Why linear phase real FIR filters are symmetric

Diogo Antunes

July 2021

1 Introduction

The aim of this text is to justify the necessary nature of well known symmetry characteristics of real and finite impulse response filters with generalised linear phase, using arguments that are hopefully intuitive, while relying on common complex analysis results. This text does not explore practical and useful implementation details like [1] does. Instead, this text focuses only on justifying why the symmetry properties of linear phase FIR filters are necessary conditions for linear phase. In detail, the filters considered are those that satisfy the 3 following conditions:

The filter's impulse response, $h[n]$, must be finite, meaning that

$$\exists_{L\in\mathbb{Z}}\ |n| > L \Rightarrow h[n] = 0. \tag{1}$$

The filter's impulse response must be real, so the filter's frequency response, $H(j\Omega)$, must satisfy

$$H(j\Omega) = H^*(-j\Omega). \tag{2}$$

Finally, the filter must have generalised linear phase as defined in [1], so its frequency response must be of the form

$$H(j\Omega) = A(j\Omega)\mathrm{e}^{j(\beta - k\Omega)}, \tag{3}$$

where $A(j\Omega)$ is real valued (possibly taking positive and negative values) and k and β are real constants. Later it will be shown that k and β necessarily take specific values.

The goal behind the use of condition (3) is to include a broader class of filters than just linear phase filters, while maintaining the symmetry properties. Notice that it means the phase response has constant group delay in all frequencies except for those where $A(j\Omega)$ changes sign and so the phase has a discontinuity of π radians. It turns out, as seen later, that (1) implies $A(j\Omega)$ is continuous, and so the filter's phase response has constant group delay between the filter's zeros, for example, in the passband.

Figure 1 shows one example of a filter that satisfies the 3 conditions.

2 FIR filter properties

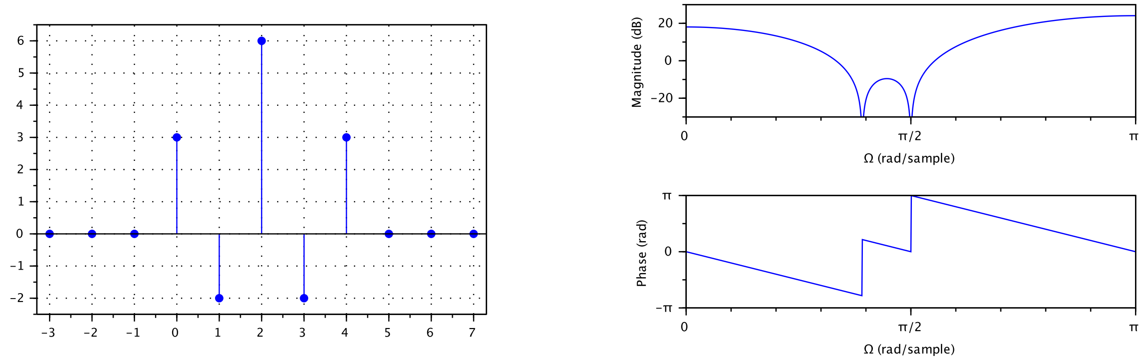
In this section, properties for arbitrary FIR and real FIR filters are obtained.

A FIR filter's transfer function can be obtained as the Discrete Time Fourier Transform (DTFT) of its impulse response, $h[n]$,

$$H(j\Omega) = \sum_{n=-L}^{L} h[n]\mathrm{e}^{-j\Omega n}. \tag{4}$$

As a finite sum of exponentials, $H(j\Omega)$ is an holomorphic function, so it is continuous, infinitely differentiable and analytic.

$H(j\Omega)$ being continuous implies that its magnitude, $|H(j\Omega)|$, is continuous. For every frequency that $H(j\Omega) \neq 0$, its phase is also continuous. This can be expected intuitively by predicting that the only way for a complex point to move continuously and change its phase discontinuously is for it to pass through 0.



(a) Impulse response

(b) Magnitude and Phase responses

Figure 1: Example filter that satisfies (1) through (3). This filter has $k = 2$ and $\beta = 0$. Notice the discontinuities of π radians in the phase at the zeros of the filter.

