Signals and Systems – Part II: Systems

Prof Guy-Bart Stan

27th November 2019

Contents

Acknowledgement

I would like to thank Prof Anil Bharath for giving me the structure of his course on which these lecture notes are based.

Online Demos

The interested reader is referred to the following Matlab toolboxes and online demos that can help in illustrating the concepts introduced in the "Signals and Systems" course.

Matlab Toolboxes. The "*Signal Processing and Communications*" and the "*Control System Design and Analysis*" *Toolboxes of Matlab* offer many functions and Graphical User Interfaces (GUI) that can be highly valuable to illustrate the introduced concepts.

For example:

- The analysis of Linear Time-Invariant Systems is facilitated through the use of the "Linear System Analyzer" GUI.
- For the design of linear filters the following GUIs are worth having a look at:
  - "Filter Design & Analysis Tool" GUI
  - "Filter Builder" GUI
- Finally, the design of LTI controllers is facilitated through the use of the "Control System Designer" GUI.

Online Demos. Beyond these Matlab functions and GUIs, there are various interactive demos available online that can also be used to acquire a better intuition and understanding of the introduced concepts.

- Dot Product:
  - http://www.falstad.com/dotproduct/
- Fourier series:
  - http://www.falstad.com/fourier/
  - http://demonstrations.wolfram.com/FourierSeriesOfSimpleFunctions/
  - http://demonstrations.wolfram.com/RecoveringTheFourierCoefficients/
  - http://demonstrations.wolfram.com/FourierTransformPairs/
  - http://demonstrations.wolfram.com/ComparingFourierSeriesAndFourierTransform/
  - http://demonstrations.wolfram.com/FromContinuousToDiscreteTimeFourierTransformBySamplingMet
  - http://demonstrations.wolfram.com/SamplingTheorem/
- Convolution:
  - http://demonstrations.wolfram.com/DiscreteTimeConvolution/
  - http://demonstrations.wolfram.com/ConvolutionWithARectangularPulse/
- Digital Filters:
  - http://www.falstad.com/dfilter/
- General Interest Demos:
  - http://www.falstad.com/mathphysics.html

Bridge with the first part of the course

In the first part of the course, we saw that one of the key steps to be able to reconstruct a continuous-time signal from its discrete-time samples (i.e. to perform a digital to analogue conversion) consisted in passing the discretised signal through an "ideal low-pass filter". In what follows we will deepen our understanding of filters and, more generally, of systems, which we will look at from an input-output point of view, i.e. we will consider systems and filters as "black boxes" that transform input signals into output signals, without caring, initially, about the content of this black box. This input-output point of view is at the core of analogue and digital filter design and signal processing.

1 Systems

In this introduction, we will give a broad overview of the essential features of most common "systems".

1.1 General definition

- A system may be thought of as a **Black Box** (B.B.) with one or more input terminals and one or more output terminals.
- This Black Box could be:
  - a mechanical system
  - an electrical system
  - a chemical system
  - a biological system
  - ...

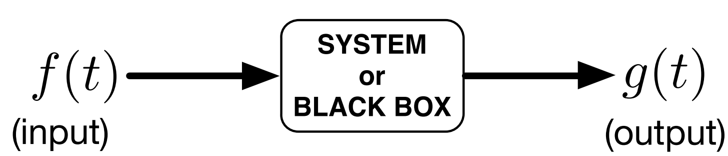
  or could be an imaging, or an audio device/process.

How do we describe systems or B.B.?

Usually we want to be able to define a mapping between the inputs and the outputs of the B.B., i.e. we don't want to know what is inside the B.B. or how it is built, and, instead, we care about the relationship between its inputs and its outputs. There are different types of system classes and subclasses, but we here focus on four main classes that are used widely in systems engineering and signal processing.

2 Types of systems

System or Black Box



Most Common Systems Types

A	**Linear**	**Nonlinear**
B	**Time Invariant**	**Time Variant**
C	**Causal**	**Acausal**
D	**Open Loop**	**Closed Loop**

Remark 1. *For time-invariant B.B., time doesn't affect the content of the B.B.*

Remark 2. *For closed loop B.B., the output signal(s) "feed(s) back" to the input signal(s).*

Remark 3. *Linear time-invariant systems have very important properties that make their analysis and design much easier than for nonlinear or time-variant systems.*

Remark 4. *Nonlinear and stochastic systems will be introduced as part of the* "Modelling in Biology" *course.*

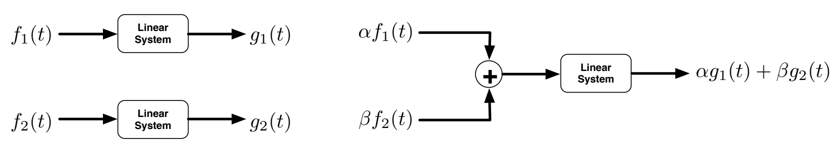
2.1 A. Linear vs Nonlinear Systems

For a linear system, the following **must** hold:

$$\text{IF } \; f_1(t) \text{ (input)} \overset{System}{\Rightarrow} g_1(t) \text{ (output)}$$

$$\text{AND } \; f_2(t) \text{ (input)} \overset{System}{\Rightarrow} g_2(t) \text{ (output)}$$

$$\text{THEN } \; \boxed{\alpha f_1(t) + \beta f_2(t) \underset{\text{Linear}}{\overset{System}{\Rightarrow}} \alpha g_1(t) + \beta g_2(t), \quad \forall f_1(t) \neq 0 \text{ and } f_2(t) \neq 0}$$



In other words, for a linear system, if the input is a linear combination of inputs, the output will be the same linear combination of the outputs corresponding to these inputs.

Remark 5. *This property must hold* ***for all*** *non-zero $f_1(t)$ and $f_2(t)$.*

Remark 6. *A system for which the linear system property does not hold is by definition* nonlinear.

Remark 7. *For any practical device, linearity is typically going to hold only for a subset of all possible signals. So, even supposedly linear amplifiers (for example) are typically only so for a certain range of input signals.*

Remark 8. *Linearity is often desirable. For example, a lot of audio systems are designed to be linear. More specifically, the level of quality of Hi-Fi audio systems is measured in terms of their linearity, i.e. their ability to truthfully reproduce sounds as they were recorded. For Hi-Fi systems, nonlinearites are called "distortions".*

However, some systems are intrinsically nonlinear (e.g. biological systems) or are designed to behave nonlinearly (e.g. switch or relay systems).

2.2 B. Time Invariant vs Time Variant Systems

For time invariant systems, the mapping between the input and the output does not depend on the time at which the input signal starts. Mathematically time invariance amounts to the following property:

IF $f(t)$ (input) $\overset{System}{\Rightarrow}$ $g(t)$ (output)

THEN $\boxed{f(t-\tau) \underset{\text{Time Invariant}}{\overset{System}{\Rightarrow}} g(t-\tau), \quad \forall \tau \in \mathbb{R}}$

In other words, for a time invariant system, if the input signal is time-shifted by τ then the output signal will be shifted by the same time-shift τ.

Remark 9. *A pure amplification, $g(t) = Af(t)$, where A is a constant is an example of a time invariant system (it is also linear).*

Mathematically, time-invariance is easy to see:

- *Time-shifting the input in the expression of the output gives: $Af(t-\tau)$*
- *Time-shifting the output gives: $g(t-\tau) = Af(t-\tau)$*

It is clear that these two signals are the same. Therefore, this system is time invariant.

