Physics Experiments

Signal Processing and Noise

Jelmer J.T. Wagenaar

Leiden University

2024

Physics Experiments: Signal Processing and Noise
Jelmer J.T. Wagenaar
jelmerjt@gmail.com

First edition (Dutch): August 2012
Second edition (Dutch): August 2013
Third edition: August 2014
Fourth edition: August 2015
Fifth edition: January 2018
Sixth edition: January 2019
Seventh edition: January 2020
Eighth edition: December 2020
Nineth edition: August 2022
Tenth edition: July 2023
Eleventh edition: August 2024

In collaboration with Tjerk H. Oosterkamp, Bas Hensen, and Sense Jan van der Molen.
Translated to English by Thomas O'Reilly in 2014.
Updated by Jaap Kautz in 2018.
Chapter 0 was written by Michiel de Dood, Jaap Kautz and Paul Logman in 2019 and adapted by Jelmer Wagenaar in 2022.
Bálint Bosman helped with editing this book in 2022.

Contents

Introduction

When you perform an experiment, you get the opportunity to test a scientific hypothesis by directly interacting with nature. To properly set up and perform a measurement is a challenge which you have already been introduced to in student lab work (practicum). An important part of a good experiment, however, deals with the way a signal is detected and subsequently treated. For that, it is not only important to understand the measurement process itself (a physical interaction!), but also to assess the possible sources of noise and interference and to know how to deal with them. The course Physics Experiments deals with these issues and can be seen as a bridge between student lab work and a full-blown research project. Interestingly, you will find that doing this correctly requires a significant mathematical foundation. The main mathematical ingredients are linear differential equations and Fourier transforms. This reader aims to present a mathematical and physical basis to the physics students at Leiden University, who typically have a solid mathematical understanding. However, the philosophy of the full course is broader. Apart from exercise classes, you will also be using your newly acquired knowledge in practice during a set of lab experiments. In this way, we aim for you to be ready for an inspiring and high-level Bachelor or Master project by the time you have finished this course.

As for the history of this document: Tjerk Oosterkamp, the lecturer of the course Signal Processing and Noise until 2012, wanted to have a reader in order to accommodate the need for a solid starting point in this subject. As a student, Jelmer Wagenaar followed this course and he wrote the first version of this reader in 2012. Sense Jan van der Molen has been the lecturer of Signal Processing and Noise since 2013 and it was decided in 2014 to offer the course in English. Hence, Thomas O'Reilly who followed the subject as a student, translated the course material to English. In 2018 Jaap Kautz updated the reader and in 2019, Michel Orrit was the lecturer of the course Physics Experiments 2. This course is similar to the course Signal Processing and Noise, but with some of the material of the first two chapters already taught in Physics Experiments 1. In 2019-2020, Jelmer Wagenaar started lecturing both courses Physics Experiments 1 and 2. In 2022, we added the reader of Physics Experiments 1 written in 2019 by Michiel de Dood, Jaap Kautz and Paul Logman and adapted by Jelmer Wagenaar in 2020 to this reader as a new chapter 0. In 2023, before leaving the university and starting full-time as teacher at a secondary school, Jelmer updated the reader a final time with some fixed mistakes in the definition of noise spectra in chapter 3.

The current reader presents a summary of most of the material and forms the basis for further understanding. The use of illustrations and short definitions allows for the quick referencing of concepts. However, the reader is rather concise and hence, we also refer to several books that can help you to gain insight in this important subject. In past years, we have received feedback from students on ways in which the reader may be improved. Your suggestions for improving the reader further are sincerely welcome. Hopefully this reader will be a strong basis that will provide new insight in an efficient manner.

Jelmer Wagenaar, July 2023

Chapter 0

Fourier transformations

Introduction

In this chapter you will learn to apply the mathematical method of Fourier analysis to physics research. The big conceptual step that you need to make is to describe time-dependent signals in the frequency domain by thinking of these signals as being composed of infinitely many harmonic functions. Instead of thinking about the signal and the response of a physical system as a function of time, you will learn to think about signals and physical quantities as an amplitude and phase in the frequency domain. With time you will appreciate the frequency domain as it allows to analyze the response of a system one frequency at a time. This is a much easier problem than dealing with all frequencies at once. More importantly, it is often possible to develop some intuition in the frequency domain or to zoom in on a small effect at a particular frequency on top of a large background. For instance, we know that a harmonic oscillator or pendulum tends to oscillate at its natural frequency.

Another good example is given by figure 1 that looks like a sinusoidal wave with some additional higher frequency component on top of the signal. The same signal in the frequency domain reveals a large amplitude, low frequency peak and a second higher frequency peak with a smaller amplitude.

Linear systems

The underlying reason why the method of Fourier analysis is so successful is that most physical phenomena are linear: If we increase the input (or stimulus) by a constant factor then the output (or response) of the system increases by that same constant factor. Many real-world systems respond in some way to stimuli, and are characterized by an input-output relationship. An electrical system, such as an amplifier in a stereo set, is an example of such a system – it has signal inputs from CD players, radio, etc., and outputs to speakers. This is an example of a 'linear system', because its output is proportional to the input. The system may become nonlinear if you turn-up the amplifier way to loud and

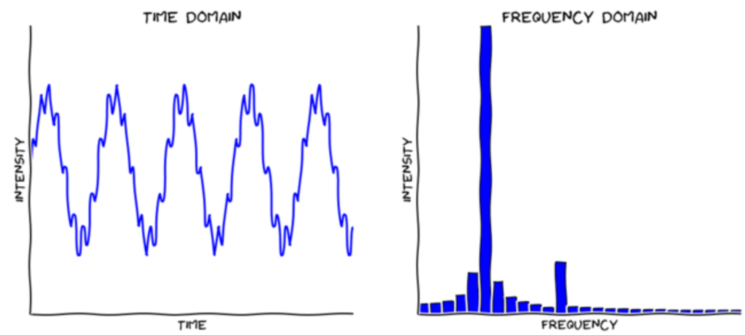


Figure 1: *An example of the use of Fourier analysis with intensity plotted in the time domain (left) and frequency domain (right). The frequency domain plot clearly reveals that the signal contains a slow and fast oscillation. (Image from* `https://lstsal.wordpress.com/2015/02/28/comparing-two-audio-files/`*)*

saturate the output. This produces sound waves that have their tops chopped off because there is a limit to how much power the amplifier can generate. You may hear different sounds than were on the original recording if the volume is too high (they tend to be screechy).

Linear systems can be more complicated than an amplifier, however. The general definition of a linear system is any system whose output $F(x)$ satisfies the following equation:

$$F(a\ g(t) + b\ h(t)) = a\ F(g(t)) + b\ F(h(t)) \tag{1}$$

where a and b are constant real numbers. The functions $g(t)$ and $h(t)$ are arbitrary functions of an independent variable t (commonly time is used a independent variable in applications, but it can also be a position). The definition shows that constant multiplicative scale factors can be 'factored-out'. More importantly, it tells us that the response of a linear system to a sum of inputs is the sum of the responses to each individual input separately.

Examples of linear systems

1. Wave propagation such as sound waves or electromagnetic waves
2. Electrical circuits composed of resistors, capacitors and inductors. These circuits can be designed as a frequency filter that only passes low frequencies and reject high, or vice-versa.
3. Mechanical motion of masses with springs and dashpots (to include damping)

4. Limiting cases of non-linear systems for small inputs are linear: Even if the output of a system is not exactly linear, it will often be approximately linear for small enough inputs.

5. Most optical systems are linear – lenses and optical fibers all have linear responses to how much light is put in. They may distort the input light in other ways (delays, changing the angle depending on color), but they are described by the linear equations above. Some optical systems are non-linear: for instance they change the frequency of the light that comes in (fluorescent materials for example).

6. The process of differentiation and integration are both linear systems

7. A convolution as a mathematical operation where each value in the output is expressed as a sum of values in the input multiplied by a set of weighing coëfficients. This operation plays an important role if one samples an ideal signal using a measurement apparatus with a finite resolution, e.g. the finite response time (inverse bandwidth) of an oscilloscope or the point-spread function of a telescope.