Rigorously, phase can be apparently tricky, but the problems are just superficial. One could define a discontinuous phase function of complex $z \neq 0$, $\angle(z)$, where $\angle(z)$ returns phase values in the range $]-\pi, \pi]$. This phase function is discontinuous in the negative part of the real axis, where the phase jumps between π and $-\pi$. Given a continuous function $f(\Omega) \in \mathbb{C}$, where $\Omega \in \mathbb{R}$, then $\phi(\Omega) = \angle(f(\Omega))$ will also be continuous when $f(\Omega)$ is neither negative nor zero. At the points where $f(\Omega)$ is negative, the phase might have a jump of 2π. These discontinuities can be removed by adding or subtracting 2π to the phase at these discontinuity points, resulting in a phase function that is not limited to a range and is continuous where $f(\Omega) \neq 0$.

Another way to think about phase rigorously might be to define a set of numbers in which $a = a + 2\pi$, reflecting the sense that they *mean* the same phase, avoiding the problem of discontinuity for negative inputs of $\angle(z)$ in the first place by making $-\pi = \pi$. Whatever formulation is used, the phase of a continuous complex function $f(\Omega)$ of a real variable Ω can be made continuous where $f(\Omega) \neq 0$.

Notice that where $f(\Omega) = 0$, it doesn't have a unique phase. One could leave the phase undefined, or simply hold the previous value of $\phi(\Omega)$, in order to keep the phase function continuous where possible. For example, the phase of Ω^2 at $\Omega = 0$ could be specified as any number c and it would still be true that $0^2 = 0\mathrm{e}^{jc}$, but holding the phase 0 keeps $\phi(\Omega)$ continuous in $\Omega \in \mathbb{R}$. For Ω^3 however, no specified phase for $\Omega = 0$ can make the phase continuous, as the limits from either side are different. One could call the phase undefined or discontinuous at that point. The term discontinuities will be used to refer to these points where the phase limits from each side of the zero of $f(\Omega)$ are different.

In this text, the rigorous details of the formulation of phase are ignored, as they are irrelevant to the resulting complex number $A(j\Omega)\mathrm{e}^{j\phi(j\Omega)}$.

$H(j\Omega)$ being analytic means that it is locally equal to its Taylor series. Consequently, near a frequency Ω_0 such that $H(j\Omega_0) = H_0$, $H(j\Omega)$ can be approximated asymptotically with its first derivative that is not 0 at $\Omega = \Omega_0$. For $\Omega \to \Omega_0$, the transfer function approaches

$$H(j\Omega) \to H_0 + c_m(\Omega - \Omega_0)^m,$$

where $m > 0$ is the order of the first non-zero derivative and c_m is the corresponding Taylor's coefficient. Since Ω is real, $H_0 + c_m(\Omega - \Omega_0)^m$ is restrained to the line that passes through $H_0 + c_m$ and H_0. This means that at every frequency, the trajectory of $H(j\Omega)$ has a tangent line, so no sudden changes in direction.

Applying this result specifically to points where $H_0 = 0$, as Ω passes through 0, if m is even, $H(j\Omega)$ will approach and leave 0 from the same phase, so the phase stays continuous. If m is odd, $H(j\Omega)$ will leave 0 from the opposite direction it approached it from, so the phase will have a discontinuity of π radians.

We conclude that a FIR filter's magnitude response is continuous and that its phase response is continuous, except at the zeros of the magnitude, where it might have discontinuities of π radians. As a consequence, a

FIR filter's transfer function is of the form

$$H(j\Omega) = A(j\Omega)\mathrm{e}^{j\phi(j\Omega)}, \tag{5}$$

where $A(j\Omega)$ is real and continuous, it specifies the function's magnitude and changes sign whenever the total phase has a discontinuity, and $\phi(j\Omega)$ is a continuous and real function that specifies the continuous variation of the total phase response.

If the filter has real coefficients, then $H(j\Omega) = H^*(-j\Omega)$ and $\angle(H(j\Omega)) = -\angle(H(-j\Omega))$. In this text, filters are separated into 2 Types according to the continuity of the phase at $\Omega = 0$, which will be referred to as Type C (continuous) and Type D (discontinuous):

- Type C: The phase is continuous at $\Omega = 0$, so $\angle(H(j0)) = 0$ or $\angle(H(j0)) = \pi$. These are the only two values that satisfy $\angle(H(j0)) = -\angle(H(j0))$, where one must remember that the phases π and $-\pi$ are the same.

- Type D: The phase is discontinuous at $\Omega = 0$. Since only discontinuities of π radians are possible, $\angle(H(j0^+)) = \pm\frac{\pi}{2}$. These are the possible values since $\angle(H(j0^+)) = \angle(H(j0^-))\pm\pi = -\angle(H(j0^+))\pm\pi$.