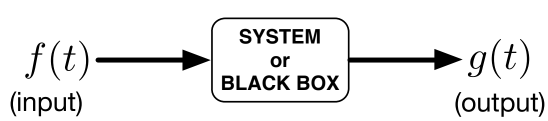
On the contrary, a system defined by $g(t) = A(t)f(t)$ is time variant (and still linear) since the amplification $A(t)$ is here a function of time, i.e. the point in time at which you consider the system makes a difference for the output! Mathematically, this is again easy to see:

- *Time-shifting the input in the expression of the output gives: $A(t)f(t-\tau)$*
- *Time-shifting the output gives: $g(t-\tau) = A(t-\tau)f(t-\tau)$*

It is clear that these two signals are not the same. Therefore, this system is time variant.

2.3 C. Causal vs Acausal Systems

A system is said to be *causal* if the output of the system, $g(t)$, is only dependent on the values of the input to the system, $f(t)$, for times up until the current point in time, t.



For all t, $g(t)$ can only depend on $f(t-\tau)$ for values of $\tau \geq 0$. In other words, for causal systems, τ can never take negative values.

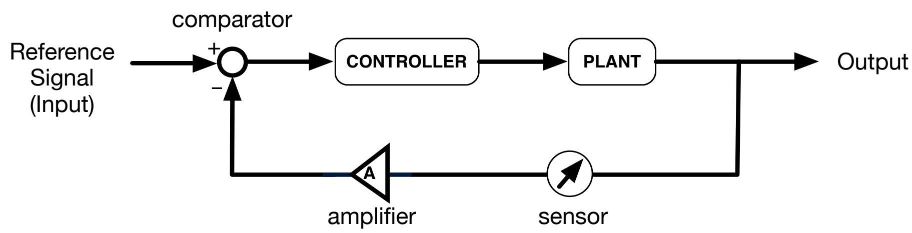
Remark 10. *A causal system cannot depend on an input that is ahead in time of the current input. For example, if the relationship between the input and the output is given by $g(t) = f(t-\tau)$ and we consider $\tau = -1$ and $t = 10$ then $f(t-\tau) = f(11)$ and $g(t) = g(10)$. This system is thus not causal as its output at the current time, $t = 10$, depends on the input at a future point in time, $t = 11$.*

Remark 11. *All systems that exist in real-life are causal.*

Remark 12. *If the system is only dependent on $f(t+\tau)$ for $\tau \geq 0$, the system is* acausal. *A system that is* acausal *requires the ability to look ahead in time and provide the output $g(t)$ as a function of future inputs, $f(t+\tau)$, $\tau \geq 0$. Acausal systems do not exist in real-life but can be used for offline processing (based on input-output data that have already been collected and stored).*

2.4 D. Open Loop vs Closed Loop Systems

In a closed loop system (some proportion of) the output signal is fed back to the input signal.

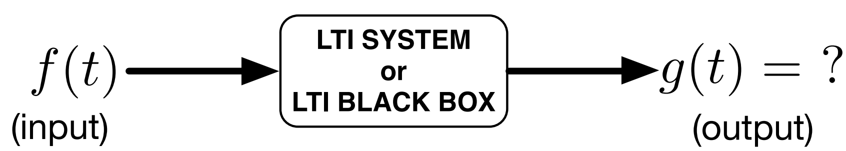


Examples of closed-loop systems include:

- Thermostats for automatic temperature regulation around a desired reference temperature, irrespective of temperature perturbations (e.g. caused by open window(s) or door(s)).
- (Adaptive) Cruise Control for automatic speed regulation in cars (or automatic distance regulation with respect to other vehicles on the road), irrespective of the slope of the road or number of passengers.
- Escalator speed regulation for maintaining constant speed, irrespective of the number of people on the escalator.
- Microphone & amplifier in feedback and the "Larsen" effect: the feedback path is physically implemented through the propagation of sound waves through air from the loudspeaker to the microphone.

  The amplification level is critical to trigger/avoid the "Larsen" effect.

3 LTI Systems



Is there a way to describe the output $g(t)$ of an LTI system/B.B. in terms of its input $f(t)$ and some "core characteristic"?

Definition of an LTI System
An LTI system is a system that is both linear and time invariant.

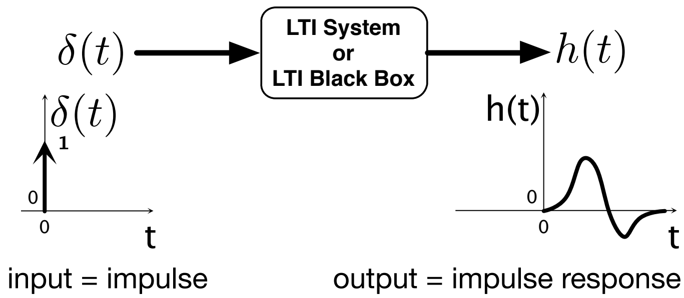
Why are LTI systems important?
LTI systems are important because any LTI system can be *completely characterised* by a "signal" known as its *impulse response*.

Definition of the impulse response of a system
The impulse response of a system is the output of the system obtained in response to a δ-"function" ("impulse") at its input.

Remark 13. *Sometimes people talk also about LSI systems. An LSI system is a system that is both linear and* space *invariant. LSI systems are an extension of LTI systems to 2D imaging systems.*

Remark 14. *As we will see next, a system that is both linear and time invariant (an LTI system) inevitably performs* convolution *to produce its output from a given input.*

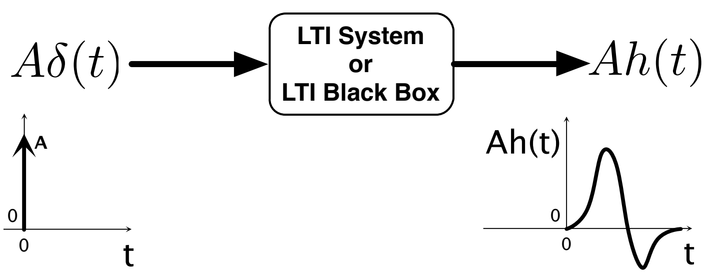
3.1 LTI Systems and the Impulse Response



The ouptut of an LTI system to an impulse signal, i.e. a δ-“function”, is the *impulse response* of that LTI system.

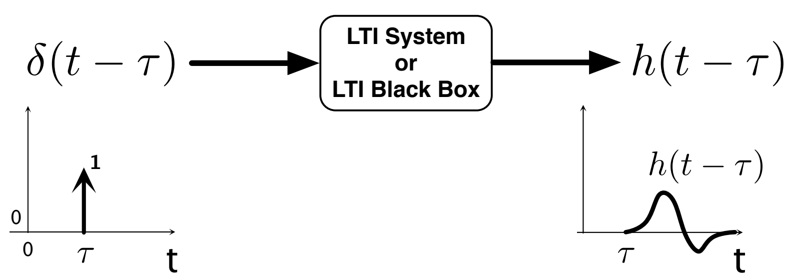
Linearity:

If the system is linear and the input is scaled by some constant, then the output will be scaled by the same constant.



Time invariance:

Similarly, if the system is time invariant and we delay the input by τ, then the output will also be delayed by the same amount τ.