Fourier analysis can be used to construct the output of a linear system to an arbitrary input by thinking of this input as a sum of harmonic functions. This is a very powerful and useful method if the input-output relation for harmonic functions is easier to obtain and if we have a general method that can decompose an arbitrary function as a sum of harmonic functions.

Signal processing

Building your intuïtion in the frequency domain and thinking about signals as being composed of different frequency components is extremely useful in the field of signal processing. Consider the additive property of a linear system: The response of the system to input $x(t)$ results in an output $y(t)$. Then from Eqn 1 it follows that

$$\begin{aligned} x_1(t) &\rightarrow y_1(t) \\ x_2(t) &\rightarrow y_2(t) \\ c_1\ x_1(t) + c_2\ x_2(t) &\rightarrow c_1\ y_1(t) + c_2\ y_2(t) \end{aligned}$$

In electronics, circuits of capacitors, inductors and resistors can be designed to filter the input and/or output of a measurement device and reduce noise. To design such filters either requires solving a differential equation to describe the output voltage $V_{out}(t)$ as a function of the input voltage $V_{in}(t)$. An alternative description is to consider the signal as a sum of sinusoidal functions, i.e. to use

$$\begin{aligned} x_1(t) &= a_1 \sin(\omega_1 t) \\ x_2(t) &= a_2 \sin(\omega_2 t) \end{aligned}$$

where a_1 and a_2 indicate the amplitude of the two sinusoidal functions with frequencies ω_1 and ω_2. If the system is linear the response $y_1(t)$ and $y_2(t)$ is given by

$$
\begin{aligned}
y_1(t) &= \kappa_1 \cdot a_1 \sin(\omega_1 t + \phi_1) \\
y_2(t) &= \kappa_2 \cdot a_2 \sin(\omega_2 t + \phi_2)
\end{aligned}
$$

where κ_1 and κ_2 are multiplicative factors that depends on the frequency ω and the phase ϕ_1 and ϕ_2 take into account the possible phase-shift in the system. Instead of using the explicit time dependence we could describe the system by only using the frequencies ω_1 and ω_2. We will see later that a more general relation holds and that we can describe any signal as an infinite sum of sinusoidal functions. This is an important step because it allows to design electronic circuits in the frequency domain based on the output voltage $V(\omega)$ [1]. The response of the system can be decomposed in the frequency response of the individual components in the circuit. Because the system is linear simple rules exist to add the response of the different components.

Fourier analysis and applications in the physics B.Sc. program

The behavior of linear systems is part of the first year physics curriculum and you have already been exposed to simple electronic circuits, the harmonic oscillator, optical interference and diffraction. You know how to deal with these systems without being aware of the methods of Fourier analysis that can be applied to all of them. The mathematics of Fourier Series and Fourier Transforms is taught in the course 'Analyse 3 NA' for physics and astronomy students. This course uses the book K.F. Riley, M.P. Hobson, and S.J. Bence, *Mathematical Methods for Physics and Engineering*, 3^{rd} edition, (Cambridge University Press, New York, 2016). Fourier Series and Transforms are covered in Chapters 12 and 13 of this textbook (pages 421–458) and you will become familiar with the formal mathematical aspects of this material in the mathematics course.

The challenge for a physics student is to recognize a linear system and come up with a mathematical description of the system. With time and through practice the use of Fourier analysis will evolve and become part of your intuition as a physicist. Throughout the second and third year of your B.Sc. education it is assumed that you know about Fourier analysis and that you know how to apply these methods in a physics context.

[1] This response is a complex quantity and contains a frequency dependent amplitude and phase response

0.1 The Fourier Series

We first consider a situation encountered in physics where we study the output of a linear system to any periodic input signal. For instance we could investigate the output of an electronic circuit as a response to a periodic input voltage given by

$$V(t) = V(t+T)$$

where T is the time period of the periodic input voltage. A very similar situation occurs when we consider a mechanical system of a mass on a spring subject to a periodically varying external force given by

$$F(t) = F(t+T)$$

To find the response of the linear system we use the superposition principle that implies that (for a reasonably well-behaved function) it is possible to write the response to an arbitrary input as a sum of responses to simple harmonic (sine and cosine) functions with a fixed frequency. For this to work we must assume that any periodic function can be written as a sum of (infinitely many) cosine and sine functions. This is a much easier problem to solve because the harmonic function are much easier to deal with than an arbitrary function. Our strategy will thus be to solve the response one frequency at a time and then write the response to an arbitrary function as a series expansion in simple harmonic functions.

The series expansion of a periodic signal in terms of cosine and sine functions is called the Fourier Series. Let us first consider the real-valued function $x(t)$ that is periodic in time with a period T:

$$x(t) = x(t+T)$$

This function can be expressed as a Fourier Series defined by

$$\begin{aligned} x(t) = \frac{a_0}{2} &+ a_1 \cos \omega t + a_2 \cos 2\omega t + a_3 \cos 3\omega t + \ldots \\ &+ b_1 \sin \omega t + b_2 \sin 2\omega t + b_3 \sin 3\omega t + \ldots \\ x(t) = \frac{a_0}{2} &+ \sum_{n=1}^{\infty} a_n \cos n\omega t + b_n \sin n\omega t \qquad (2) \end{aligned}$$

where we have introduced the angular frequency

$$\omega = \frac{2\pi}{T}$$

Note that the sum runs over discrete frequencies given by $n\omega$ with n a positive integer. These functions have a periodicity given by T/n and are all periodic with period T. The case $n = 0$ reduces the cos-function to a constant and has been taken out of the summation. The extra factor $\frac{1}{2}$ for the Fourier coefficient a_0 will prove to be convenient in the definition of the coefficients a_n and b_n.

Exercise 1. The Fourier Series contains an infinite sum of periodic functions that are either *odd* or *even*. Show that any arbitrary function $x(t)$ can always be written as a sum of *even* and *odd* functions.

Most examples in this reader focus on signals that are periodic in time because most of the easily accessible examples are based on a physical observable that varies with time. This could be the position $x(t)$ of a mass on a spring, or the time-dependent voltage $V(t)$ in an electronic circuit. If these signals are periodic in time, the time-dependent variable can be represented as sum over discrete frequencies ω_n. There are however important physical phenomena that are periodic in the spatial coordinate. Consider for instance a wave with an amplitude $A(x)$. At a fixed time, we know that this function should be periodic in space as well with a periodicity given by the wavelength λ

$$A(x) = A(x + \lambda)$$

We could equally well write down a Fourier Series using coefficients a_n and b_n belonging to a sum of cosines and sines in terms of discrete spatial frequencies.

$$A(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos k_n x + b_n \sin k_n x$$

where we have introduced the wavenumber or spatial frequency $k_n = n\frac{2\pi}{\lambda}$. The signal that is periodic in space can be represented as a sum over discrete wavevectors k_n.

To illustrate that a set of harmonic functions, with discrete ω_n or k_n, can complete describe the physics of a system we consider as an example the harmonics of a vibrating string. With the string fixed on both sides the edges of the string should be the nodes of a sinusoidal wave. Because the string is fixed on both sides the fundamental mode of the string is a sinusoidal function that has a periodicity that is twice the length of the string. The Fourier Series considers a periodic function.

