Transform analysis of LTI systems

Oppenheim and Schafer, Second edition pp. 240–339.

For LTI systems we can write

$$y[n] = x[n] * h[n] = \sum_{k=-\infty}^{\infty} x[k]h[n-k].$$

Alternatively, this relationship can be expressed in the z-transform domain as

$$Y(z) = H(z)X(z),$$

where $H(z)$ is the **system function**, or the z-transform of the system impulse response.

Recall that a LTI system is completely characterised by its impulse response, or equivalently, its system function.

1 Frequency response of LTI systems

The frequency response $H(e^{j\omega})$ of a system is defined as the gain that the system applies to the complex exponential input $e^{j\omega n}$. The Fourier transforms of the system input and output are therefore related by

$$Y(e^{j\omega}) = H(e^{j\omega})X(e^{j\omega}).$$

In terms of magnitude and phase,

$$|Y(e^{j\omega})| = |H(e^{j\omega})||X(e^{j\omega})|$$
$$\sphericalangle Y(e^{j\omega}) = \sphericalangle H(e^{j\omega}) + \sphericalangle X(e^{j\omega}).$$

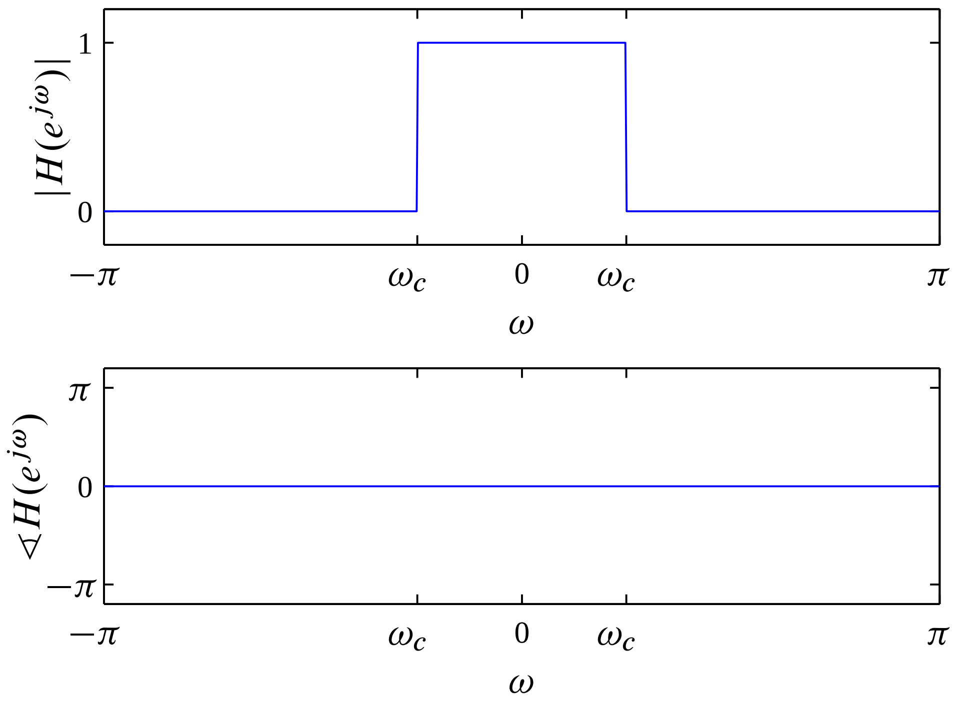
In this case $|H(e^{j\omega})|$ is referred to as the **magnitude response** or **gain** of the system, and $\sphericalangle H(e^{j\omega})$ is the **phase response** or **phase shift**.

1.1 Ideal frequency-selective filters

Frequency components of the input are suppressed in the output if $|H(e^{j\omega})|$ is small at those frequencies. The **ideal lowpass filter** is defined as the LTI system with frequency response

$$H_{\text{lp}}(e^{j\omega}) \begin{cases} 1 & |\omega| \leq \omega_c \\ 0 & \omega_c < |\omega| \leq \pi. \end{cases}$$

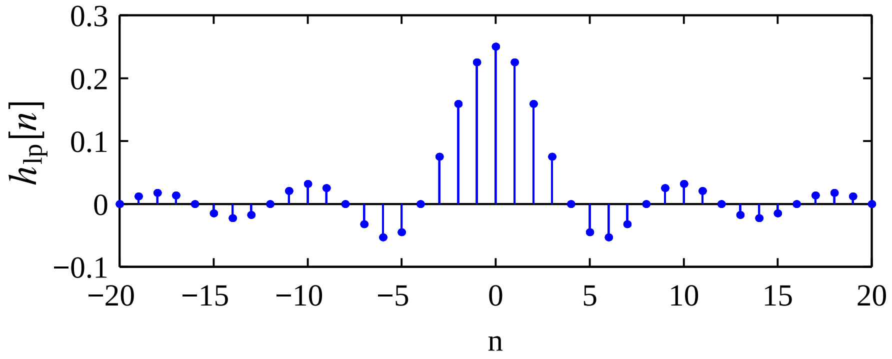
Its magnitude and phase are



This response, as for all discrete-time signals, is periodic with period 2π. Its impulse response (for $-\infty < n < \infty$) is

$$\begin{aligned} h_{\text{lp}}[n] &= \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} e^{j\omega n} d\omega = \frac{1}{2\pi} \left[\frac{1}{jn} e^{j\omega n} \right]_{-\omega_c}^{\omega_c} \\ &= \frac{1}{\pi n} \frac{1}{2j} (e^{j\omega_c n} - e^{-j\omega_c n}) = \frac{\sin(\omega_c n)}{\pi n}, \end{aligned}$$

which for $\omega_c = \pi/4$ is



The ideal lowpass filter is noncausal, and its impulse response extends from $-\infty < n < \infty$. The system is therefore not computationally realisable. Also, the phase response of the ideal lowpass filter is specified to be zero — this is a problem in that causal ideal filters have nonzero phase responses.