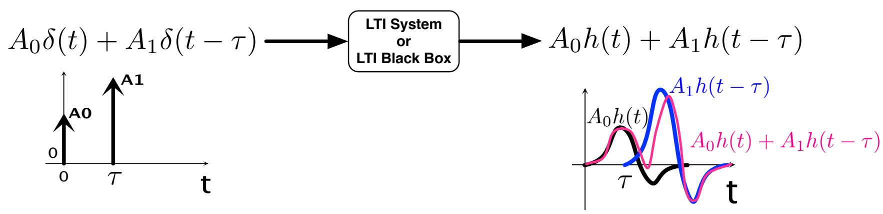
Linearity + Time Invariance (LTI):

As a consequence of the combination of linearity and time invariance, if the input of an LTI system is:

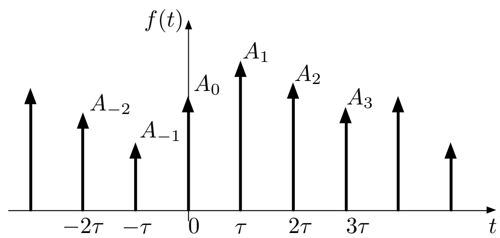
$$f(t) = A_0\delta(t) + A_1\delta(t-\tau)$$

then the output of this LTI system will be:

$$g(t) = A_0h(t) + A_1h(t-\tau)$$



What happens if there are more δ-"functions" at the input?



With more delayed δ-"functions" at the input, the output is straightforward to deduce from what we just saw.

If the input of an LTI system has the following definition:

$$f(t) = \sum_{n=-\infty}^{\infty} A_n \delta(t - n\tau)$$

Then the output of that LTI system will be (by linearity and time invariance):

$$g(t) = \sum_{n=-\infty}^{+\infty} A_n h(t - n\tau)$$

Each scaled and time-delayed δ-"function" gives rise to a correspondingly scaled and time-delayed impulse response. The output is then simply the summation of all these scaled and time-delayed impulse responses.

Compare this last expression for the output with the expression for convolution:

$$f_1(t) * f_2(t) = \int_{-\infty}^{+\infty} f_1(\tau) f_2(t-\tau) d\tau$$

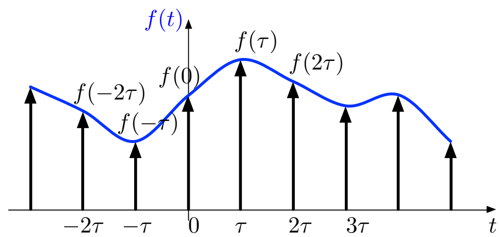
In particular, for $f_1(t) = f(t)$ and $f_2(t) = h(t)$, we get:

$$f(t) * h(t) = \int_{-\infty}^{+\infty} f(\tau) h(t-\tau) d\tau$$

As we can see, τ appears in both the expression of $g(t)$ and in that of $f(t) * h(t)$. However, the expression of the output of the LTI system, $g(t)$, to a "train" of δ-"functions" involves a summation whereas the expression of the convolution involves an integration.

Let's have a look at the train of δ-"functions" again. If we assume that the magnitudes of the δ-functions, A_n, correspond to the time-sampled values of the continuous-time function $f(t)$, with a sampling period τ, we then have: $A_0 = f(0)$, $A_1 = f(\tau)$, $A_2 = f(2\tau)$, ..., $A_n = f(n\tau)$, ... [1]

[1] This is quite clear from what you saw when we introduced sampling: If $f(t)$ is a continuous-time function, then sampling $f(t)$ using a sampling period of τ amounts to multiply $f(t)$ by a "train" of δ-"functions" where the δ-"functions" are delayed by multiples of τ: $f_{\text{sampled}} = f(t)\sum_{-\infty}^{+\infty}\delta(t-n\tau) = \sum_{-\infty}^{+\infty} f(t)\delta(t-n\tau) = \sum_{-\infty}^{+\infty} f(n\tau)\delta(t-n\tau)$.



Therefore, $f(t)$ and $g(t)$ can be expressed as:

$$f(t) = \sum_{-\infty}^{+\infty} f(n\tau)\delta(t - n\tau)$$

and

$$g(t) = \sum_{-\infty}^{+\infty} f(n\tau)h(t - n\tau).$$

As $\tau \to 0$, the summations become integrations,[2]
and we thus obtain:

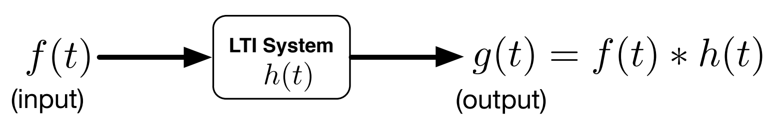
$$\begin{aligned} f(t) &= \int_{-\infty}^{+\infty} f(\tau)\delta(t-\tau)d\tau \quad (n\tau \text{ gets replaced by } \tau) \\ &= f(t) * \delta(t) \quad (\text{by the sifting property of the } \delta\text{-“function”}) \end{aligned}$$

and, similarly:

$$g(t) = \int_{-\infty}^{+\infty} f(\tau)h(t-\tau)d\tau = f(t) * h(t)$$

Output of an LTI system in terms of its impulse response
For any LTI system, the output $g(t)$ can always be expressed as the *convolution* of the input $f(t)$ with the impulse response of this LTI system $h(t)$:

$$\boxed{g(t) = f(t) * h(t) = h(t) * f(t)} \tag{1}$$



Consequences and meaning of (1):

- Any system which is linear and time invariant is **completely characterised** by its **impulse response, $h(t)$, i.e. for *any* input signal $f(t)$, the output signal of an LTI system is given by $g(t) = f(t) * h(t) = h(t) * f(t)$ where $h(t)$ is the impulse response of the considered LTI system**
- We can predict the response of an LTI system to any input given the knowledge of its impulse response.

Remark 15. *An example of the use of impulse responses of LTI systems in industry is given by image analysis of fMRI scans to identify task-specific functional regions within the brain or by 3D-audio virtual reality systems:*

[2] As $\tau \to 0$, the spacing between the samples decreases until the samples are very tightly packed. Summing over these very tightly packed samples thus begins to look like computing the area under the (continuous-time) curve, which corresponds to an integration.

- *Search for task-specific functional regions within the brain using fMRI scans:*
  - *Using modelling or experimental approaches, the brain haemodynamic impulse response (or haemodynamic transfer function) of a patient performing a specific task (e.g. tapping his index finger on a table) can be predicted/measured.*
  - *Knowing this haemodynamic impulse response, the problem amounts to searching time-lapse movies of fMRI scans for the set of pixels whose intensities change in time according to the* a priori *predicted brain heamodynamic (impulse) response.*
- *3D-audio virtual reality systems require the availability of two impulse responses:*
  - *The impulse response of the acoustical space that one wants to immerse the listener into, e.g. a cathedral, a lecture hall, or an opera house. This impulse response, once convolved with a sound input signal (e.g. a piece of music or a speech), will produce a sound output that corresponds to the signal you would have obtained by directly recording the input signal played in the chosen acoustical space (e.g. cathedral, lecture hall, opera house, etc.).*
  - *The binaural impulse responses of the listener (or of an "average listener") to a sound emitted from a particular direction around the head of that listener. These binaural impulse responses, once convolved with a sound input signal, will "spatialise" the sound output, i.e. the resulting sound output, when listened to using headphones, will be perceived as if coming from a sound source emitting the sound input signal from the direction used in the measurement of the binaural impulse responses.*

  Convolution of any sound signal (input) with these impulse responses (e.g. the input sound signal is convolved first with the acoustical space impulse response, and the resulting signal is then convolved with the binaural impulse responses) allows to both "colour" and spatialise the sound signal so as to give the impression to the listner that he/she is acoustically immersed in a chosen 3D environment. This is currently being used for creating 3D audio virtual reality systems by the gaming industry and in high-end civilian and military simulators.

