r e^{-j\omega d}$$

(41)

where $R_0(e^{j\omega})$ is real appears to be of the form (5), $H(z)$ will usually have branch points on the unit circle. Hence, to maintain analyticity of $H(z)$ along the unit circle, different Riemann surfaces, or branches, must be entered, rendering $R_0(e^{j\omega})^r$ no longer real over the entire range $(-\pi, \pi)$ [4]. Notable exceptions occur when $r$ is a positive integer, when

$$H_0(z) = (1 + z^{-1})^k, \quad k = 1, 2, 3, \cdots$$

(42)

or when $r = 1/p$ and

$$H(z) = [H_0(z)]^p.$$  

(43)

If only FIR linear phase filters are selected for which $H_0(e^{j\omega})$ is of the form in (41) and $R_0(e^{j\omega})$ is always real and positive, it is possible to remain on one branch for $[R_0(e^{j\omega})]^r$. However, the inverse transform corresponding to the stable filter is generally noncausal, and the causal inverse transform is generally unstable.

7) $s_r(n)$ with $r = 1/2$ has the interesting property that $s_r(n) * s_r(n)$ is zero everywhere except $n = 0$ and $n = 1$, despite $s_r(n)$'s infinite extent. Therefore, if $s_r(n)$ were truncated at $N$ and convolved with itself, the result would be zero except at $n = 0$, $n = 1$, and $n \geq N$. In continuous time, $s_r(t) * s_r(t)$ for $r = 1/2$ is oscillatory with zero crossings at $t = -1, +2, +3, \cdots$ and is one at $t = 0$ and $t = 1$.

8) The signal $s_r(t)$ has a mainlobe width of $2 + r$. This, coupled with property 5, shows $s_r(t) * s_r(t)$ has a mainlobe width of $\nu + \theta + 2$ with zero crossings every one time unit.

9) The results shown can serve to give the evaluation of the below-listed definite integral, which we were unable to find in any table:

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \left(2 \cos \frac{Q}{2}\right)^r e^{j\omega t} d\Omega$$

$$= \frac{\Gamma(1 + r)}{\Gamma(1 + r/2 + i) \Gamma(1 + r/2 - i)}.$$  

(44)
V. Summary

We have explored an interesting family of functions for which the \( \sin \pi t / \pi t \) function is a special case. We have shown that the sampled versions of these signals are causal and linear phase and, under certain conditions, are bounded, square integrable, and absolutely integrable. We have not explored to any great extent the possible applications, although some immediately come to mind. Band-limited extrapolation is one possible application area. The main focus of this paper has been of a theoretical nature. Although the nonrationality of the systems described render them unrealizable, the theoretical tools developed are similar to those that use the \( \sin \phi / \phi \) function and could potentially be equally useful.

Appendix

We wish to show the ranges over which the functions \( s_r(t) \) and \( s_r(n) \) are bounded, \( L_2 \) stable, and \( L_1 \) stable.

To show boundedness, we define a function

\[
g_r(t) = \frac{s_r(t)}{\Gamma(1 + r)} = \frac{1}{\Gamma(1 + r) \Gamma(1 + r - t)}
\]

for convenience. (Boundedness need only be considered for \( t > 0 \) due to symmetry.) We can rewrite \( g_r(t) \) as

\[
g_r(t) = \frac{(r - t + 1)(r - t + 2) \cdots (\phi - 1)}{(r + 1) \cdots \Gamma(r) \Gamma(\phi)} (A1)
\]

where \( \theta = t - \text{int}(t) + 1 \) and \( \phi = 1 + r - t + \text{int}(t) \).

\[
|g_r(t)| = \left| \frac{r - t + 1}{r + 1} \right| \left| \frac{r - t + 2}{r + 1} \right| \cdots \frac{1 - \phi}{\Gamma(\theta) \Gamma(\phi)} \cdot (A2)
\]

For \( r \geq -1 \), all terms in the numerator can be matched with a term in the denominator which is at least as large, rendering it impossible to be unbounded. However, if \( r < -1 \), \( |g_r(t)| \) can be arbitrarily large for large enough \( t \) and is hence unbounded.