The **ideal highpass filter** is

$$H_{\mathrm{hp}}(e^{j\omega}) = \begin{cases} 0 & |\omega| \leq \omega_c \\ 1 & \omega_c < |\omega| \leq \pi. \end{cases}$$

Since $H_{\mathrm{hp}}(e^{j\omega}) = 1 - H_{\mathrm{lp}}(e^{j\omega})$, its frequency response is

$$h_{\mathrm{hp}}[n] = \delta[n] - h_{\mathrm{lp}}[n] = \delta[n] - \frac{\sin(\omega_c n)}{\pi n}.$$

1.2 Phase distortion and delay

Consider the ideal delay, with impulse response

$$h_{\mathrm{id}}[n] = \delta[n - n_d]$$

and frequency response

$$H_{\mathrm{id}}(e^{j\omega}) = e^{-j\omega n_d}.$$

The magnitude and phase of this response are

$$|H_{\mathrm{id}}(e^{j\omega})| = 1,$$
$$\sphericalangle H_{\mathrm{id}}(e^{j\omega}) = -\omega n_d, \qquad |\omega| < \pi.$$

The phase distortion of the ideal delay is therefore a linear function of ω. This is considered to be a rather mild (and therefore acceptable) form of phase distortion, since the only effect is to shift the sequence in time. In other words, a filter with linear phase response can be viewed as a cascade of a zero-phase filter, followed by a time shift or delay.

In designing approximations to ideal filters, we are therefore frequently willing to accept linear phase distortion. The ideal lowpass filter with phase distortion would be defined as

$$H_{\mathrm{lp}}(e^{j\omega}) = \begin{cases} e^{-j\omega n_d} & |\omega| \le \omega_c \\ 0 & \omega_c < |\omega| \le \pi, \end{cases}$$

with impulse response

$$h_{\mathrm{lp}}[n] = \frac{\sin(\omega_c(n - n_d))}{\pi(n - n_d)}.$$

A convenient measure of linearity of the phase is the **group delay**, which relates to the effect of the phase on a narrowband signal. Consider the narrowband input $x[n] = s[n]\cos(\omega_0 n)$, where $s[n]$ is the envelope of the signal. Since $X(e^{j\omega})$ is nonzero only around $\omega = \omega_0$, the effect of the phase of the system can be approximated around $\omega = \omega_0$ by

$$\sphericalangle H(e^{j\omega}) \approx -\phi_0 - \omega n_d.$$

Thus the response of the system to $x[n] = s[n]\cos(\omega_0 n)$ is approximately $y[n] = s[n - n_d]\cos(\omega_0 n - \phi_0 - \omega_0 n_d)$. The time delay of the envelope $s[n]$ of the narrowband signal $x[n]$ with Fourier transform centred at ω_0 is therefore given by the negative of the slope of the phase at ω_0. The group delay of a system is therefore defined as

$$\tau(\omega) = \mathrm{grd}[H(e^{j\omega})] = -\frac{d}{d\omega}\left\{\arg[H(e^{j\omega})]\right\}.$$

The deviation of the group delay away from a constant indicates the degree of nonlinearity of the phase. Note that the phase here must be considered as a continuous function of ω.

2 System response for LCCD systems

Ideal filters cannot be implemented with finite computation. Therefore we need approximations to ideal filters. Systems described by LCCD equations

$$\sum_{k=0}^{N} a_k y[n-k] = \sum_{k=0}^{M} b_k x[n-k]$$

are useful for providing one class of approximation.

The properties of this class of system are best developed in the z-transform domain. The z-transform of the equation is

$$\sum_{k=0}^{N} a_k z^{-k} Y(z) = \sum_{k=0}^{M} b_k z^{-k} X(z),$$

or equivalently

$$\left(\sum_{k=0}^{N} a_k z^{-k}\right) Y(z) = \left(\sum_{k=0}^{M} b_k z^{-k}\right) X(z).$$

The system function for a system that satisfies a difference equation of the required form is therefore

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{k=0}^{M} b_k z^{-k}}{\sum_{k=0}^{N} a_k z^{-k}} = \left(\frac{b_0}{a_0}\right) \frac{\prod_{k=1}^{M}(1 - c_k z^{-1})}{\prod_{k=1}^{N}(1 - d_k z^{-1})}.$$

Each factor $(1 - c_k z^{-1})$ in the numerator contributes a zero at $z = c_k$ and a pole at $z = 0$. Each factor $(1 - d_k z^{-1})$ contributes a zero at $z = 0$ and a pole at $z = d_k$.

The difference equation and the algebraic expression for the system function are equivalent, as demonstrated by the next example.