3.2 Transfer Function: the Fourier Transform of the Impulse Response

We can also look at the expression (1) in the frequency domain by taking the Fourier transform of both sides:

$$\begin{aligned} FT\{g(t)\} &= FT\{f(t) * h(t)\} \\ &= FT\{f(t)\}FT\{h(t)\} \end{aligned}$$

which implies:

Output of an LTI system in terms of its transfer function

$$\boxed{G(j\omega) = F(j\omega)H(j\omega) = H(j\omega)F(j\omega)} \quad (2)$$

where $H(j\omega) = FT\{h(t)\}$ is the *transfer function* of the LTI system.

Consequences and meaning of (2):

- The Fourier transform of the impulse response of an LTI system is the **transfer function** of that LTI system.
- Any system which is linear and time invariant is **completely characterised** by its **transfer function**.

- We can predict the response of an LTI system to any input given the knowledge of its transfer function.

- If we wish to determine the amplification factor given to different frequencies of the input signal of an LTI system, we only need to measure its impulse response, $h(t)$, compute its associated transfer function using $H(j\omega) = FT\{h(t)\}$ and then $|H(j\omega)|$ gives the amplification (or attenuation) factor at the chosen angular frequency $\omega = 2\pi\nu$ where ν is the frequency in Hz and ω is the angular frequency in rad/s. (Note that, as we will emphasise later, a full characterisation of the transfer function requires more than just the frequency-dependent amplification factor, also known as magnitude response, $|H(j\omega)|$. For a full characterisation, you would also need to consider the frequency-dependent *phase* response, i.e. $\angle H(j\omega)$).

3.3 Intermediate Summary

The **impulse response** or the **transfer function** of a Linear Time Invariant (LTI) system *each* **completely characterise** the input-output properties of that system.

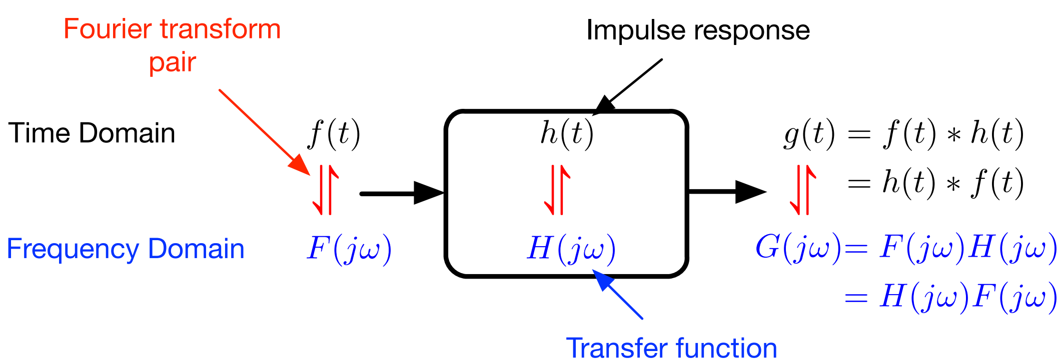
Given the input to an LTI system, the output can be determined:

- **In the time domain:** as the **convolution** of the impulse response and the input.

- **In the frequency domain:** as the **multiplication** of the transfer function and the Fourier transform of the input.

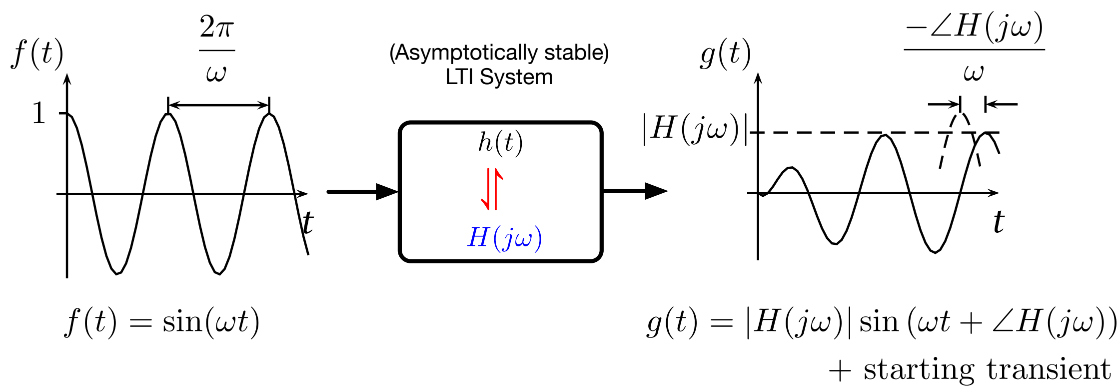
They are **related** as follows: The transfer function is the Fourier transform of the impulse response.

Remark 16. *From* (2)*, we can see that the transfer function of an LTI system is, by definition,* $H(j\omega) = \frac{G(j\omega)}{F(j\omega)}$*, i.e.* the Fourier transform of the output signal divided by the Fourier transform of the input signal.



3.4 Response of LTI systems to sinusoidal inputs

If a pure sinusoid is input into an (asymptotically stable) LTI system, then the output will also settle down, eventually, to a pure sinusoid. This *steady-state* output will have the same frequency as the input but will have a different magnitude and phase. The dependence of the magnitude and phase on the frequency of the input is called the *frequency response* of the system.



The frequency response $H(j\omega)$ is a *complex-valued* function of the angular frequency $\omega = 2\pi\nu$ (in rad/s) where ν is the frequency (in Hz). At each angular frequency ω, the complex number $H(j\omega)$ can be represented either in terms of its real and imaginary parts, or in terms of its magnitude (or "gain"), $|H(j\omega)|$, and phase, $\angle H(j\omega)$.[3]

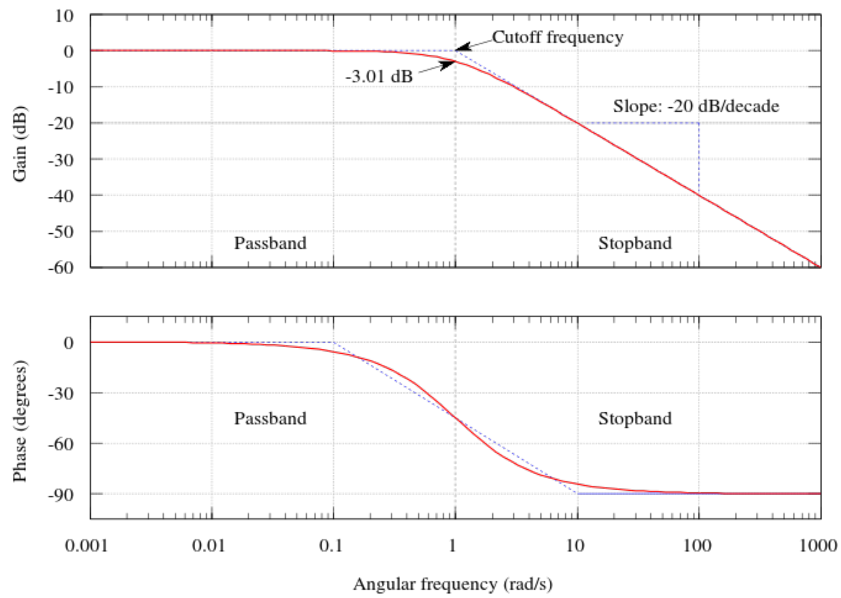
Remark 17. *The frequency response of an LTI system provides,* for each value of the angular frequency ω*, a direct understanding of the change in magnitude and phase imposed on* any *signal passing through that LTI system. This is a consequence of the definition of the transfer function of the system, i.e.* $G(j\omega) = F(j\omega)H(j\omega)$*, which implies that* $|G(j\omega)| = |F(j\omega)||H(j\omega)|$ *and* $\angle G(j\omega) = \angle F(j\omega) + \angle H(j\omega)$*. Please also note that any signal can be decomposed into a sum of sinusoidal signals through a Fourier series decomposition. Therefore, understanding the response of an LTI system to a sinusoidal input is essential to understanding its response to any input signal.*