As a result, any filter with a real and finite impulse response has a phase response that starts either at 0, π (Type C) or $\pm\frac{\pi}{2}$ (Type D). One important practical detail is that Type D filters must have a zero at $\Omega = 0$ in order to have the phase discontinuity at $\Omega = 0$.

3 Real FIR filters with generalised linear phase

In this section, properties for real FIR filters with generalised linear phase, conditions (1) through (3), are obtained.

It is possible to separate the filter's transfer function into a filter without delay and an ideal delay, writing $H(j\Omega) = A(j\Omega)\mathrm{e}^{j(\beta-k\Omega)}$ as $G(j\Omega)\mathrm{e}^{-jk\Omega}$, where $G(j\Omega) = A(j\Omega)\mathrm{e}^{j\beta}$. Since $H(j\Omega)$ and $\mathrm{e}^{-jk\Omega}$ are both conjugate symmetric, so does $G(j\Omega)$ satisfy

$$G(j\Omega) = G^*(-j\Omega). \tag{6}$$

For filters of Type C, since the total phase at $\Omega = 0$ is either 0 or π, $H(j0)$ is real and β can be taken as 0. This means that $G(j\Omega) = A(j\Omega)$, so $G(j\Omega)$ is real valued, which coupled with (6) means that $G(-j\Omega) = G(j\Omega)$. This implies that the corresponding impulse response, $g[n]$, is symmetric around $n = 0$, since

$$g[-n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} G(j\Omega)\mathrm{e}^{-j\Omega n}d\Omega = \frac{1}{2\pi}\int_{-\pi}^{\pi} G(-j\Omega)\mathrm{e}^{-j\Omega n}d\Omega,$$

using the substitution $\Omega' = -\Omega$, this results in

$$g[-n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} G(j\Omega')\mathrm{e}^{j\Omega' n}d\Omega' = g[n]. \tag{7}$$

For filters of Type D, since the total phase at $\Omega = 0$ is $\pm\frac{\pi}{2}$, $H(j0)$ is imaginary and β has to be $\pm\frac{\pi}{2}$. $A(j\Omega)$ must change sign at $\Omega = 0$ in order to $G(j\Omega) = A(j\Omega)\mathrm{e}^{\pm j\pi/2}$ satisfy (6). $G(j\Omega)$ is purely imaginary, which coupled with (6), implies that $G(-j\Omega) = -G(j\Omega)$. Using the same argument as in (7), the corresponding impulse response, $g[n]$, is anti-symmetric around $n = 0$, so

$$g[-n] = -g[n]. \tag{8}$$

To summarise, real FIR filters with generalised linear phase are equivalent to the series of a filter $G(j\Omega)$ with an ideal delay of k samples. For Type C filters, $\beta = 0$ and $G(j\Omega)$ has an even impulse response. For Type D filters, $\beta = \pm\frac{\pi}{2}$ and $G(j\Omega)$ has an odd impulse response.

$h[n]$ is the result of applying the delay of transfer function $\mathrm{e}^{-jk\Omega}$ to $g[n]$, which is said to be a delay of k samples, even if $k \notin \mathbb{Z}$. The meaning of a delay of a real number of samples is given in Annex A. Simply put, the filter with frequency response $\mathrm{e}^{-jk\Omega}$ corresponds to creating the continuous, bandlimited interpolation of the discrete time input, delaying the equivalent of k samples and then re-sampling the continuous signal into discrete time.

If $k \in Z$, $\mathrm{e}^{-jk\Omega}$ corresponds simply to a discrete time delay, and so $h[n] = g[n-k]$, which means that $h[n]$ is symmetric (Type C) or anti-symmetric (Type D) around $n = k$.

For non-integer k, it helps the define the continuous and bandlimited interpolation of $g[n]$, with period $T = 1$, as

$$g_{\mathrm{c}}(t) = \sum_{n=-\infty}^{+\infty} g[n]\mathrm{sinc}\,(t-n)\,. \tag{9}$$