Example: second-order system
Given the system function

$$H(z) = \frac{(1 + z^{-1})^2}{(1 - \frac{1}{2}z^{-1})(1 + \frac{3}{4}z^{-1})},$$

we can find the corresponding difference equation by noting that

$$H(z) = \frac{1 + 2z^{-1} + z^{-2}}{1 + \frac{1}{4}z^{-1} - \frac{3}{8}z^{-2}} = \frac{Y(z)}{X(z)}.$$

Therefore

$$(1 + \frac{1}{4}z^{-1} - \frac{3}{8}z^{-2})Y(z) = (1 + 2z^{-1} + z^{-2})X(z),$$

and the difference equation is

$$y[n] + \frac{1}{4}y[n-1] - \frac{3}{8}y[n-2] = x[n] + 2x[n-1] + x[n-2].$$

2.1 Stability and causality

A difference equation does not uniquely specify the impulse response of a LTI system. For a given system function (expressed as a ratio of polynomials), each possible choice of ROC will lead to a different impulse response. However, they will all correspond to the same difference equation.

If a system is causal, it follows that the impulse response is a right-sided sequence, and the region of convergence of $H(z)$ must be outside of the outermost pole.

Alternatively, if we require that the system be stable, then we must have

$$\sum_{n=-\infty}^{\infty} |h[n]| < \infty.$$

For $|z| = 1$ this is identical to the condition

$$\sum_{n=-\infty}^{\infty} |h[n]z^{-n}| < \infty,$$

so the condition for stability is equivalent to the condition that the ROC of $H(z)$ include the unit circle.

Example: determining the ROC

The frequency response of the LTI system with difference equation

$$y[n] - \frac{5}{2}y[n-1] + y[n-2] = x[n]$$

is

$$H(z) = \frac{1}{1 - \frac{5}{2}z^{-1} + z^{-2}} = \frac{1}{(1 - \frac{1}{2}z^{-1})(1 - 2z^{-1})}.$$

There are three choices for the ROC:

- **Causal:** ROC outside of outermost pole $|z| > 2$ (but then not stable).
- **Stable:** ROC such that $\frac{1}{2} < |z| < 2$ (but then not causal).
- If $|z| < \frac{1}{2}$ then the system is neither causal nor stable.

For a causal and stable system the ROC must be outside the outermost pole and include the unit circle. This is only possible if all the poles are inside the unit circle.

2.2 Inverse systems

The system $H_i(z)$ is the inverse system to $H(z)$ if

$$G(z) = H(z)H_i(z) = 1,$$

which implies that

$$H(z) = \frac{1}{H_i(z)}.$$

The time-domain equivalent is

$$g[n] = h[n] * h_i[n] = \delta[n].$$

The question of which ROC to associate with $H_i(z)$ is answered by the convolution theorem — for the previous equation to hold the regions of convergence of $H(z)$ and $H_i(z)$ must overlap.

Example: inverse system for first-order system
Let $H(z)$ be

$$H(z) = \frac{1 - 0.5z^{-1}}{1 - 0.9z^{-1}}$$

with ROC $|z| > 0.9$. Then $H_i(z)$ is

$$H_i(z) = \frac{1 - 0.9z^{-1}}{1 - 0.5z^{-1}}.$$

Since there is only one pole, there are only two possible ROCs. The choice of ROC for $H_i(z)$ that overlaps with $|z| > 0.9$ is $|z| > 0.5$. Therefore, the impulse response of the inverse system is

$$h_i[n] = (0.5)^n u[n] - 0.9(0.5)^{n-1} u[n-1].$$

In this case the inverse is both causal and stable.

A LTI system is stable and causal with a stable and causal inverse if and only if both the poles and zeros of $H(z)$ are inside the unit circle — such systems are called **minimum phase** systems.

The frequency response of the inverse system, if it exists, is

$$H(e^{j\omega}) = \frac{1}{H_i(e^{j\omega})}.$$

Not all systems have an inverse. For example, there is no way to recover the frequency components above the cutoff frequency that were set to zero by the action of the lowpass filter.

2.3 Impulse response for rational system functions

If a system has a rational transfer function, with at least one pole that is not cancelled by a zero, then there will always be a term corresponding to an infinite length sequence in the impulse response. Such systems are called **infinite impulse response (IIR)** systems.

On the other hand, if a system has no poles except at $z = 0$ (that is, $N = 0$ in the canonical LCCDE expression), then

$$H(z) = \sum_{k=0}^{M} b_k z^{-k}.$$

In this case the system is determined to within a constant multiplier by its zeros, so the impulse response has a finite length:

$$h[n] = \sum_{k=0}^{M} b_k \delta[n-k] = \begin{cases} b_n & 0 \le n \le M \\ 0 & \text{otherwise} \end{cases}$$

In this case the impulse response is finite in length, and the system is called a **finite impulse response (FIR)** system.

Example: a first-order IIR system
Given a causal system satisfying the difference equation

$$y[n] - ay[n-1] = x[n],$$

the system function is

$$H(z) = \frac{1}{1 - az^{-1}} = \frac{z}{z-a}, \qquad |z| > |a|.$$

The condition for stability is $|a| < 1$. The inverse z-transform is

$$h[n] = a^n u[n].$$

Example: a simple FIR system
Consider the truncated impulse response

$$h[n] = \begin{cases} a^n & 0 \leq n \leq M \\ 0 & \text{otherwise.} \end{cases}$$