3.5 Bode magnitude and phase diagrams

The frequency response can be captured through *Bode* diagrams, which consist in two separate graphs:

- one of $\mathbf{20}\log_{\mathbf{10}}|\mathbf{H}(\mathbf{j}\omega)|$ (in decibels, dB, i.e. $20\log_{10}$ axis) **vs** ω (in rad/s, $\log_{10}$ axis), i.e. the *Bode* magnitude *diagram*.[4]
- one of $\angle\mathbf{H}(\mathbf{j}\omega)$ (in degrees or radians, linear axis) **vs** ω (in rad/s, $\log_{10}$ axis), i.e. the *Bode* phase *diagram*.

[3] $|H(j\omega)|e^{j\angle Hj\omega}$ is the polar representation of the complex number $H(j\omega)$.
[4] In some books, the Bode *magnitude* diagram is called the Bode *gain* diagram.



Remark 18. *As we will see in what follows, Bode diagrams are relatively straightforward to sketch to a high degree of accuracy, are compact and give an indication of the frequency ranges within which different levels of performance in terms of the frequency response can be achieved.*

Remark 19. *By changing the frequency of the sinusoidal signal at the input of an LTI system and recording the corresponding sinusoidal output, one can build Bode diagrams point by point (frequency by frequency) by simply comparing the output and input magnitudes, and the output and input phases.*

However, as we will see, simple rules allow to approximately sketch Bode diagrams quite easily and reveal important aspects key to the analysis and design of LTI filters.

In what follows we will try to get a better understanding of Bode diagrams and of their use for the design of LTI filters in the frequency domain.

4 LTI Filter Design in the Frequency Domain

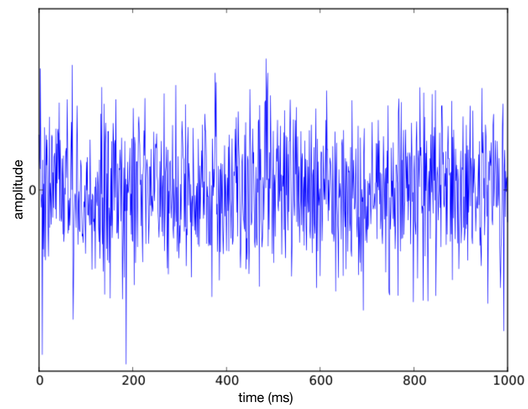
Linear filtering represents a large application class for LTI systems. In what follows, we will emphasise the duality between time-domain and frequency-domain representations of signals (Bode diagrams) and their use for the design and realisation of basic LTI filters. We will, in particular, focus on:

- Broad-spectrum signals: Pseudo-random noise (e.g. Maximum Length Sequences) vs Sweep signals (e.g. Sine Sweeps) vs impulse.
- Low-Pass, High-Pass, Band-Pass, and Band-Stop Filters and (examples of) their corresponding transfer functions.
- Butterworth Filter: the general form of the Butterworth filter can be used as one way of specifying a transfer function and therefore an impulse response.

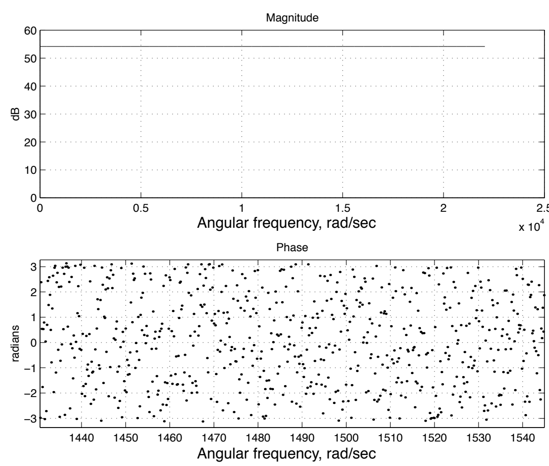
4.1 Broad-Spectrum Signals: Maximum Length Sequence and Sine Sweep signals

Exercise 1. *What do you think the Bode magnitude and phase diagrams for an impulse signal (δ-"function") would look like?* Hint: *Consider the Fourier transform of the δ-"function". What does that tell you about the Bode magnitude and phase diagrams of a δ-"function"?*

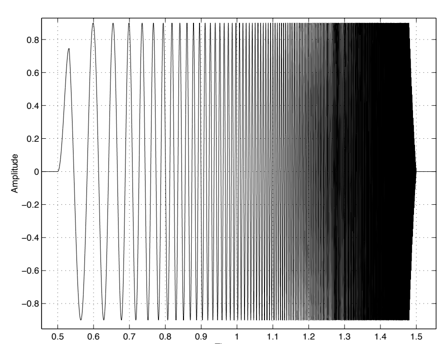
A **Maximum Length Sequence** (MLS) signal of any desired length can be easily generated to approximate a white noise signal:



The Bode magnitude and phase diagrams of an MLS signal look like this:



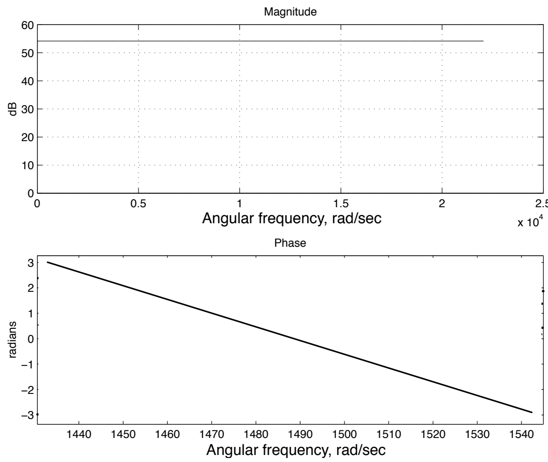
A **Sine Sweep (or Chirp)** signal is another example of a broad-spectrum signal:



Exercise 2. *What do you think the Bode magnitude and phase diagrams for a linear sine sweep would look like?* Hint: The slope of the phase diagram gives you information about the "*group delay*", i.e. the lag (or time delay) it takes for a frequency to "appear" in the time signal (e.g. in the sine sweep here). The more negative the slope on the Bode phase diagram at a given frequency

$\hat{\omega}$, the later frequencies "close to" that frequency $\hat{\omega}$, i.e. $\hat{\omega} + \epsilon$ with $0 < \epsilon \ll 1$, will appear in the time signal (e.g. in the sine sweep).[5]