This means that $g[n] = g_{\mathrm{c}}(n)$. Define also the corresponding signal $h_{\mathrm{c}}(t)$ for $h[n]$. Considering that $h[n]$ is $g[n]$ delayed k samples, then $h_{\mathrm{c}}(t) = g_{\mathrm{c}}(t-k)$. Note that since $h[n]$ is finite, meaning limited in time, $h_{\mathrm{c}}(t)$ is a finite sum of sinc functions

$$h_{\mathrm{c}}(t) = \sum_{n=-L}^{+L} h[n]\mathrm{sinc}\,(t-n)\,. \tag{10}$$

As previously concluded, $g[n]$ is either even (Type C) or odd (Type D), and so $g_{\mathrm{c}}(t)$ will share that property. As $h_{\mathrm{c}}(t) = g_{\mathrm{c}}(t-k)$, $h_{\mathrm{c}}(t)$ is symmetric (Type C) or anti-symmetric (Type D) around $t = k$, so

$$h_{\mathrm{c}}(t) = \pm h_{\mathrm{c}}(2k-t). \tag{11}$$

Note that $h[n] = h_{\mathrm{c}}(n) = 0$ for $n > L$, from (1), so it is also that $h_{\mathrm{c}}(2k-n) = 0$ for $n > L$. If $2k \notin \mathbb{Z}$, so will $2k - n \notin \mathbb{Z}$, and these zeros will not occur at integer times. Consequently, $h_{\mathrm{c}}(t)$ must have infinite zeros at times that the individual sinc functions aren't trivially 0, so they must cancel each other out. This is impossible with a finite number of sinc functions, as a sum of N sinc functions in the typical equally spaced arrangement can only have $N-1$ non-trivial zeros.

This is briefly proved in Annex B, by showing that non-trivial zeros of signals like $h_{\mathrm{c}}(t)$, sums of N sinc functions, are also zeros of a polynomial of degree $N-1$.

As a result, it is impossible for $h_{\mathrm{c}}(t)$ to have infinite non-trivial zeros, which in our case are zeros for times $t \notin \mathbb{Z}$. As infinite such zeros are necessary if $2k \notin \mathbb{Z}$, $2k$ must be an integer.

Since $2k \in \mathbb{Z}$, then from (11) it follows that $h[n]$ will also be (anti-)symmetrical around k, as

$$h[n] = h_{\mathrm{c}}(n) = \pm h_{\mathrm{c}}(2k-n) = \pm h[2k-n]. \tag{12}$$

4 Conclusion

This text shows, using common results of complex analysis, well known properties of real FIR filters with generalised linear phase, according to (1) through (3). Specifically, it is seen that such a filter must have $2k \in \mathbb{Z}$ and be of one of two types.

The first type has $\beta = 0$ and its impulse response, $h[n]$, is symmetric around $n = k$, so $h[n] = h[2k-n]$.

The second type has $\beta = \pm\frac{\pi}{2}$ and an anti-symmetric impulse response around $n = k$, so $h[n] = -h[2k-n]$.

References

[1] Discrete-Time Signal Processing: Alan V. Oppenheim & Ronald W. Schafer 2010 Pearson, 3rd Edition.

Annex A

Ideal non-integer delay

A delay of integer k samples has a frequency response of $\mathrm{e}^{-jk\Omega}$. The same frequency response but with more general, real k can be interpreted as a delay of k samples to the equivalent bandlimited continuous signal. More precisely, the system with transfer function $\mathrm{e}^{-jk\Omega}$ is equivalent to creating the continuous and bandlimited, to the Nyquist frequency, interpolation of the discrete input signal with any period T, applying a continuous time delay of kT seconds and finally re-sampling the output with the same sampling period T.

To check this is the case, consider starting with a discrete signal $x[n]$ with Discrete Time Fourier Transform $X[j\Omega]$. Then create the corresponding bandlimited continuous signal with period $T = 1/f_\mathrm{s} = 2\pi/\omega_\mathrm{s}$, given by

$$x_T(t) = \sum_{n=-\infty}^{+\infty} x[n]\mathrm{sinc}\left(\frac{t-nT}{T}\right), \tag{13}$$

with Fourier Transform given by

$$\begin{aligned} X_T(j\omega) = \mathcal{F}\left\{x_T(t)\right\} = \mathcal{F}\left\{\sum_{n=-\infty}^{+\infty} x[n]\mathrm{sinc}\left(\frac{t-nT}{T}\right)\right\} = \\ = \sum_{n=-\infty}^{+\infty} x[n]\mathcal{F}\left\{\mathrm{sinc}\left(\frac{t-nT}{T}\right)\right\} = \sum_{n=-\infty}^{+\infty} x[n]\mathcal{F}\left\{\mathrm{sinc}\left(\frac{t}{T}\right)\right\}\mathrm{e}^{-j\omega nT} = \\ = \sum_{n=-\infty}^{+\infty} x[n]T\mathrm{rect}\left(\frac{\omega}{\omega_\mathrm{s}}\right)\mathrm{e}^{-j\omega nT} = T\mathrm{rect}\left(\frac{\omega}{\omega_\mathrm{s}}\right)\sum_{n=-\infty}^{+\infty} x[n]\mathrm{e}^{-j\omega nT} = \\ = T\mathrm{rect}\left(\frac{\omega}{\omega_\mathrm{s}}\right)X(j\omega T). \end{aligned} \tag{14}$$

As intended, we can check the continuous signal created is bandlimited to $\omega \in [-\omega_\mathrm{s}/2,\ \omega_\mathrm{s}/2]$.