0.1.1 Fourier coefficients a_n and b_n cosines and sines

We have defined the series expansion of a periodic function in terms of sines and cosines, which allows to decompose any periodic function in a set of simple harmonic functions with discrete frequencies

$$\omega_n = n\frac{2\pi}{T}, \quad n \in \mathbb{N}$$

We are left with the mathematical task to calculate the Fourier coefficients a_n and b_n. These coefficients can be calculated from the equation that defines the Fourier Series if we assume that the series exists and that it is convergent. We do not prove this, but simply state that the conditions to have a smooth and continuous function that describes a physical phenomenon is a sufficient condition for the series to exist and to be convergent. The coefficient a_0 can be found by integrating the left and right hand sides of equation 2 that defines the Fourier Series.

$$\begin{aligned}
x(t) &= \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos n\omega t + b_n \sin n\omega t \Leftrightarrow \\
\int_{t_0}^{t_0+T} x(t)dt &= \int_{t_0}^{t_0+T} \left(\frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos n\omega t + b_n \sin n\omega t \right) dt \Leftrightarrow \\
\frac{a_0}{2}T &= \int_{t_0}^{t_0+T} x(t) \Leftrightarrow \\
a_0 &= \frac{2}{T} \int_{t_0}^{t_0+T} x(t)dt
\end{aligned}$$

The other Fourier coefficients can be found if we multiply the left and right hand side of equation 2 by $\cos n\omega t$ or $\sin n\omega t$ and integrate both sides. For example the coefficient a_n can be found via

$$\begin{aligned}
x(t)\cos n\omega t &= \cos n\omega t \left(\frac{a_0}{2} + \sum_{m=1}^{\infty} a_m \cos m\omega t + b_m \sin m\omega t \right) \Leftrightarrow \\
\int_{t_0}^{t_0+T} x(t)\cos n\omega t dt &= \int_{t_0}^{t_0+T} \cos n\omega t \left(\frac{a_0}{2} + \sum_{m=1}^{\infty} a_m \cos m\omega t + b_m \sin m\omega t \right) dt \Leftrightarrow \\
\int_{t_0}^{t_0+T} x(t)\cos n\omega t dt &= \int_{t_0}^{t_0+T} a_n \cos^2 n\omega t dt = \frac{a_n T}{2}
\end{aligned}$$

In the last step we used that all but one of the integrals in the summation evaluate to zero.[2] This leads to the following integrals that define the Fourier coefficients:

$$\begin{aligned}
a_0 &= \frac{2}{T} \int_{t_0}^{t_0+T} x(t)dt && (3) \\
a_n &= \frac{2}{T} \int_{t_0}^{t_0+T} x(t)\cos n\omega t dt,\ n \in \mathbb{N} && (4) \\
b_n &= \frac{2}{T} \int_{t_0}^{t_0+T} x(t)\sin n\omega t dt,\ n \in \mathbb{N} && (5)
\end{aligned}$$

---

[2]To show this you can set $t_0 = -T/2$ and use the fact that $\sin m\omega t \cos n\omega t$ is an odd function. The integral $\int \cos m\omega t \cos n\omega t$ can be rewritten using complex exponentials, i.e. use $2\cos x = \exp ix + \exp -x$.

Exercise 2. *Fourier Series of triangular wave* The Fourier Series of the triangular wave shows fast convergence and only a few terms are necessary to obtain a good approximation of this waveform. The triangular wave is defined by the periodic function

$$y(t) = \begin{cases} A(1+\frac{4t}{T}) & -\frac{T}{2} \leq t < 0 \\ A(1-\frac{4t}{T}) & 0 \leq t \leq \frac{T}{2} \end{cases}$$

Determine the Fourier Series of $y(t)$.

The Fourier Series defined using the coefficients a_n and b_n are most useful if we deal with real-valued functions and if we can exploit that the function is either an *even* or *odd* function. In this case many of the Fourier coefficients will be zero. That this is often true can be seen in figure 2 that shows the Fourier Series for a square wave input. Compared to the triangular wave in the example it should be noted that the convergence of the square wave is much slower and that there are significant deviations (overshoot and undershoot) at $t = 0$ and $t = T/2$ where the original function is discontinuous. These deviations persist as we try to approximate a discontinuous function with a sum of continuous functions. [3]

Exercise 3. *Fourier Series of the square wave* Consider the periodic square wave function $y(t)$ defined by

$$y(t) = \begin{cases} A & 0 \leq t \leq T/2 \\ -A & T/2 \leq t \leq T \end{cases}$$

Determine the Fourier coefficients of the Fourier Series defined by

$$y(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(n \cdot 2\pi\nu t) + \sum_{n=1}^{\infty} b_n \sin(n \cdot 2\pi\nu t)$$

where we introduced the frequency $\nu = \frac{1}{T}$. Start by showing that all coefficients a_n are equal to zero. Show that the coefficients $b_n = 0$ for even n.

Exercise 4. Suppose we shift the origin of the square-wave by $T/2$, i.e. we define $\tilde{y}(t) = y(t - T/4)$ to create a function that is symmetric around the y-axis ($y(t) = y(-t)$). Use the result of the previous exercise. What can you say about the Fourier coefficients a_n and b_n of $\tilde{y}(t)$?

---

[3] These non-vanishing overshoots and undershoots are known as Gibbs phenomenon `https://en.wikipedia.org/wiki/Gibbs_phenomenon` and can be solved with additional mathematical methods. In physics, the input is always a continuous function as there is a smallest timescale at which the input becomes continuous. This removes the Gibbs phenomenon because there is a maximum frequency in the problem that is relevant.

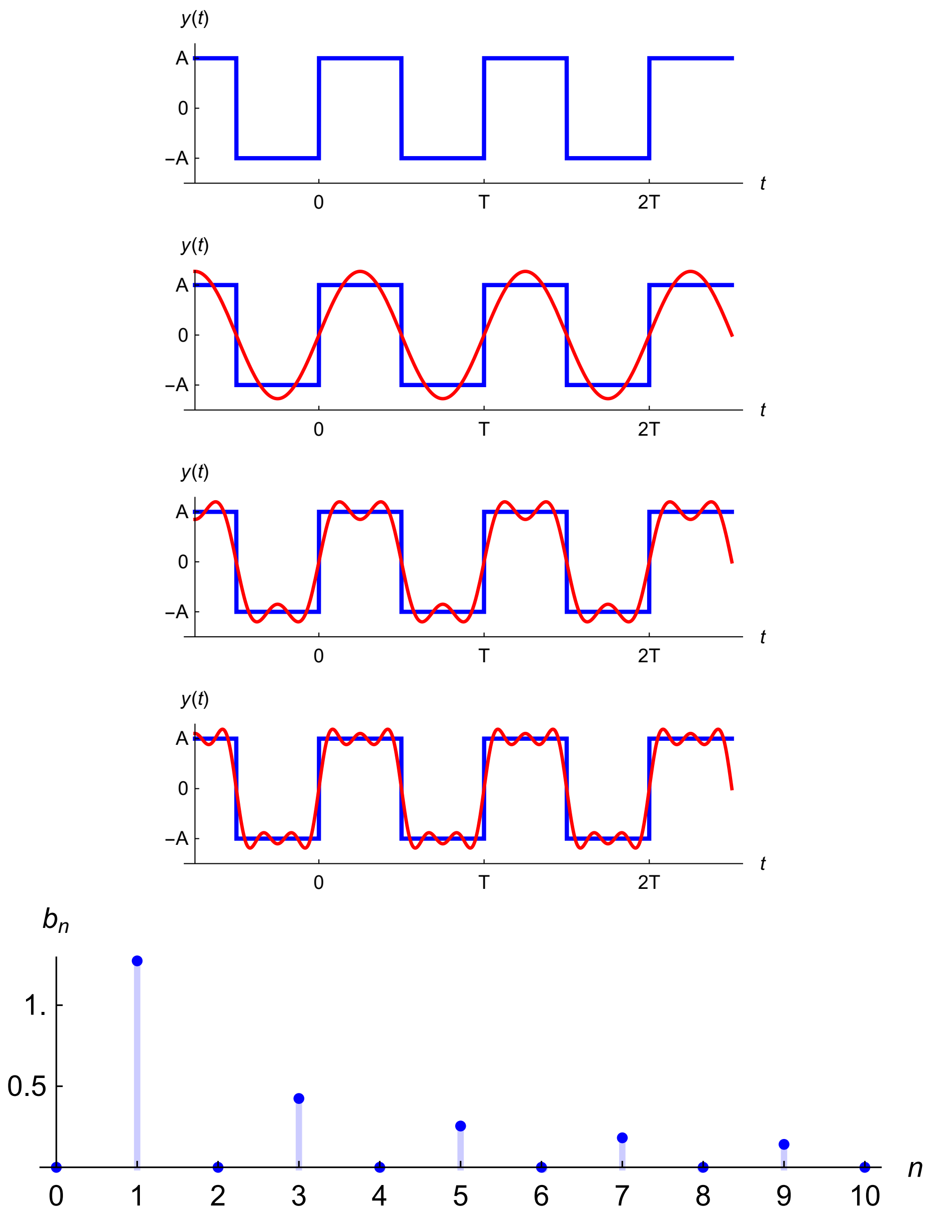


Figure 2: *Fourier Series of the square wave $y(t)$ defined as an odd function (blue curve). Progressive approximations from top to bottom using a sum of one, two and three sine functions. The bottom figure shows the importance of higher order Fourier coefficients b_n as a function of n.*

0.1.2 Fourier coefficients c_n: complex exponentials

The Fourier Series with coefficients a_n and b_n are defined using a series expansion in sines and cosines. This requires three slightly different expressions to calculate the coefficients. An alternative, more concise, notation of Fourier Series exists that makes use of complex exponentials. This is the one most frequently encountered Fourier Series in physics textbooks because the complex exponential is easier to handle than the sine and cosine functions.