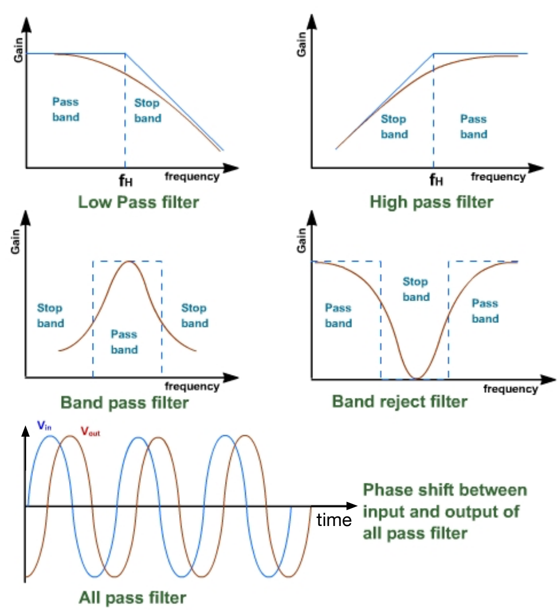
The Bode magnitude and phase diagrams of a linear sine sweep signal look like this:



As we can see from the broad spectrum signals above, the Bode magnitude diagram is not sufficient to uniquely specify a signal. White noise (or Maximum Length Sequence), sine sweep, and impulse (δ-"function") signals have identical Bode magnitude diagrams. However, their Bode phase diagrams are very different. This emphasises that *both* Bode magnitude and phase diagrams need to be specified if one wants, on their basis, to be able to uniquely specify an associated signal (either in the time or in the frequency domain).

4.2 Bode Magnitude Diagrams for Basic Filters

There are five major types of LTI filters. Hereafter, we provide a characterisation of the first 4 in terms of their Bode magnitude diagrams.



[5]The reason for considering frequencies "close to" the considered frequency $\hat{\omega}$ comes directly from the definition of the group delay: $\tau_g = -\frac{d(\angle H(j\omega))}{d\omega}$, since the derivative of a curve (here, the Bode phase curve) at given a point gives only local information about that curve (i.e. the information given by the derivative of a curve at a point $\hat{\omega}$ is only valid for points "close to" the point $\hat{\omega}$).

Remark 20. *Bode magnitude* AND *phase diagrams are useful to build an understanding of the action of filters in the frequency domain.*

The actual implementation of a filter typically requires to have access to the analytical form of the transfer function associated with the desired Bode diagrams. More specifically, the type of implementation (analogue or digital) of a filter dictates the type of transfer function (continuous-time or discrete-time) that needs to be considered. Once the type of transfer function is chosen, the frequency domain design of an analogue or digital filter is typically done by trying to find a transfer function that approximates at best the desired Bode magnitude *diagram. The Bode* phase *diagram is then imposed by this chosen transfer function. This is why many text books on analogue or digital filter design typically only show the Bode magnitude diagrams and then immediately show a transfer function that can approximate this Bode magnitude response. Please do remember, however, that the expression of a transfer function imposes both Bode magnitude and Bode phase diagrams, and that you always need both Bode magnitude and phase information to fully specify a signal or a filter in the frequency domain.*

4.3 Examples of Transfer Functions for Basic Filters

Hereafter, we provide some examples of first order transfer functions for low-pass and high-pass filters. Furthermore, cascade or parallel combination of high-pass and low-pass filters can be used to obtain basic band-pass and band-stop filters:

- First Order Low-Pass Filter Transfer Function: $K\frac{1}{1+\tau j\omega}$ where $\omega_c = \frac{1}{\tau}$ is the cutoff angular frequency of the Low-Pass filter.
- First Order High-Pass Filter Transfer Function: $K\frac{\tau j\omega}{1+\tau j\omega}$ where $\omega_c = \frac{1}{\tau}$ is the cutoff angular frequency of the High-Pass filter.
- Band-Pass = cascade of High-Pass and Low-Pass filters where the cutoff angular frequency of the High-Pass is smaller than the cutoff angular frequency of the Low-Pass.[6]
- Band-Stop = parallel combination of Low-Pass and High-Pass filters where the cutoff angular frequency of the Low-Pass is smaller than the cutoff angular frequency of the High-Pass.

Exercise 3. *Plot the Bode diagrams of the above three filters (low-pass, high-pass, and band-pass) for values of the parameters (K and cutoff angular frequencies) that you chose yourself.*

4.4 Sketching Bode diagrams

Basic idea: Consider a transfer function written as a ratio of factorised polynomials, e.g.

$$H(j\omega) = \frac{a_1(j\omega)a_2(j\omega)}{b_1(j\omega)b_2(j\omega)}$$

Clearly:

$$\log_{10}|H(j\omega)| = \log_{10}|a_1(j\omega)| + \log_{10}|a_2(j\omega)| - \log_{10}|b_1(j\omega)| - \log_{10}|b_2(j\omega)|,$$

so we can compute the Bode magnitude curve by simply adding and subtracting magnitudes corresponding to terms in the numerator and denominator. Similarly:[7]

[6] *Cascade* here refers to the output of one filter being connected to the input of another. If a filter, $h(t)$, is composed of two filters, $h_1(t)$ and $h_2(t)$, in cascade, then $h(t) = h_1(t) * h_2(t)$ or equivalently, $H(j\omega) = H_1(j\omega)H_2(j\omega)$. For a Band-Pass filter composed of the cascade of a High-Pass filter and a Low-Pass filter, we would thus have: $K_1\frac{\tau_1 j\omega}{1+\tau_1 j\omega} \cdot K_2\frac{1}{1+\tau_2 j\omega}$ with $\frac{1}{\tau_1} < \frac{1}{\tau_2}$.

[7] The phase of the product of two complex numbers is equal to the sum of the phases of these complex numbers. This is easily seen from the polar representation of complex numbers: $\angle\left(ae^{jb}be^{jd}\right) = \angle\left(abe^{j(b+d)}\right) = b + d$.

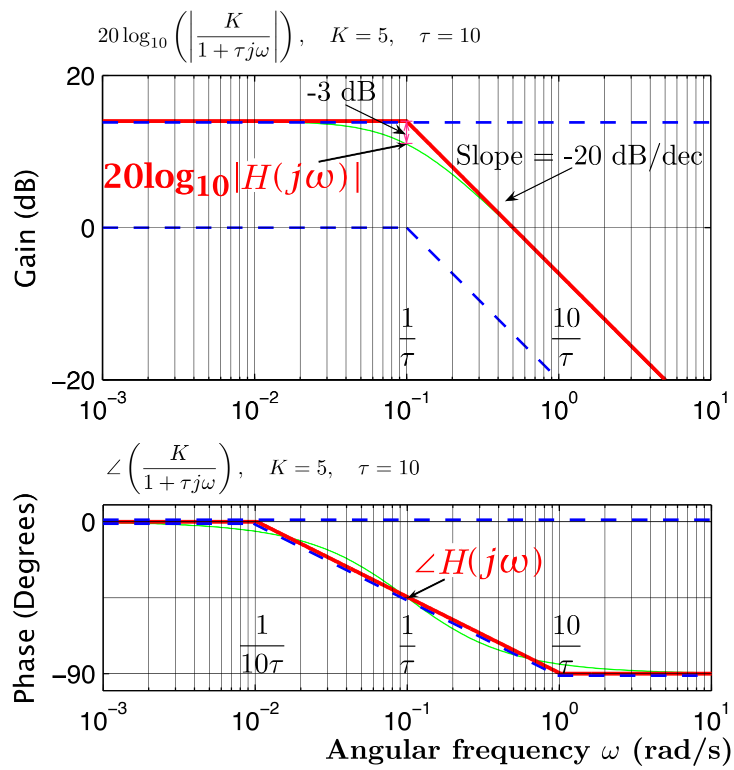
Furthermore, the phase of the inverse of a complex number is equal to minus the phase of that complex number. Again, this is easily seen from the polar representation of complex numbers: $\angle\left(\frac{1}{ae^{jb}}\right) = \angle\left(\frac{1}{a}e^{-jb}\right) = -b$.

$$\angle H(j\omega) = \angle a_1(j\omega) + \angle a_2(j\omega) - \angle b_1(j\omega) - \angle b_2(j\omega)$$

and so the Bode phase curve can be determined in an analogous fashion.