Now apply the continuous delay of kT seconds to obtain the delayed signal with Fourier Transform $Y_T(j\omega) = X_T(j\omega)\mathrm{e}^{-jkT\omega}$.

Finally, re-sample the delayed signal with sampling period T to obtain the discrete output $y[n] = y_T(nT)$ with Discrete Time Fourier Transform given by

$$\begin{aligned} Y(j\Omega) = \sum_{n=-\infty}^{+\infty} y_T(nT)\mathrm{e}^{-j\Omega n} = \sum_{n=-\infty}^{+\infty} \frac{1}{2\pi}\int_{-\infty}^{+\infty} Y_T(j\omega)\mathrm{e}^{j\omega nT}d\omega\ \mathrm{e}^{-j\Omega n} = \\ = \sum_{n=-\infty}^{+\infty} \frac{1}{2\pi}\int_{-\omega s/2}^{+\omega_\mathrm{s}/2} TX(j\omega T)\mathrm{e}^{-jkT\omega}\mathrm{e}^{j\omega nT}d\omega\ \mathrm{e}^{-j\Omega n} = \\ = \frac{1}{2\pi}\int_{-\omega s/2}^{+\omega_\mathrm{s}/2} TX(j\omega T)\mathrm{e}^{-jkT\omega}\sum_{n=-\infty}^{+\infty}\mathrm{e}^{j(\omega T-\Omega)n}d\omega. \end{aligned}$$

Noticing that $\sum_{n=-\infty}^{+\infty}\mathrm{e}^{j(\omega T-\Omega)n}$ is the DTFT of $\mathrm{e}^{j\omega Tn}$ and knowing it is equal to $2\pi\sum_{l=-\infty}^{\infty}\delta(\Omega-\omega T-2l\pi)$, $Y(j\Omega)$ can be obtained as

$$Y(j\Omega) = \int_{-\omega s/2}^{+\omega_\mathrm{s}/2} TX(j\omega T)\mathrm{e}^{-jkT\omega}\sum_{l=-\infty}^{\infty}\delta(\Omega-\omega T-2l\pi)d\omega =$$

$$= \sum_{l=-\infty}^{\infty} \int_{-\omega s/2}^{+\omega_s/2} TX(j\omega T) \mathrm{e}^{-jkT\omega} \delta(\Omega - 2l\pi - \omega T) d\omega.$$

Changing integration variable to $\Omega' = \omega T + 2l\pi$, to more clearly use the Dirac delta's integration property, the result is

$$Y(j\Omega) = \sum_{l=-\infty}^{\infty} \int_{-\pi+2l\pi}^{\pi+2l\pi} TX(j(\Omega' - 2l\pi)) \mathrm{e}^{-jk(\Omega'-2l\pi)} \delta(\Omega - \Omega') \frac{d\omega}{d\Omega'} d\Omega' =$$

$$= \sum_{l=-\infty}^{\infty} \int_{-\pi+2l\pi}^{\pi+2l\pi} TX(j(\Omega' - 2l\pi)) \mathrm{e}^{-jk(\Omega'-2l\pi)} \delta(\Omega - \Omega') \left(\frac{1}{T}\right) d\Omega'.$$

Since both $X(j\Omega)$ and the exponential have period 2π, the terms $2l\pi$ can be ommited, giving

$$Y(j\Omega) = \sum_{l=-\infty}^{\infty} \int_{-\pi+2l\pi}^{\pi+2l\pi} X(j\Omega') \mathrm{e}^{-jk\Omega'} \delta(\Omega - \Omega') d\Omega'.$$

Notice that all the integration regions are contiguous and they span from $-\infty$ to ∞, so the sum of integrals can be replaced by a single integral spanning $[-\infty, \infty]$, which reduces to

$$Y(j\Omega) = \int_{-\infty}^{\infty} X(j\Omega') \mathrm{e}^{-jk\Omega'} \delta(\Omega - \Omega') d\Omega' = X(j\Omega) \mathrm{e}^{-jk\Omega}.$$

To conclude, for any real k, the overall transfer function of the system that creates the continuous, bandlimited interpolation of a discrete time signal, delays the equivalent of k samples and then samples the continuous signal again into discrete time, is simply $\mathrm{e}^{-jk\Omega}$. Examples are given in Figure 2.