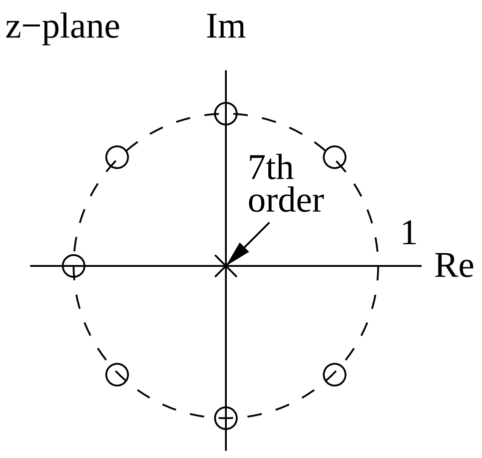
The system function is

$$H(z) = \sum_{n=0}^{M} a^n z^{-n} = \frac{1 - a^{M+1} z^{-M-1}}{1 - az^{-1}}.$$

The zeros of the numerator are at

$$z_k = ae^{j2\pi k/(M+1)}, \qquad k = 0, 1, \ldots, M.$$

With a assumed real and positive, the pole at $z = a$ is cancelled by a zero. The pole-zero plot for the case of $M = 7$ is therefore given by



The difference equation satisfied by the input and output of the LTI system is the convolution

$$y[n] = \sum_{k=0}^{M} a^k x[n-k].$$

The input and output also satisfy the difference equation

$$y[n] - ay[n-1] = x[n] - a^{M+1}x[n-M-1].$$

3 Frequency response for rational system functions

If a stable LTI system has a rational system function, then its frequency response has the form

$$H(e^{j\omega}) = \frac{\sum_{k=0}^{M} b_k e^{-j\omega k}}{\sum_{k=0}^{N} a_k e^{-j\omega k}}.$$

We want to know the magnitude and phase associated with the frequency response. To this end, it is useful to express $H(e^{j\omega})$ in terms of the poles and zeros of $H(z)$:

$$H(e^{j\omega}) = \left(\frac{b_0}{a_0}\right) \frac{\prod_{k=1}^{M}(1 - c_k e^{-j\omega})}{\prod_{k=1}^{N}(1 - d_k e^{-j\omega})}.$$

It follows that

$$|H(e^{j\omega})| = \left|\frac{b_0}{a_0}\right| \frac{\prod_{k=1}^{M}|1 - c_k e^{-j\omega}|}{\prod_{k=1}^{N}|1 - d_k e^{-j\omega}|}.$$

Therefore $|H(e^{j\omega})|$ is the product of the magnitudes of all the **zero factors** of $H(z)$ evaluated on the unit circle, *divided by* the product of the magnitudes of all the **pole factors** evaluated on the unit circle.

The **gain** in dB of $H(e^{j\omega})$, also called the **log magnitude**, is given by

$$\text{Gain in dB} = 20 \log_{10} |H(e^{j\omega})|,$$

which for a rational system function takes the form

$$20\log_{10}|H(e^{j\omega})| = 20\log_{10}\left|\frac{b_0}{a_0}\right| + \sum_{k=1}^{M} 20\log_{10}|1-c_k e^{-j\omega}| - \sum_{k=1}^{N} 20\log_{10}|1-d_k e^{-j\omega}|.$$

Also

$$\text{Attenuation in dB} = -\text{Gain in dB}.$$

Thus a 60dB attenuation at frequency ω corresponds to $|H(e^{j\omega})| = 0.001$. Also,

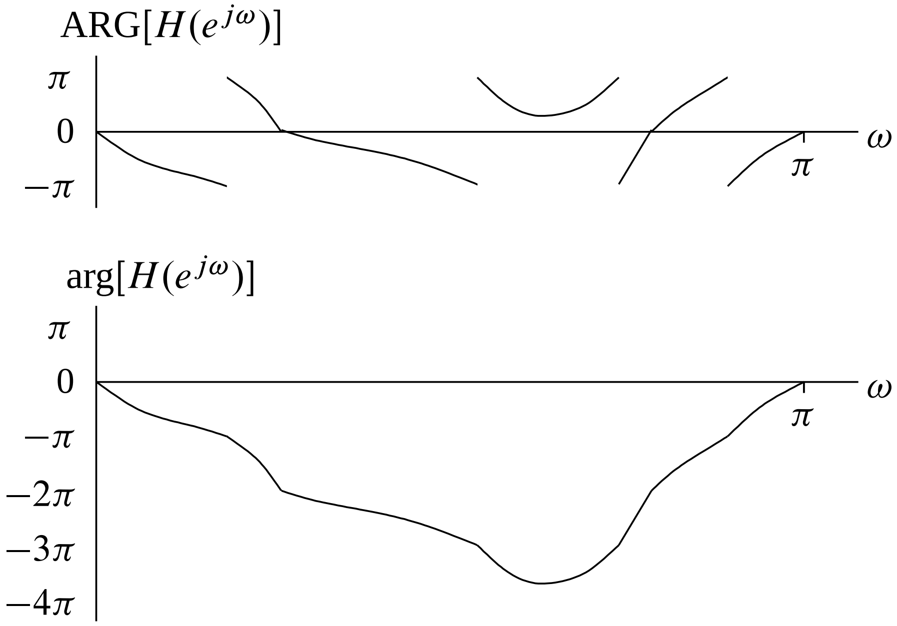
$$20\log_{10}|Y(e^{j\omega})| = 20\log_{10}|H(e^{j\omega})| + 20\log_{10}|X(e^{j\omega})|.$$

The phase response for a rational system function is

$$\sphericalangle H(e^{j\omega}) = \sphericalangle\left[\frac{b_0}{a_0}\right] + \sum_{k=1}^{M} \sphericalangle[1-c_k e^{-j\omega}] - \sum_{k=1}^{N} \sphericalangle[1-d_k e^{-j\omega}].$$

The zero factors contribute with a plus sign and the pole factors with a minus.