4.4.1 First order low-pass filter

The Bode magnitude and phase diagrams for a low-pass filter with $K = 5$ and $\tau = 10$ are given hereafter:



In what follows, we explain how such a diagram can be sketched.

Consider the transfer function of a first order low-pass filter:

$$H(j\omega) = K\frac{1}{1+\tau j\omega}, \quad K > 0, \quad \tau > 0$$

Bode Magnitude

$$\begin{aligned} 20\log_{10}|H(j\omega)| &= 20\log_{10}\left|K\frac{1}{1+\tau j\omega}\right| \\ &= 20\log_{10}|K|\frac{1}{|1+\tau j\omega|} \\ &= 20\log_{10}|K| + \cancel{20\log_{10}(1)} - 20\log_{10}|1+\tau j\omega| \\ &= 20\log_{10}|K| - 20\log_{10}\sqrt{1+(\tau\omega)^2} \end{aligned}$$

The Bode magnitude diagram for the transfer function $K\frac{1}{1+\tau j\omega}$ is thus the sum of the Bode magnitude diagrams $20\log_{10}|K|$ and $-20\log_{10}\sqrt{1+(\tau\omega)^2}$.

Asymptotic behaviours of $-20\log_{10}\sqrt{1+(\tau\omega)^2}$

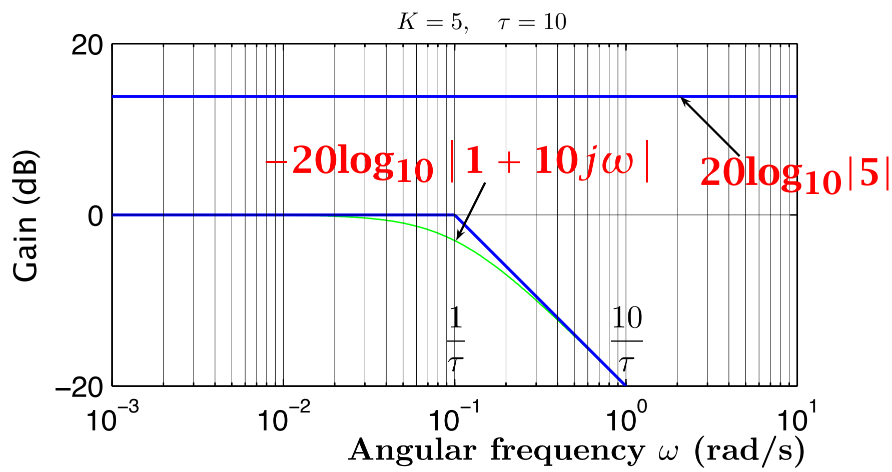
- $\omega \ll \frac{1}{\tau} \Leftrightarrow (\tau\omega)^2 \ll 1 \Leftrightarrow -20\log_{10}\sqrt{1+(\tau\omega)^2} \approx -20\log_{10}(1) = 0$ (horizontal asymptote)
- $\omega \gg \frac{1}{\tau} \Leftrightarrow (\tau\omega)^2 \gg 1 \Leftrightarrow -20\log_{10}\sqrt{1+(\tau\omega)^2} \approx -20\log_{10}(\tau\omega)$ ("diagonal" asymptote)

Remark 21. *The slope of the "diagonal" asymptote* $-20\log_{10}(\tau\omega)$ *is* $-20dB/decade$*. This is easily seen by considering 2 angular frequencies that are a* decade *away from each other, i.e. one angular frequency is* 10 *times larger than the other one. For example, consider the two angular frequencies,* ω_0 *and* $10\omega_0$*, where* ω_0 *is arbitrarily chosen. The value at* $10\omega_0$ *on the asymptote is* $-20\log_{10}(\tau 10\omega_0) = -20\log_{10}(10) - 20\log_{10}(\tau\omega_0) = -20 - 20\log_{10}(\tau\omega_0)$*, i.e.* $20dB$ *less than the value at* ω_0*.*

In particular, on the asymptote, at $\omega = \omega_0 = \frac{1}{\tau}$*, we have that* $-20\log_{10}(\tau\omega_0) = 0dB$*, whereas at* $\omega = 10\omega_0 = \frac{10}{\tau}$*, we have* $-20\log_{10}(\tau\frac{10}{\tau}) = -20dB$*, i.e.* $20dB$ *less than the value at* $\omega_0 = \frac{1}{\tau}$*.*

Remark 22. *The two asymptotes intersect at* $\omega = \frac{1}{\tau}$*. This can be easily seen by considering the value of* ω *at which the two asymptotes are equal, i.e. the value of* ω *at which* $-20\log_{10}(\tau\omega) = 0$*, which is* $w = \frac{1}{\tau}$*.*

Remark 23. *At* $\omega = \frac{1}{\tau}$*, we have* $20\log_{10}|H(j\omega)| = 20\log_{10}(K) - 20\log_{10}\left(\sqrt{2}\right) = 20\log_{10}(K) - 3.01\ dB$*, i.e.* $3.01dB$ *less than the value at which the asymptotes intersect.*



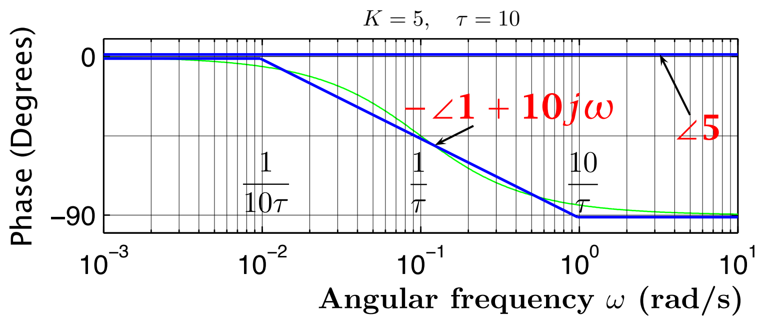
Bode Phase

$$\begin{aligned}\angle H(j\omega) &= \angle K - \angle(1+\tau j\omega)\\ &= 0 - \arctan\left(\frac{\tau\omega}{1}\right)\\ &= -\arctan(\tau\omega)\end{aligned}$$

Asymptotic behaviours of $-\arctan(\tau\omega)$

- $\omega \ll \frac{1}{\tau} \Leftrightarrow \tau\omega \ll 1 \Leftrightarrow -\arctan(\tau\omega) \approx -\arctan(0) = 0$
- $\omega \gg \frac{1}{\tau} \Leftrightarrow \tau\omega \gg 1 \Leftrightarrow -\arctan(\tau\omega) \approx -\arctan(+\infty) = -90°$

Remark 24. *At* $\omega = \frac{1}{\tau}$*, we have* $-\arctan(\tau\omega) = -\arctan(1) = -45°$*.*



Remark 25 (Comments on Bode sketching)**.** *An alternative technique for sketching Bode diagrams would be to ignore the asymptotes, and just calculate and plot the true Bode magnitude and phase over a grid of frequencies. This is* not *recommended for a number of reasons. Firstly, a lot more points are required to get the same accuracy. Secondly, the structure of the diagram is then lost. A practising filter design engineer or control engineer will often prefer a good sketch, with the asymptotes shown, to an accurate computer generated diagram – since this gives a better idea of how things can be changed to improve the behaviour of the filter or of the controlled system.*

In practice, sketching Bode diagrams is about producing a drawing showing the straight line asymptotes and approximations and a rough approximation to the true gain and phase by rounding the corners appropriately.