The complex Fourier Series is defined by [4]

$$\begin{aligned} x(t) &= c_0 + \sum_{n=1}^{\infty}(c_n e^{in\omega t} + c_{-n}e^{-in\omega t}) \\ &= \sum_{-\infty}^{\infty} c_n e^{in\omega t} \end{aligned}$$

with complex Fourier coefficients c_n defined as

$$c_n = \frac{1}{T}\int_{-T/2}^{T/2} x(t)e^{-in\omega t}dt,\ n \in \mathbb{Z}$$

Note that the summation now runs over both positive and negative values of n so that the series expansion contains both positive and negative frequencies. One can easily convert the Fourier Series written in sines and cosines (equation 2) using Euler's formula:

$$e^{ix} = \cos x + i \sin x$$

You will find that a_n and b_n can be written in terms of c_n and c_{-n} according to:

$$\begin{aligned} a_n &= c_n + c_{-n} = 2\Re(c_n) \\ b_n &= i(c_n - c_{-n}) = -2\Im(c_n) \end{aligned}$$

With $\Re(c_n)$ the real part and $\Im(c_n)$ the imaginary part of c_n. An alternative formulation that is often encountered is to use the amplitude and the phase instead of the complex number c_n by writing $c_n = |c_n|e^{i\phi_n}$

$$\begin{aligned} C_n = |c_n| &= \frac{1}{2}\sqrt{a_n^2 + b_n^2} \\ \phi_n = \arg c_n &= \arctan\frac{-b_n}{a_n} \end{aligned}$$

[4] The definition here links the Fourier coefficient c_n to an angular frequency $n\omega$. The opposite definition, that links the coefficient c_n to a negative angular frequency $-n\omega$ is also encountered. Typically, this introduces a few extra minus signs in the equations. Mixing the two definitions will lead to an arbitrary number of sign errors and should be avoided.

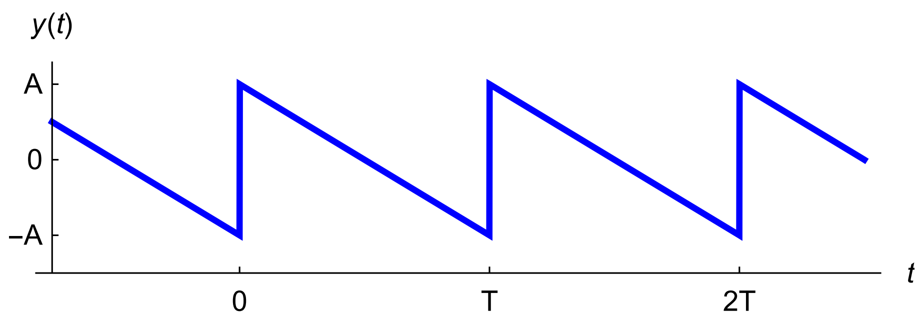


Figure 3: *Plot of the sawtooth waveform*

It is important to keep in mind that measurable quantities in physics are always real, which places bounds on the coefficients. For a real-valued function $x(t)$ the Fourier coefficients a_n and b_n are real-valued as well. Using the complex notation the coefficients c_n and c_{-n} satisfy $c_n = c^*_{-n}$. The response of a physical system can be thought of as an amplitude response using the coefficient $C_n = |c_n|$ and a phase ϕ_n.

Exercise 5. Given is the sawtooth periodic waveform defined by

$$y(t) = A\left(1 - \frac{2t}{T}\right)$$

Calculate the complex Fourier coefficients c_n for the sawtooth waveform given by $y(t)$

0.2 The Fourier Transform

The Fourier Series is limited to functions that are strictly periodic. Functions that are periodic in time (space) can be represented as a discrete sum over discrete frequencies (wavevectors) that are an integer multiple of the fundamental frequency (wavector) defined by the periodicity of the underlying function. To describe the response of physical systems it would be advantageous to extend the description to non-periodic systems and replace the discrete sum over frequencies (wavectors) by an integral.

The result is known as the Fourier Transform which gives a description in terms of a continuous spectrum of frequencies (wavevectors). The physics of a linear system is described by a (set of coupled) linear differential equation(s) as a function of time or space coordinates. These equations can be Fourier Transformed to change the differential equation in the time (space) domain into an algebraic equation in the frequency (wavevector) domain that defines the response function $H(\omega)$, this will be discussed in chapter 4.

0.2.1 Fourier Transform as a limiting case of the Fourier Series

A formal way to introduce Fourier Transforms is to consider the limiting case of a Fourier Series when the periodicity T becoming infinite. As long as we observe the response of a system for a finite time we should be unable to tell if the system is periodic or not. We take the limit $T \to \infty$ of the complex Fourier Series with coefficients c_n. To avoid problems with the notation we use the symbol $\omega_0 = \frac{2\pi}{T}$ for the fundamental harmonic. The limit $T \to \infty$ is the same as taking the limit $\omega_0 \downarrow 0$. Note that the frequency ω in the final integral expressions is given by $n\omega_0$ in the discrete Fourier sum. This yields

$$
\begin{aligned}
x(t) &= \lim_{T\to\infty} \sum_{-\infty}^{\infty} c_n e^{in\omega_0 t} \\
&= \lim_{T\to\infty} \sum_{-\infty}^{\infty} e^{in\omega_0 t} \left(\frac{1}{T} \int_{-1/2T}^{1/2T} x(t) e^{-in\omega_0 t} dt \right) \\
&= \lim_{\omega_0 \downarrow 0} \sum_{-\infty}^{\infty} e^{in\omega_0 t} \left(\frac{1}{2\pi}\omega_0 \int_{-1/2T}^{1/2T} x(t) e^{-in\omega_0 t} dt \right) \\
&= \int_{-\infty}^{\infty} e^{i\omega t} \left(\int_{-\infty}^{\infty} x(t) e^{-i\omega t} dt \right) \frac{1}{2\pi} d\omega \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega t} \left(\int_{-\infty}^{\infty} x(t) e^{-i\omega t} dt \right) d\omega \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) e^{i\omega t} d\omega
\end{aligned}
$$

The Fourier Transform $X(\omega)$ of $x(t)$ is given by:

$$
\boxed{X(\omega) = \mathcal{F}[x(t)](\omega) = \int_{-\infty}^{\infty} x(t) e^{-i\omega t} dt}
$$

Here the symbol $\mathcal{F}$ denotes a Fourier Transform. Similarly the inverse transform

$$
\boxed{x(t) = \mathcal{F}^{-1}[X(\omega)](t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) e^{i\omega t} d\omega}
$$

is also a Fourier Transform that translates from the frequency domain back to the time-domain.

0.2.2 Alternative definitions of the Fourier Transform

Unfortunately, the way we introduced the Fourier Transform is not the only way to define the transform and at least three different conventions are used in literature. Mathematicians (e.g. in AN3NA) prefer a more symmetric definition of the transformations (where both transformations include a factor $\frac{1}{\sqrt{2\pi}}$). We will use the same notation as we use in Physics Experiments 2. The reason for our convention is that it is easier to convert between different Fourier Transforms (Fourier Series, Fourier Transform, Discrete Fourier Transform, Discrete Time Fourier Transform). Furthermore, we will often explicitly perform transformations from the time to the Fourier domain, and perform mathematical operations in the Fourier domain. Our convention causes much fewer factors 2π in our mathematical operations. Also, some relations, as Parseval's theorem, are more in line with our convention.

In this course, as in PE2, we will give a lot of attention to keep everything in line with our chosen convention. However, one cannot simply combine results from different sources or use library functions in a programming language such as python without verifying the convention. Also note that in previous years, the symmetric convention was used in this course, so it could be that there are some leftovers in the solutions of exercises and/or in this reader.