In the above equation, the phase of each term is ambiguous, since any integer multiple of 2π can be added at each value of ω without changing the value of the complex number. When calculating the phase with a computer, the angle returned will generally be the **principal** value $\text{ARG}[H(e^{j\omega})]$, which lies in the range $-\pi$ to π. This phase will generally be a discontinuous function, containing jumps of 2π radians whenever the phase wraps. Appropriate multiples of 2π can be added or subtracted, if required, to yield the continuous phase function $\arg[H(e^{j\omega})]$.



3.1 Frequency response of a single pole or zero

Consider a single zero factor of the form

$$(1 - re^{j\theta}e^{-j\omega})$$

in the frequency response. The magnitude squared of this factor is

$$\begin{aligned} |1 - re^{j\theta}e^{-j\omega}|^2 &= (1 - re^{j\theta}e^{-j\omega})(1 - re^{-j\theta}e^{j\omega}) \\ &= 1 + r^2 - 2r\cos(\omega - \theta), \end{aligned}$$

so the log magnitude in dB is

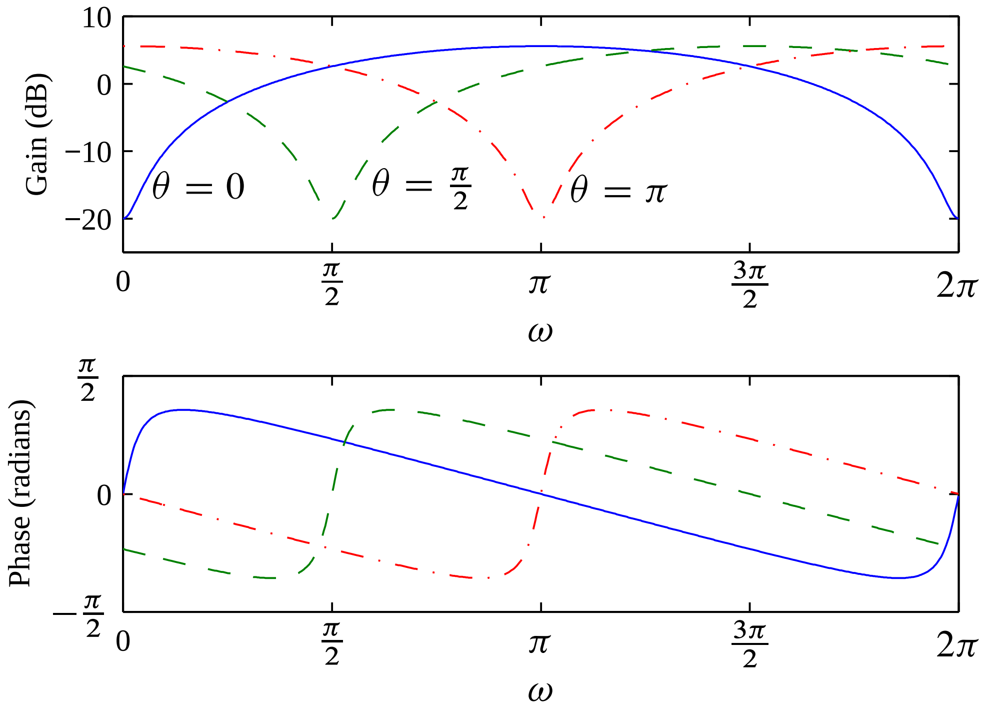
$$20\log_{10}|1 - re^{j\theta}e^{-j\omega}| = 10\log_{10}[1 + r^2 - 2r\cos(\omega - \theta)].$$

The principle value of the phase for the factor is

$$\text{ARG}[1 - re^{j\theta}e^{-j\omega}] = \arctan\left[\frac{r\sin(\omega - \theta)}{1 - r\cos(\omega - \theta)}\right].$$

These functions are periodic in ω with period 2π.

The following plot shows the frequency response for $r = 0.9$ and three different values of θ:



Note that

- The gain dips at $\omega = \theta$. As θ changes, the frequency at which the dip occurs changes.
- The gain is maximised for $\omega - \theta = \pi$, and for $r = 0.9$ the magnitude of the resulting gain is

$$10 \log_{10}(1 + r^2 + 2r) = 20 \log_{10}(1 + r) = 5.57\text{dB}.$$

- The gain is minimised for $\omega = \theta$, and for $r = 0.9$ the resulting gain is

$$10 \log_{10}(1 + r^2 - 2r) = 20 \log_{10} |1 - r| = -20\text{dB}.$$

- The phase is zero at $\omega = \theta$.

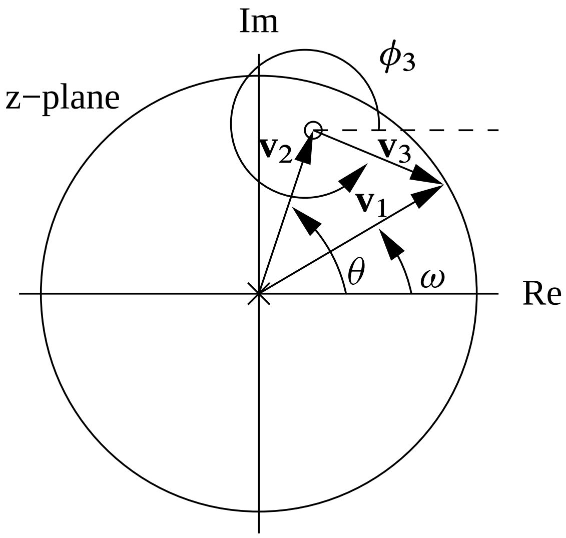
Note that if the factor $(1 - re^{j\theta}e^{j\omega})$ occurs in the denominator, thereby representing a pole factor, then the entire analysis holds with the exception that the sign of the log magnitude and the phase changes.