4.4.2 First order high-pass filter

The Bode magnitude and phase diagrams for a high-pass filter can be obtained similarly.

Exercise 4. *Sketch the Bode diagrams for a first order high-pass filter, i.e. for a filter with the following transfer function:* $H(j\omega) = K\frac{\tau j\omega}{1+\tau j\omega}$.

4.4.3 Band-Pass and Band-Stop filters

The Bode diagrams for a cascade of first order LTI filters (such as those considered in the creation of a Band-Pass filter) can be obtained by considering the Bode diagrams composition rules outlined above (see Section 4.4).

Exercise 5. *Sketch the Bode diagrams for a band-pass filter.*

4.5 The Butterworth filter

In what follow we will show how the rather theoretical work that we have been doing on LTI systems is related to the construction of analogue electronic circuits realising these systems.

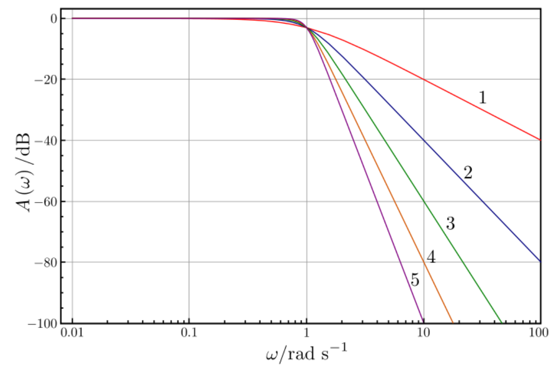
Butterworth filters can be used to specify transfer functions corresponding to low-pass, high-pass, band-pass, or band-stop filters. As a result of this frequency-domain filter design process, once the expression of the desired transfer function is known (e.g. analytic expression of the Butterworth transfer function), it can be realised using analogue electronic components or implemented as purely digital filters. Matlab can be used to easily perform the design of a Butterworth filter (using the Matlab command `butter`[8]).

Consider the following *Butterworth filter* Bode magnitude specification:

$$|H(j\omega)|^2 = \frac{1}{1 + (\omega/\omega_c)^{2N}} \tag{3}$$

where N determines the order of the filter and ω_c determines the cutoff angular frequency.

[8]The Matlab command `help butter` or, even more informatively, the Matlab documentation on the function `butter`, provide interesting information on the general transfer function form of Butterworth filters and how Matlab can be used to easily design low-pass, high-pass, band-pass or band-stop Butterworth filters.



The above figure is a plot of the Bode magnitude for Butterworth low-pass filters of orders $N = \{1, 2, 3, 4, 5\}$ with cutoff angular frequency $\omega_c = 1\ rad/s$. Note that the slope is $-20N$ dB/decade where N is the filter order.

Remark 26. *As you can see, the higher the order N of the Butterworth filter, the closer the Bode magnitude response gets to that of an ideal low-pass filter.*

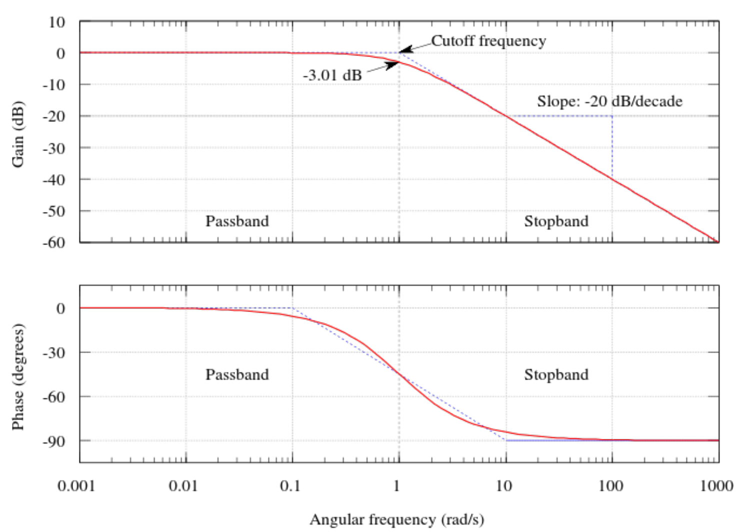
Let us consider the case of a Butterworth filter of order $N = 1$ with a cutoff angular frequency $\omega_c = 1\ rad/s$, so that

$$|H(j\omega)|^2 = \frac{1}{1+\omega^2}$$

Because this is only a specification on the Bode *magnitude* of the filter, one still has to decide on what the Bode *phase* specification will be.

Remark 27. *There are some standard options for specifying the phase, which depend on how the filter will be implemented (e.g. analogue implementation or digital implementation). For the time being, we will just look at a "standard" analogue filter implementation. This imposes the analytical expression of the (continuous-time) transfer function and thereby the Bode phase plot. For an analogue implementation of this Butterworth filter, we will here consider the transfer function: $H(j\omega) = \frac{1}{1+j\omega}$.*

An example of the magnitude and Bode phase plots for an analogue Butterworth filter of order $N = 1$ with cutoff freqency $\omega_c = 1\ rad/s$ is provided hereafter. As you can see the Bode diagrams are exactly those that we considered for first order low-pass filters.



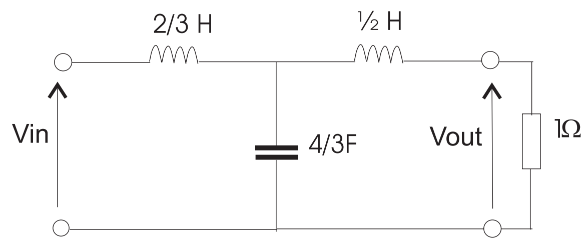
4.6 Passive Filter Implementation of the Butterworth Filter

The simplest implementation of a Butterworth filter is surprisingly familiar to anyone who has done even the tiniest bit of electronics. A Butterworth filter can be constructed by an "LC ladder" network, where capacitors and inductors are strung together.

One example of a passive circuit implementation for a third-order Butterworth filter, with $N = 3$ and $\omega_c = 1\ rad/s$:

$$|H(j\omega)|^2 = \frac{1}{1+\omega^6},$$

might look like this:



The above schematic representation represents an electronic circuit that can be used to implement a 3^{rd} order passive Butterworth filter.

By changing the values of the electronic components in this LC ladder network we can obtain different cut-off frequencies. As we saw previously, the Bode magnitude response of the corresponding Butterworth filter is not quite the same as that of the ideal low-pass filter ($rect_1(\cdot)$) but it approximates it increasingly well as the order of the Butterworth filter N increases.