0.2.3 Parseval's theorem

Parseval's theorem states that

$$\int_{-\infty}^{\infty} |f(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |F(\omega)|^2 d\omega = \int_{-\infty}^{\infty} |F(\nu)|^2 d\nu$$

With the frequency ν related to the angular frequency ω with $\omega = 2\pi\nu$.

Exercise 6. (a) Given is the complex function $g(k,x) = \mathrm{e}^{ikx} = \cos kx + i \sin kx$. Show that $\cos kx$ en $\sin kx$ can be written as complex exponentials

(b) Calculate $\frac{\mathrm{d}}{\mathrm{d}x} \cos(kx)$ en $\frac{\mathrm{d}}{\mathrm{d}x} \sin(kx)$

(c) Calculate $\frac{\mathrm{d}}{\mathrm{d}x} \mathrm{e}^{i(kx+\varphi)}$. Explain in your own words why it is more advantageous to use the function $\frac{\mathrm{d}}{\mathrm{d}x} \mathrm{e}^{i(kx+\varphi)}$ compared to $\frac{\mathrm{d}}{\mathrm{d}x} \cos(kx+\varphi)$ when using the functions as a trial function to solve a differential equation.

0.2.4 Fourier Transform of the derivative

In Exercise 5 we have seen that calculating the derivative of the complex exponential is the same as multiplying by the function by a factor (ik). This simple rule can be exploited to calculate the Fourier Transform of the derivative of a function.

Let us now consider a function $f(t)$ in the time-domain with a Fourier Transform $F(\omega)$ as a function of frequency ω defined by

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t}dt$$

The Fourier Transformation of the first order derivative of $f(t)$ can be represented as $F^{(1)}(\omega)$ and can be calculated using integration by parts

$$\begin{aligned}
F^{(1)}(\omega) &= \int_{-\infty}^{\infty} \frac{df(t)}{dt} e^{-i\omega t}dt \\
&= e^{-i\omega t} f(t)|_{-\infty}^{\infty} + i\omega \int_{-\infty}^{\infty} f(t)e^{-i\omega t}dt \\
&= i\omega F(\omega)
\end{aligned}$$

In the last line we have used that $f(t) = 0$ when $t \to \pm\infty$ (see next section why this can be assumed). Similarly, higher order derivatives can be calculated and yield

$$F^{(n)}(\omega) = (i\omega)^n F(\omega)$$

The derivative of a function is thus a simple multiplication in the Fourier domain. This can be used to reduce any linear differential equation to a linear equation in the Fourier domain.

0.2.5 Fourier Transforms frequently encountered in physics

The Fourier Transform of many functions is known and tables of Fourier Transforms can be found easily on the internet (e.g. use Wolfram alpha `http://www.wolframalpha.com` and search for 'Fourier Transform sin(x)'). The table here is restricted to a few one-dimensional transforms that are frequently encountered in physics.

Conditions for the existence of the Fourier Transform are complicated to state exactly. A sufficient condition for the existence of the Fourier Transform of $x(t)$ is that $x(t)$ is 'absolutely integrable', i.e.

$$||x(t)|| \triangleq \int_{-\infty}^{\infty} |x(t)|dt < \infty$$

Another sufficient condition is that the function $x(t)$ is 'square integrable', defined as

$$||x(t)||^2 \triangleq \int_{-\infty}^{\infty} |x(t)|^2 dt < \infty$$

For real-world physics problems and signals there is never a question of existence, of course because the function will always go to zero at $\pm\infty$. However, when dealing with idealized signals and models we often use functions, such as the sine and cosine function, that go on forever in time. These functions do pose normalization difficulties that can be resolved (see e.g. exercise 2.2) if we use Dirac's delta function defined via

$$\int_{-\infty}^{\infty} f(x)\delta(x-a)dx = f(a)$$

The property above defines the Dirac's delta function (defining property).

Exercise 7. (a) Determine the Fourier Transform of Dirac's delta function $\delta(t-\tau)$ using the defining property of the function.

(b) Use the inverse Fourier Transform on your previous answer to show the following useful property:

$$\delta(t-\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-i(t-\tau)\omega} d\omega$$

(c) Find the Fourier Transform of the functions $\cos(\omega_0 t)$ en $\sin(\omega_0 t)$ using Euler's formula.

We give the following useful table of Fourier Transforms

$$\begin{aligned}
\mathcal{F}(\alpha e^{i\omega_0 t}) &= 2\pi\delta(\omega-\omega_0) \\
\mathcal{F}(\cos(\omega_0 t)) &= \pi\left(\delta(\omega-\omega_0)+\delta(\omega+\omega_0)\right) \\
\mathcal{F}(\sin(\omega_0 t)) &= \frac{\pi}{i}\left(\delta(\omega-\omega_0)-\delta(\omega+\omega_0)\right) \\
\mathcal{F}(\alpha\delta(t-\tau)) &= \alpha e^{-i\omega\tau} \\
\mathcal{F}(\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{t^2}{2\sigma^2}}) &= e^{-\frac{\sigma^2\omega^2}{2}} \\
\mathcal{F}(e^{-\alpha|t|}) &= \frac{2\alpha}{\alpha^2+\omega^2}
\end{aligned}$$

From the table we see that Fourier Transform of a Gaussian function is again a Gaussian function with a standard deviation that is inversely proportional to the standard deviation, i.e. the width of the Fourier Transform is given by $\frac{1}{\sigma}$. As a consequence the product of the width of the original Gaussian and its Fourier Transform is constant. This is known as a 'Fourier relation' and plays an important role in the wave description of physical phenomena.

To be able to deal with an exponential decay we used the absolute value $|t|$ to ensure that the function is absolutely integrable. The Fourier Transforms of this exponential fucntion is known as the Lorentzian function and is frequently encountered as the frequency response of a resonant phenomenon. This function is found as the response of damped harmonic oscillators, the electronic RLC circuit and the emission spectrum of light emitted by an atom.

Chapter 1

Electronic circuits

1.1 Linear time-invariant systems

In this chapter we will examine systems that transform an input signal $x(t)$ into an output signal $y(t)$. In this reader we will focus on systems that are made up of electronic components such as resistors, inductors and capacitors but much of the theory is also applicable to mechanical systems. A special class of systems are linear time-invariant systems.

- **Linear system:** If $x_1(t) \to y_1(t)$ and $x_2(t) \to y_2(t)$ then the system is linear when $Ax_1(t) + Bx_2(t) \to Ay_1(t) + By_2(t)$.
- **Time-invariant system:** If $x(t) \to y(t)$ then the system is time invariant if $x(t-\tau) \to y(t-\tau)$.

Many systems can, for small values of the input x, be approximated as linear. Take for example the equation of motion for a mass-spring system which can be approximated to $F = -ku$ for small values of the displacement u. When a system is linear and time-invariant, the differential equations that characterize that system are also linear. This allows us to easily determine the eigenfunctions of the system.

Wikipedia: **`http://en.wikipedia.org/wiki/LTI_system_theory`**

1.1.1 Eigenfunctions of linear time-invariant systems

Consider a linear time-invariant system that transforms an input signal $x(t)$ into an output signal $y(t)$ which is described by a linear differential equation of the form:

$$a_0x + a_1\dot{x} + a_2\ddot{x} + ... + a_nx^{(n)} = b_0y + b_1\dot{y} + b_2\ddot{y} + ... + b_my^{(m)}\,. \tag{1.1}$$

We know from calculus and linear algebra that we may solve this differential equation by substituting $x(t) = e^{st}$, where s is a complex number, into the

Figure 1.1: *A system transforms an input signal $x(t)$ in to an output signal $y(t)$.*

equation. The functions e^{st} form a basis of eigenfunctions for the system. A solution for $y(t)$ is then given by:

$$y(t) = \frac{a_0 + a_1 s + a_2 s^2 + ... + a_n s^n}{b_0 + b_1 s + b_2 s^2 + ... + b_m s^m} e^{st} \overset{\text{def}}{=} H(s) e^{st} . \tag{1.2}$$

In the last line we introduced the transfer function of the system $H(s)$. When the input signal can be written as a sum of functions of the form e^{st}, i.e. in a basis of eigenfunctions, then we can find the output signal using this transfer function. Note that the transfer function enables a direct multiplication of the Fourier Transform of the input signal with a polynomial to obtain the Fourier Transform of the output. In the time domain, with the functions $x(t)$ and $y(t)$, such relation did not exist.