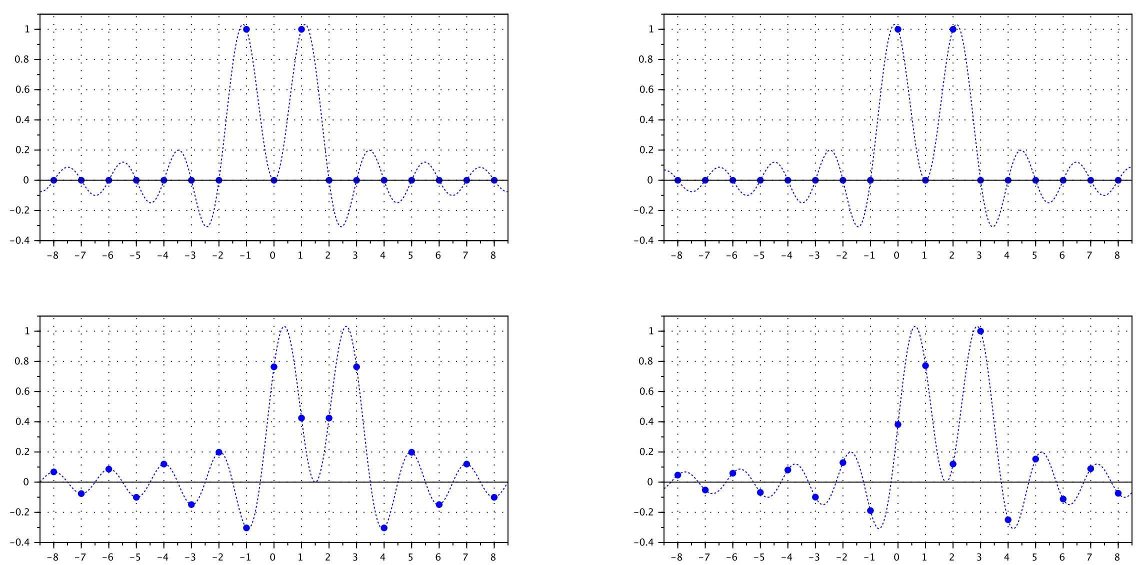


Figure 2: Examples of ideal delays: a) original signal, b) 1 sample delay, c) 1.5 sample delay and d) 1.75 sample delay. Both the discrete signal and the corresponding continuous interpolation are plotted.

Annex B

Non trivial zeros of a finite sum of sinc functions

Consider a finite sum of N weighted sinc functions, equally spaced in time such that at the centre of one sinc, all the others are 0,

$$x(t) = \sum_{n=1}^{N} a[n]\mathrm{sinc}(t-n). \tag{15}$$

Such a function has an infinite number of trivial zeros in the times that all the sinc functions are 0, which occur at integer times. However, it cannot have more than $N-1$ zeros for times $t \notin \mathbb{Z}$, at which the individual sinc functions are not 0.

To prove this is the case, we can check that such zeros of $x(t)$ are the zeros of a polynomial of degree $N-1$. Specifically,

$$x(t) = \sum_{n=1}^{N} a[n]\mathrm{sinc}(t-n) = \sum_{n=1}^{N} a[n]\frac{\sin(\pi(t-n))}{\pi(t-n)} = \sum_{n=1}^{N} a[n]\frac{\sin(\pi t)(-1)^n}{\pi(t-n)}.$$

Since $t \notin Z$, then $\sin(\pi t) \neq 0$. Dividing the equation by $\sin(\pi t)/\pi$, we get

$$x(t) = 0 \Leftrightarrow \sum_{n=1}^{N} a[n]\frac{(-1)^n}{t-n} = 0.$$

Multiplying the equation by $(t-n)$ for each value of $n \in [1, N]$, we get

$$x(t) = 0 \Leftrightarrow \sum_{n=1}^{N} a[n](-1)^n \prod_{m\neq n}(t-m) = 0. \tag{16}$$

As a result, any non-integer zero of $x(t)$ must also be a zero of (16), which is a polynomial of order $N-1$ and so has only $N-1$ complex zeros. As a consequence, $x(t)$ must also have $N-1$ complex zeros such that $t \notin \mathbb{Z}$.