The frequency response can be sketched from the pole-zero plot using a simple geometric construction. Note firstly that the frequency response corresponds to an evaluation of $H(z)$ on the unit circle. Secondly, the complex value of each pole and zero can be represented by a vector in the z-plane from the pole or zero to a point on the unit circle.

Take for example the case of a single zero factor

$$H(z) = (1 - re^{j\theta} z^{-1}) = \frac{z - re^{j\theta}}{z} \qquad r < 1,$$

which corresponds to a pole at $z = 0$ and a zero at $z = re^{j\theta}$.



If the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3 = \mathbf{v}_1 - \mathbf{v}_2$ represent respectively the complex numbers $e^{j\omega}$, $re^{j\theta}$, and $e^{j\omega} - re^{j\theta}$, then

$$|H(e^{j\omega})| = |1 - re^{j\theta} e^{-j\omega}| = \left| \frac{e^{j\omega} - re^{j\theta}}{e^{j\omega}} \right| = \frac{|\mathbf{v}_3|}{|\mathbf{v}_1|} = |\mathbf{v}_3|.$$

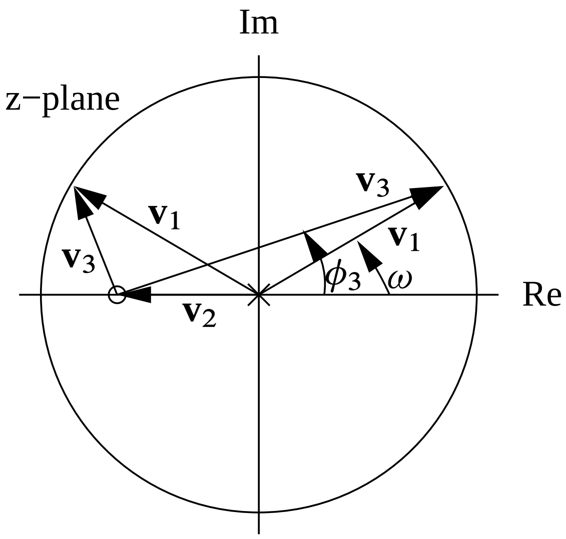
The phase is

$$\begin{aligned} \sphericalangle H(e^{j\omega}) &= \sphericalangle(1 - re^{j\theta} e^{-j\omega}) = \sphericalangle(e^{j\omega} - re^{j\theta}) - \sphericalangle(e^{j\omega}) \\ &= \sphericalangle(\mathbf{v}_3) - \sphericalangle(\mathbf{v}_1) = \phi_3 - \phi_1 = \phi_3 - \omega. \end{aligned}$$

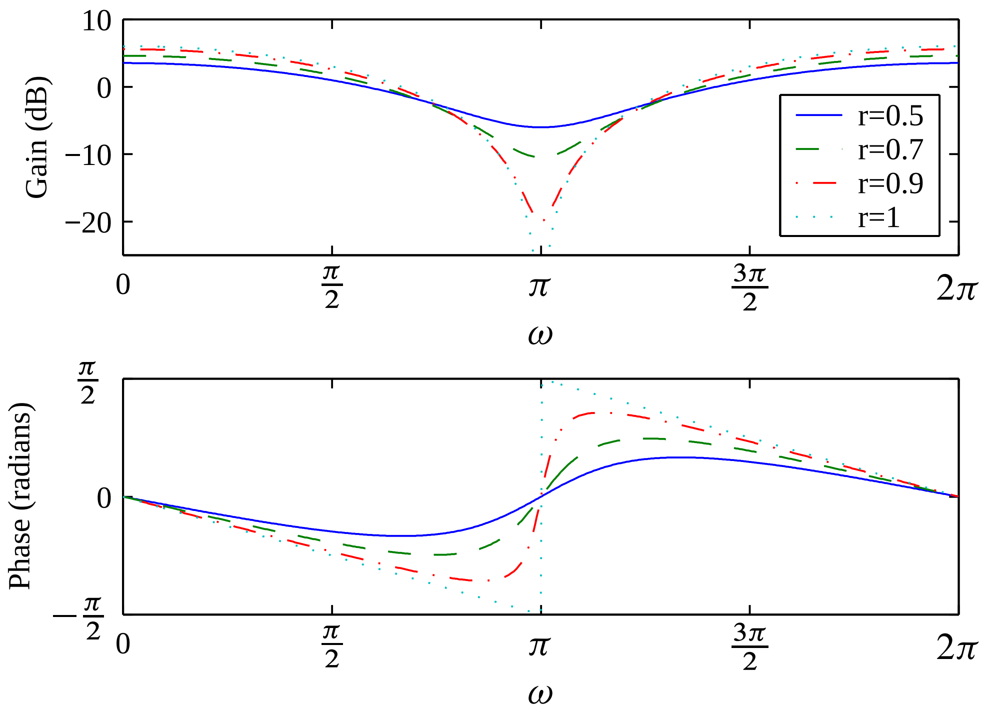
A vector such as $\mathbf{v}_3$ from a zero to the unit circle is referred to as a **zero**

vector, and a vector from a pole to the unit circle is called a **pole vector**.