A signal that does not diverge with time can be Fourier transformed. If this is the case eigenfunctions of the form $e^{i\omega t}$, where ω is a real number, will be sufficient to describe the system. It is only in Chapter 4 that we will use the more general transform, the Laplace Transform, to enable us to work with input signals that do diverge (they continually grow with time).

Wikipedia: **`http://en.wikipedia.org/wiki/LTI_system_theory`**

1.2 The Fourier Transform

In the previous section we saw that an input signal $e^{i\omega t}$ gives an output signal of the same form up to a complex factor. This invites us to make use of the Fourier Transform to further examine the behavior of those signals. Let us define $X(\omega)$, the Fourier Transform of $x(t)$ by:[1]

$$X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-i\omega t}\,\mathrm{d}t\,, \tag{1.3a}$$

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)e^{i\omega t}\,\mathrm{d}\omega\,, \tag{1.3b}$$

where ω is called the angular frequency and is related to the frequency f as $\omega = 2\pi f$. Note that the factor $1/2\pi$ in the reverse transform is a matter of convention. Alternatively some textbooks use a factor of $1/\sqrt{2\pi}$ in front both transforms. Unfortunately this means that one cannot blindly combine Fourier related identities from different sources. We shall examine the Fourier Transform and a few special cases of Fourier analysis in more detail in Chapter 2.

Using the linearity of our system and Equation 1.2 with $s = i\omega$ we can write the output signal $y(t)$ as:

$$y(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} H(\omega)X(\omega)e^{i\omega t}\,\mathrm{d}\omega\,, \tag{1.4a}$$

$$Y(\omega) = H(\omega)X(\omega)\,. \tag{1.4b}$$

Evidently the output signal $Y(\omega)$ in the frequency domain is determined completely by the transfer function $H(\omega)$ and $X(\omega)$.

Wikipedia: **`http://en.wikipedia.org/wiki/Fourier_transform`**

[1]The prefactors of the integrals may be defined differently, so long as the product of the two is $\frac{1}{2\pi}$.

1.3 Impedance

In the case of an electronic circuit the transfer function can be easily obtained. To do this we must examine what happens to the differential equations describing the capacitor and inductor when we apply a Fourier Transform to them. We will then look at an example of how to determine the transfer function of a simple electronic circuit.

1.3.1 The capacitor

The voltage $u_C(t)$ across a capacitor is proportional to the charge accumulated inside it due to the current $i_C(t)$:

$$u_C(t) = \frac{1}{C}\int_{-\infty}^{t} i_C(\tilde{t})\,\mathrm{d}\tilde{t}\,, \tag{1.5}$$

where C is the capacitance of the capacitor. Substituting the expression for the Fourier Transform (Equation 1.3b) for the voltage and current gives:

$$\begin{aligned}
\frac{1}{2\pi}\int_{-\infty}^{\infty} U_C(\omega)e^{i\omega t}\mathrm{d}\omega &= \frac{1}{C}\int_{-\infty}^{t}\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} I_C(\omega)e^{i\omega\tilde{t}}\mathrm{d}\omega\right)\mathrm{d}\tilde{t}\\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\frac{I_C(\omega)}{C}\int_{-\infty}^{t} e^{i\omega\tilde{t}}\,\mathrm{d}\tilde{t}\right)\mathrm{d}\omega\\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{I_C(\omega)}{i\omega C}e^{i\omega t}\,\mathrm{d}\omega\\
U_C(\omega) &= \frac{1}{i\omega C}I_C(\omega)\,. \qquad (1.6)
\end{aligned}$$

We can see from Equations 1.4a and 1.6 that the transfer function $H(\omega)$ for the current to the voltage across the capacitor is equal to $1/i\omega C$. For an Ohmic resistor this relationship is equal to R. We can see that for a capacitor we also obtain a simply linear relationship between the current and the voltage in the frequency domain. This (complex) "resistance" in the frequency domain is called the impedance and is written as the letter Z.

Wikipedia: `http://en.wikipedia.org/wiki/Capacitor`

1.3.2 The inductor

The voltage across an inductor $u_L(t)$ is as a function of the time derivative of the current through the inductor $i_L(t)$:

$$u_L(t) = L\frac{d}{dt}i_L(t)\,, \tag{1.7}$$

with L the inductance of the inductor. Using the time-derivative of the Fourier Transform (we leave this as an exercise to the reader) gives:

$$U_L(\omega) = i\omega L\ \cdot I_L(\omega)\,. \tag{1.8}$$

Evidently the impedance of the inductor is equal to $i\omega L$.

Wikipedia: **`http://en.wikipedia.org/wiki/Inductor`**

1.3.3 Overview of impedances

- **Impedance:** The (complex) “resistance" that indicates the relationship between the current and voltage in the frequency domain.

- **Total impedance in series:** The total impedance of components connected in series is calculated in the same manner as the total resistance of Ohmic resistors in series:

$$Z_{tot} = Z_1 + Z_2 + ... \tag{1.9}$$

- **Total impedance in parallel:** The total impedance Z_{tot} of components connected in parallel is given by:

$$\frac{1}{Z_{tot}} = \frac{1}{Z_1} + \frac{1}{Z_2} + ... \tag{1.10}$$

Component	**Differential equation**	**Impedance**
Resistor	$u_R(t) = Ri_R(t)$	R
Capacitor	$u_C(t) = \frac{1}{C}\int\ i_C(t)\,\mathrm{d}t$	$\frac{1}{i\omega C}$
Inductor	$u_L(t) = L\frac{di_L(t)}{dt}$	$i\omega L$

Table 1.1: *Overview of the impedances and differential equations for the three most important passive components used in electronic circuits.*

Wikipedia: **`http://en.wikipedia.org/wiki/Electrical_impedance`**

1.4 The transfer function

We came across the transfer function $H(\omega)$ several times in the previous sections. Equation 1.4a shows that the transfer function reveals the relationship between the outgoing signal and the incoming signal. In the case of an electronic circuit the signals are usually in the form of a voltage or a current.

By using the impedances in Table 1.1 we can greatly simplify the process of solving the differential equations that describe an electronic circuit and as a result simplify the process of determining the transfer function of that circuit.

Wikipedia: **`http://en.wikipedia.org/wiki/Transfer_function`**

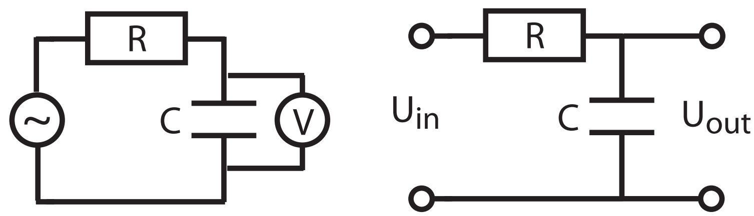


Figure 1.2: *The above circuit is an example of a low-pass filter, see Section 1.5. On the left the circuit is drawn with a variable voltage source U_{in} and a voltmeter U_{out} across the capacitor. On the right is the same circuit but drawn in the style that is commonly used in other books and will be used in this reader.*

1.4.1 Example: Calculating the transfer function

- **Question:** Give the transfer function $H(\omega) = U_{out}(\omega)/U_{in}(\omega)$ of the electronic circuit in Figure 1.2.

- **Answer:** We can use the impedances of the components in the frequency domain to determine the input and output voltage:

$$U_{out}(\omega) = \frac{1}{i\omega C} I(\omega)\,, \tag{1.11a}$$

$$U_{in}(\omega) = RI(\omega) + \frac{1}{i\omega C} I(\omega)\,. \tag{1.11b}$$

The transfer function of this system is thus:

$$H(\omega) = \frac{U_{out}(\omega)}{U_{in}(\omega)} = \frac{1}{1 + i\omega RC} = \frac{1}{1 + i\omega/\omega_{RC}}\,, \tag{1.12}$$

with $\omega_{RC} = 1/RC$ the characteristic frequency. We can see that for low values of ω the outgoing signal is equal to the incoming signal and for high values of ω the outgoing signal goes to 0. This filter is therefore often referred to as a low-pass filter, see Section 1.5.