Square integrability can be shown from the integral table formula [2]:

\[
\int_{-\infty}^{\infty} \frac{dt}{\Gamma(\alpha + t) \Gamma(\beta - t)} = \frac{\Gamma(2\alpha + 2\beta - 3)}{\Gamma(\alpha + \beta - 1)} \quad \alpha + \beta > 3/2. \quad (A3)
\]

In our case, since \( \alpha = 1 \) and \( \beta = 1 + r \), \( s_r(t) \) will be square integrable for \( r > -1/2 \).

Showing absolute integrability of \( s_r(t) \) or \( g_r(t) \) is somewhat more involved.

Due to the symmetry and boundedness of \( g_r(t) \) for \( r > -1 \), we need only consider positive values of \( t \) which are larger than some arbitrary amount \( t_0 \).

In other words, we need only to show

\[
\int_{t_0}^{\infty} |g_r(t)| \, dt < \infty. \quad (A4)
\]

Referring to (A2), we can rewrite

\[

\[
g_r(t) = \frac{(1 - 1 - r) \cdots (1 - \phi) \Gamma(1 - \phi)}{\Gamma(t + 1) \Gamma(1 - \phi)} \Gamma(\phi) \quad (A5)
\]

\[
|g_r(t)| = \left| \frac{\Gamma(t - r)}{\Gamma(t + 1) \Gamma(1 - \phi)} \right| \quad (A6)
\]

It can be shown that

\[
\frac{1}{|\Gamma(\phi) \Gamma(1 - \phi)|} \leq \pi^{-1} \quad (A7)
\]

with equality at \( \phi = 1/2 \) and values of zero at \( \phi = 0 \) and \( \phi = 1 \). It follows, therefore, that

\[
|g_r(t)| < \frac{\Gamma(t - r)}{\Gamma(t + 1)} \quad (A8)
\]

It can be shown [5] that

\[
\lim_{t \to \infty} \frac{\Gamma(t - r)}{\Gamma(t + 1)} = 1, \quad (A9)
\]

implying that

\[
\lim_{t \to \infty} \frac{\Gamma(t - r)}{\Gamma(t + 1)} \cdot t^{-r} = 1 \quad (A10)
\]

or

\[
\lim_{t \to \infty} \left| \frac{\Gamma(t - r)}{\Gamma(t + 1)} \right| \cdot t^{-r} = 1. \quad (A11)
\]

For this to be true, the term

\[
\left| \frac{\Gamma(t - r)}{\Gamma(t + 1)} \right|
\]

must asymptotically decay at a rate no slower than \( t^{-r-1} \) and no faster than \( t^{-r-1+\epsilon} \) where \( \epsilon \) is an arbitrarily small positive value. Therefore, \( s_r(t) \) and \( g_r(t) \) are absolutely integrable for \( r > 0 \) and are not for \( r < 0 \). For \( r = 0 \), \( s_r(t) = \sin \pi t / \pi t \), which we know is not absolutely integrable.

The sampled version of \( s_r(t) \), denoted \( s_r(n) \), must be bounded, square summable, and absolutely summable, at least over the range of \( r \) for which \( s_r(t) \) is bounded, square integrable, and absolutely integrable, respectively. Boundedness transferring from \( s_r(t) \) to \( s_r(n) \) is obvious. The reverse need not be true, however, and we find, for \( r = -1 \), \( s_r(n) \) is bounded. The bounds expressed for \( |s_r(t)| \) must also apply to \( |s_r(n)| \), showing \( |s_r(n)| \) to be summable for \( r > 0 \) and nonsummable for \( r < 0 \). For \( r = 0 \), however, \( |s_r(n)| \) becomes \( |\sin \pi t / \pi t| = \delta(n) \), which is summable. Similarly, the same bounds show, for
\( r > -1/2 \), \( s^2(n) \) is summable and, for \( r < -1/2 \), it is not. For \( r = -1/2 \), Parseval's relation applied to (20) shows \( s^2(n) \) is not summable for \( r = -1/2 \).

Mark A. Clements (S‘81–M‘82) received the S.B., S.M., and Sc.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, in 1976, 1978, and 1982, respectively. Since 1982 he has been with the School of Electrical Engineering at the Georgia Institute of Technology, Atlanta, where he is an Associate Professor. His current research interests include speech enhancement, automatic speech recognition, speech modeling, and aids for the hearing-impaired.

Jerrold W. Pease received the B.E.E. degree from the Georgia Institute of Technology, Atlanta, in 1986. He has worked as an Engineer for Boeing Corporation and is currently working as an independent software consultant in Huntsville, AL. His interests include communications, signal processing, and linguistics.