Consider now the pole zero system depicted below:

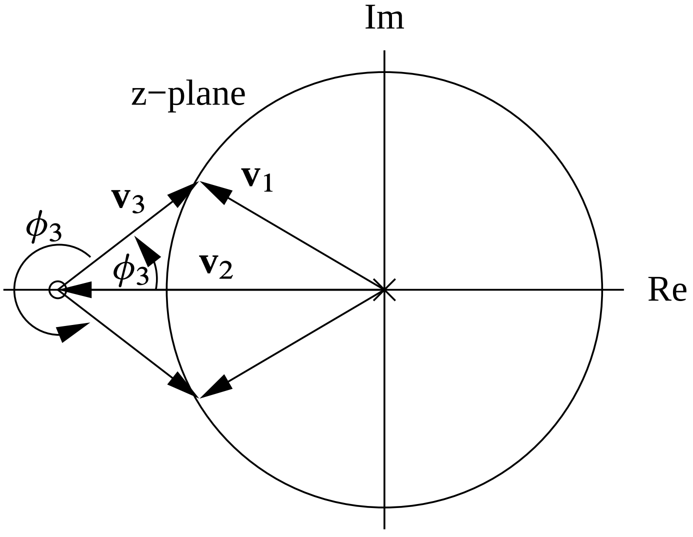


The frequency response for the single zero at different values of r and $\theta = \pi$ is

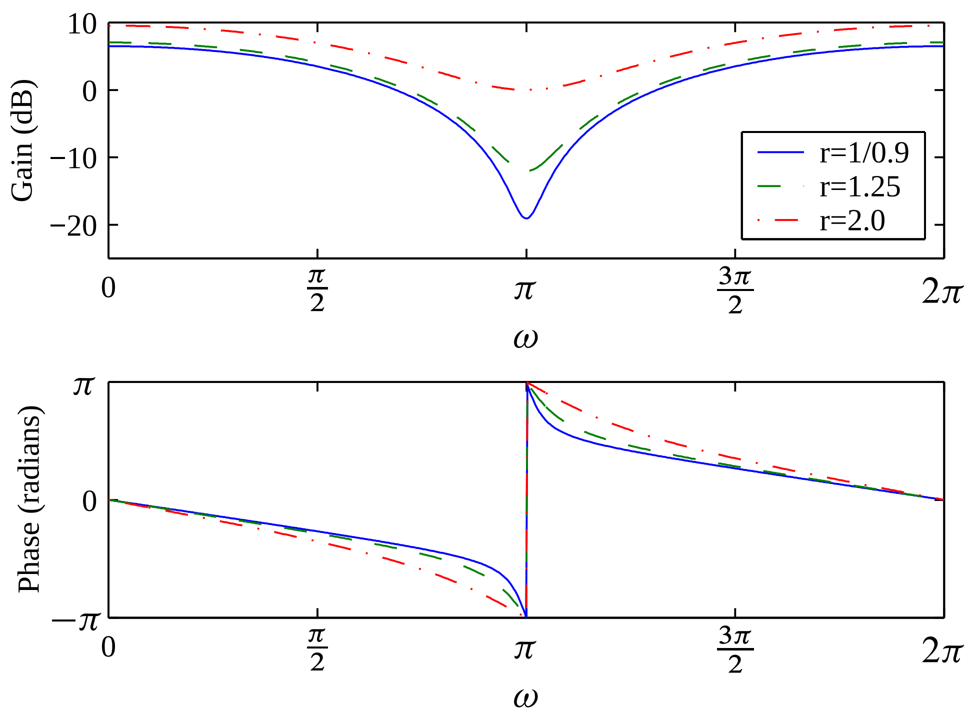


Note that the log magnitude dips more sharply as r approaches 1 (and at $\omega = \pi$ tends to $-\infty$ as r tends towards 1). The phase function has positive slope around $\omega = \theta$. This slope increases as r approaches 1, and becomes infinite for $r = 1$. In this case the phase function is discontinuous, with a jump of π radians at $\omega = \theta$.

If r increases still further, to lie outside of the unit circle,



then the frequency response becomes



3.2 Frequency response with multiple poles and zeros

In general, the z-transform of a LTI system can be factorised as

$$H(z) = \left(\frac{b_0}{a_0}\right) \frac{\prod_{k=1}^{M}(1 - c_k z^{-1})}{\prod_{k=1}^{N}(1 - d_k z^{-1})} = \left(\frac{b_0}{a_0}\right) z^{N-M} \frac{\prod_{k=1}^{M}(z - c_k)}{\prod_{k=1}^{N}(z - d_k)}.$$

Depending on whether N is greater than or less than M, the factor z^{N-M} represents either $N - M$ zeros at the origin, or $M - N$ poles at the origin. In either case, the z-transform can be written in the form

$$H(z) = K \frac{\prod_{i=0}^{M_0}(z - z_k)}{\prod_{i=0}^{N_0}(z - p_k)}$$

where $z_1, \ldots, z_{M_0}$ are the zeros, and $p_1, \ldots, p_{N_0}$ the poles of $H(z)$. This representation could also be obtained by merely factorising $H(z)$ in terms of z rather than z^{-1}.