Wikipedia: **`http://en.wikipedia.org/wiki/RC_circuit`**

1.4.2 Order of the transfer function

The transfer function is determined by solving the differential equations that describe the system. Every derivative of the differential equations introduces a power of ω as a prefactor when we substitute in $e^{i\omega t}$ as the input signal. An important characteristic of a system is the order of the differential equation because it determines the powers of ω in the transfer function.

- **Order of the transfer function:** The order of a filter or a transfer function $H(\omega)$ is equal to the order of the differential equation that describes the system.

The order of the transfer function is also equal to the number of poles of the function, which is often equivalent to the highest order of ω in the denominator of the transfer function.[2] We will see the importance of the order when we draw the transfer function in the Section 1.4.3.

Example: Order of a transfer function

- **Question:** What is the order of the transfer function associated with Figure 1.2?
- **Answer:** We solved this system without using the time-dependent differential equation for $u_{in}(t)$ and $u_{out}(t)$. The transfer function (Equation 1.12) has a single pole ($\omega = i/RC$). The order of the transfer function is thus equal to 1.

1.4.3 Bode plot

To visualize the characteristics of a system we can draw Bode plots of the transfer function. A Bode plot consists of **two** graphs: the Bode magnitude plot and the Bode phase plot.

- **Bode magnitude plot:** Plot of $20\log_{10}|H(\omega)|$, where $|H(\omega)| = \frac{|U_{out}(\omega)|}{|U_{in}(\omega)|}$, the modulus of the transfer function, is plotted against the frequency ω. The horizontal axis is on a logarithmic scale, the vertical axis is expressed in decibels (dB).
- **Bode phase plot:** Plot of the phase, or argument $\arg(H(\omega))$, of the transfer function $H(\omega)$ against the frequency ω. Note that the argument of the transfer function equals the phase difference between the input and output signal for a given frequency: $\arg(H(\omega)) = \arg(U_{out}(\omega)) - \arg(U_{in}(\omega))$. The horizontal axis is on a logarithmic scale, the vertical axis is expressed in degrees or radians.

When taking the modulus or the argument of a fraction, the following standard relations for complex numbers can be useful:

$$\left|\frac{a}{b}\right| = \frac{|a|}{|b|}\,, \tag{1.13a}$$

$$\arg\left(\frac{a}{b}\right) = \arg(a) - \arg(b)\,. \tag{1.13b}$$

The reason for choosing a log-log plot for the Bode magnitude plot will become clear when we look at Figure 1.3 where the Bode plots of the RC filter

[2] If for example both the numerator and the denominator can be divided by the same factor with ω then this is not the case.

in Figure 1.2 have been drawn. We see that for low values of ω the modulus of $H(\omega)$ is approximately equal to 1. For $\omega \gg 1/RC$ we get a slope of -20 dB per tenfold (decade) increase of the frequency in our Bode magnitude plot. [3] In the following example we will explain step by step how such a bode plot can be sketched for a given transfer function. In Section 1.5 we will see that such a Bode plot is instrumental in analyzing the filtering properties of a circuit.

Example: Drawing a Bode plot

- **Exercise:** Draw the Bode plots corresponding to the transfer function in Equation 1.12.

- **Solution:** The Bode plots in Figure 1.3 have been created using a computer. To draw a Bode plot by hand, one first determines what happens at very low and at very high frequencies and at the system's characteristic frequency. One then draws a curve going from the low frequency asymptote tot the high frequency asymptote, crossing the value found at the characteristic frequency.

  The characteristic frequency for this system is given by $\omega_{RC} = 1/RC$. We first calculate the modulus of the transfer function for the three regimes:

  – For $\omega \ll \omega_{RC}$,

$$|H(\omega)| = \left|\frac{1}{1 + i\omega/\omega_{RC}}\right| \approx 1$$
$$20 \log_{10}(|H(\omega)|) \approx 0\,.$$

  – For $\omega = \omega_{RC}$,

$$|H(\omega)| = \left|\frac{1}{1 + i}\right| = \frac{1}{\sqrt{2}}$$
$$20 \log_{10}(|H(\omega)|) \approx -3\,.$$

  – For $\omega \gg \omega_{RC}$,

$$|H(\omega)| = \left|\frac{1}{1 + i\omega/\omega_{RC}}\right| \approx \frac{1}{\omega/\omega_{RC}}$$
$$20 \log_{10}(|H(\omega)|) \approx -20 \log_{10}(\omega) + 20 \log_{10}(\omega_{RC})\,.$$

  We see that we get a slope of -20 dB per decade (tenfold increase of the frequency). [4]

  Having found both asymptotes and the characteristic frequency, the connecting curve will resemble that in Figure 1.3. Now we determine the argument of the transfer function for the same three regimes:

[3] Also approximately equal to 6 dB per doubling of the frequency (octave).

[4] While it is true that for $\omega \to \infty$ the value of $|H(\omega)|$ goes to zero, this is not very illuminating. We are interested in how fast it approaches 0.

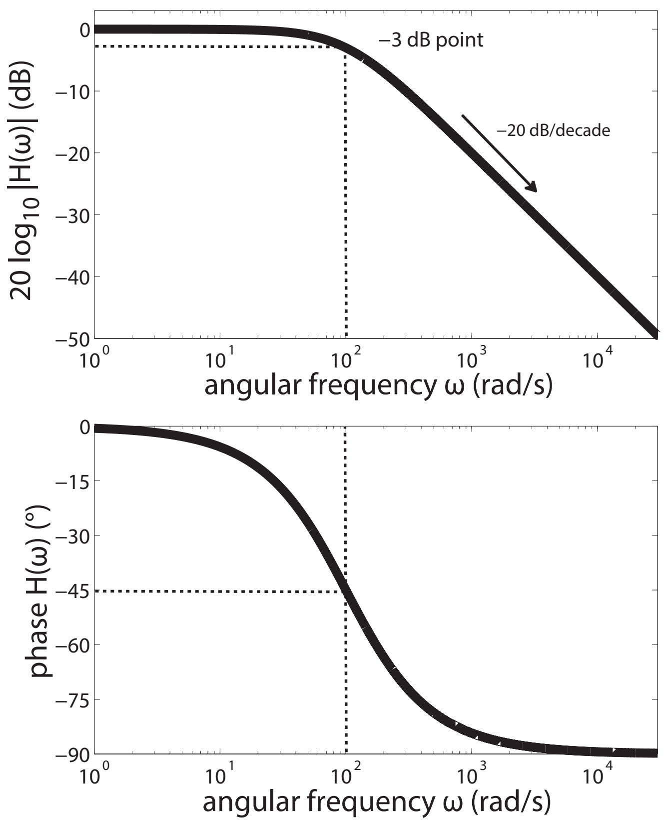


Figure 1.3: Top: *A Bode magnitude plot for the RC filter from Figure 1.2 with* $RC = 0.01s$*. The characteristic -3 dB point has been marked, this is the point where the frequency is equal to the cutoff frequency* $\omega_{RC} = 1/RC$*. We see that for first order transfer functions the amplitude decreases with 20 dB with every tenfold increase in the frequency (which is equal to a decade.)* ***Bottom:*** *A Bode phase plot for the same filter. The phase difference between the input and output voltage can be used to experimentally determine the cutoff frequency* ω_C*.*

– For $\omega \ll \omega_{RC}$,

$$\arg\left(H(\omega)\right) \approx \arg\left(\frac{1}{1}\right) = 0\,.$$

– For $\omega = \omega_{RC}$,

$$\begin{aligned}\arg\left(H(\omega)\right) &= \arg\left(\frac{1}{1+i}\right)\\ &= \arg(1) - \arg(1+i) = 0 - \frac{\pi}{4} = -\frac{\pi}{4}\,.\end{aligned}$$

In the last line we used Equation 1.13b.

– For $\omega \gg \omega_{RC}$,

$$\arg\left(H(\omega)\right) \approx \arg\left(\frac{1}{i\omega RC}\right) = -\frac{\pi}{2}\,.$$

Again, drawing a curve through the found values will give a figure similar to Figure 1.3. After marking all relevant values and axis labels, the Bode plots are complete.