The frequency response of this system is

$$H(e^{j\omega}) = K \frac{\prod_{i=0}^{M_0}(e^{j\omega} - z_k)}{\prod_{i=0}^{N_0}(e^{j\omega} - p_k)}.$$

The magnitude is therefore

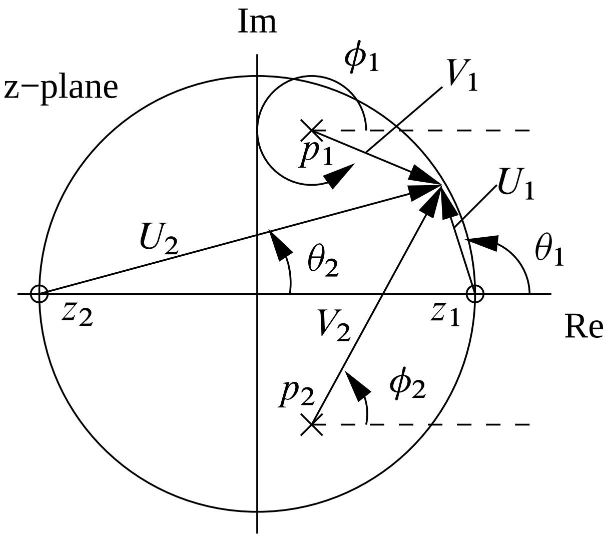
$$|H(e^{j\omega})| = |K| \frac{\prod_{i=0}^{M_0} |e^{j\omega} - z_k|}{\prod_{i=0}^{N_0} |e^{j\omega} - p_k|},$$

and the phase is

$$\sphericalangle H(e^{j\omega}) = \sum_{i=0}^{M_0} \sphericalangle(e^{j\omega} - z_k) - \sum_{i=0}^{N_0} \sphericalangle(e^{j\omega} - p_k).$$

In the z-plane, $(e^{j\omega} - z_k)$ is simply the vector from the zero z_k to the point on the unit circle. The term $|e^{j\omega} - z_k|$ is the length of this vector, and $\sphericalangle(e^{j\omega} - z_k)$ is the angle that it makes with the positive real axis. Similarly, the term $(e^{j\omega} - p_k)$ corresponds to the vector from the pole p_k to the point on the unit circle.

It follows then that the the magnitude response is the product of the lengths of the zero vectors, divided by the product of the lengths of the pole vectors. The phase response is the sum of the angles of the zero vectors, minus the sum of the angles of the pole vectors. Thus, for the two pole, two zero system



the frequency response is

$$|H(e^{j\omega})| = |K| \frac{U_1 U_2}{V_1 V_2}$$

$$\sphericalangle H(e^{j\omega}) = \theta_1 + \theta_2 - (\phi_1 + \phi_2).$$

Here K is a constant factor which cannot be determined from the pole zero diagram alone, but only serves to scale the magnitude.

4 Realisation structures for digital filters

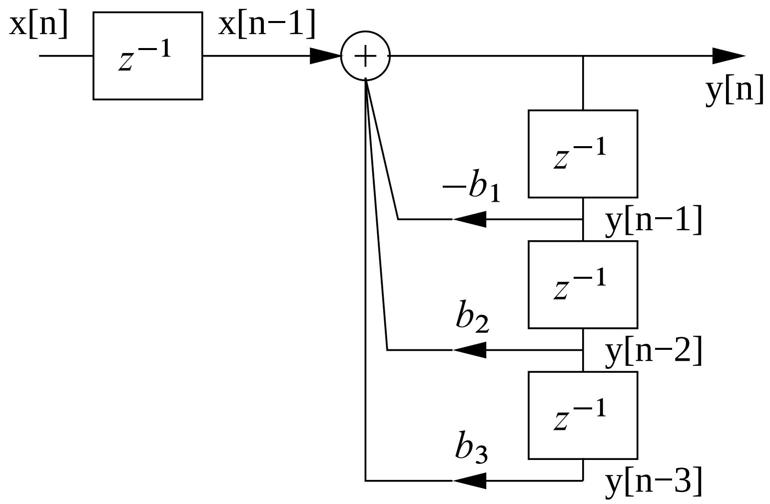
The difference equation, the impulse response, and the system function are all equivalent characterisations of the input-output relation for a LTI discrete-time system. For implementation purposes, systems described by LCCDEs can be implemented by structures consisting of an interconnection of the basic operations of addition, multiplication by a constant, and delay. The desired form for the interconnections depends on the technology to be used.

Discrete-time filters are often represented in the form of block or signal flow diagrams, which are convenient for representing the difference equations or transfer functions.

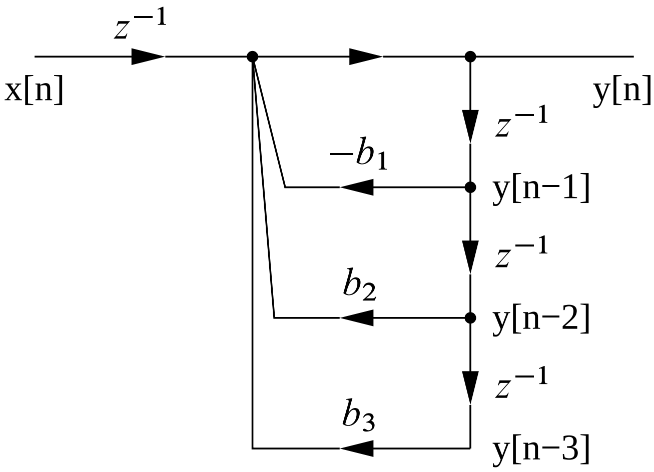
For example, the system with the difference equation

$$y[n] = x[n-1] - b_1 y[n-1] + b_2 y[n-2] + b_3 y[n-3]$$

can be represented in a block diagram form as



The symbol z^{-1} represents a delay of one unit of time, and the arrows represent multipliers (with the constant multiplication factors next to them). The equivalent signal flow diagram is



The relationship between the diagrams and the difference equation is clear.

Many alternative filter structures can be developed, and they differ mainly with respect to their numerical stability and the effects of quantisation on their performance. A discussion of these effects can be found in many DSP texts.