Wikipedia: **`http://en.wikipedia.org/wiki/Bode_plot`**

1.4.4 Polar plot

Other than illustrating the transfer function in a Bode plot, we can also illustrate the transfer function in a polar plot:

- **Polar plot:** A plot with the imaginary value of the transfer function $H(\omega)$ plotted on the vertical axes and the real value on the horizontal axis, where ω is used as a parameter with $\omega = 0$ to $\omega \to \infty$.

Figure 1.4 is a polar plot of the RC filter from Figure 1.2 with $RC = 0.01\,\mathrm{s}$, the same as in the Bode plot example. Polar plots can also be drawn to determine if a system is stable, as with the feedback-systems in Chapter 4.

Wikipedia: **`http://en.wikipedia.org/wiki/Polar_plot`**

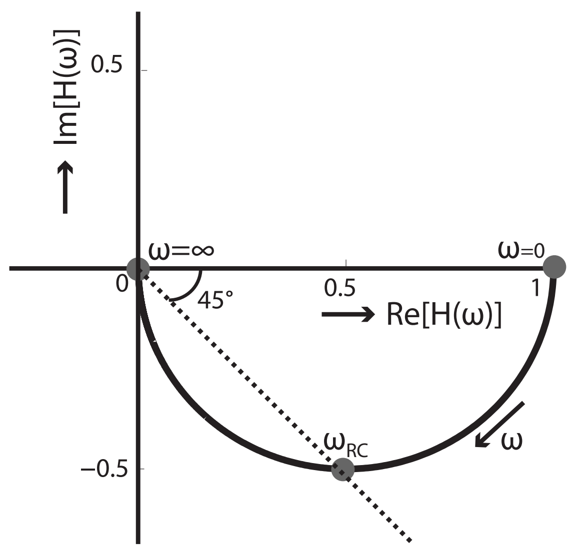


Figure 1.4: *Polar plot of the RC filter in Figure 1.2 with* $RC = 0.01\,\mathrm{s}$ *s. The imaginary value of the transfer function* $H(\omega)$ *has been plotted on the vertical axis and the real value on the horizontal axis.* ω *is used as a parameter with* $\omega = 0$ *to* $\omega \to \infty$*. You can compare this plot to the Bode plot from Figure 1.3. The radius of a polar plot is equal to the magnitude of the transfer function and the angle with the x-axis is the argument, or phase, of* $H(\omega)$*. In this polar plot, the point* $\omega = 0$ *is plotted and gives* $H(\omega) = 1$, $\omega = \omega_{RC}$ *gives* $H(\omega) = \frac{1}{1+i}$ *and* $\omega \to \infty$ *gives* $H(\omega) \to 0$.

1.5 Filters

The circuit in Figure 1.2 is called a low-pass filter. We can clearly see from Figure 1.3 that the circuit allows low-frequency signals to pass through, whereas high-frequency signals are (partially) blocked. If the resistor and capacitors are swapped around we get a high-pass filter, shown in Figure 1.6.

Filters are important when conducting experiments. Sometimes you are only interested in the amplitude of for instance the eigen-vibrations of a cantilever and want to filter out all other signals, or you want to filter out the 50 Hz noise from the mains supply but want to keep the rest of your signals. Table 1.2 gives an overview of the most common filters and shows their associated Bode magnitude plot. Notice that there is always a transition phase around the cut-off frequency ω_c. This transition gets smaller as the order of the filter increases.

Wikipedia: `http://en.wikipedia.org/wiki/Electronic_filter`

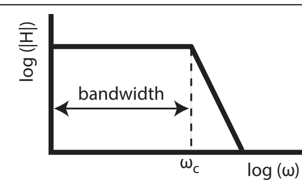
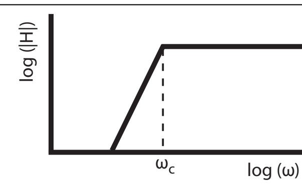
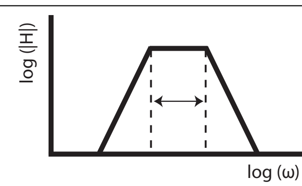
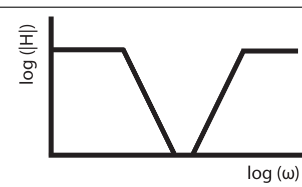
Filter	Bode magnitude plot
Low-pass filter	
High-pass filter	
Band-pass filter	
Band-stop filter	

Table 1.2: *Overview of the four main types of filters with their associated Bode magnitude plot.*

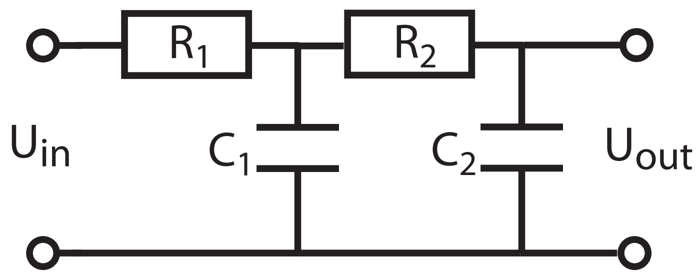


Figure 1.5: *A circuit where two first-order low-pass filters are connected to each other.*

1.5.1 Coupled filters

Figure 1.5 shows a circuit where two first-order low-pass filters are connected to each other. One would be tempted to write the transfer function of this circuit as the product of the transfer functions of the individual filters. Although this approximately holds under certain conditions, it does not hold in general:

$$H_{tot}(\omega) \neq H_1(\omega) \cdot H_2(\omega) \,. \tag{1.14}$$

Why not? When we go back to how we calculated the transfer function of the singular low-pass filter in Section 1.4.1, we made the assumption that no current could flow along the output (an ideal voltmeter). In the coupled case however, a current is able to flow from the output of the first low-pass filter. This can significantly change the output signal of the first system and with that also the total transfer function. Therefore, only when the current that flows through the second filter is so small that it does not influence the output signal of the first filter, are we allowed to just multiply the individual transfer functions to obtain the total transfer function. In audio engineering this is called impedance bridging[5]: the load impedance (the second filter) is much larger than the source impedance. The source impedance can be seen as the impedance in series with the voltage the second filter is measuring[6]. In Figure 1.5 the load impedance equals $R_2 + Z_{C_2}$ and the source impedance is Z_{C_1}. The second filter will draw little current when $R_2 + Z_{C_2} \gg Z_{C_1}$. This can often be achieved by proper design, as discussed in the example. Another way is to use an active component called an OpAmp (using a buffer amplifier, see Chapter 5).

Example: Coupled low-pass filters

- **Question:** Consider the circuit in Figure 1.5. We would like to make a second order low pass filter, so we choose both filters to have the same cut-

[5] Note that impedance *matching* is when the load draws maximum power from the source.

[6] See for more information on equivalent circuits `https://en.wikipedia.org/wiki/Thevenins_theorem`

off frequency. Show that if $R_2 \gg R_1$, we can write the transfer function for the entire circuit $H_{tot}(\omega)$ as the product of the individual transfer functions $H_1(\omega)$ and $H_2(\omega)$.

- **Answer:**

$$H_1 = \frac{1}{1 + i\omega R_1 C_1} = \frac{Z_{C_1}}{Z_{C_1} + R_1} \,,$$
$$H_2 = \frac{1}{1 + i\omega R_2 C_2} = \frac{Z_{C_2}}{Z_{C_2} + R_2} \,.$$

We use the properties of impedances (Section 1.3) to calculate the transfer function. We use the following steps and begin with the voltage across the resistor R_1:

$$U_{R_1} = U_{in} \frac{R_1}{R_1 + \frac{1}{\frac{1}{Z_{C_1}} + \frac{1}{R_2 + Z_{C_2}}}} \,,$$
$$U_{out} = U_{in} \left(1 - \frac{R_1}{R_1 + \frac{1}{\frac{1}{Z_{C_1}} + \frac{1}{R_2 + Z_{C_2}}}} \right) \cdot \frac{Z_{C_2}}{R_2 + Z_{C_2}} \,,$$
$$H_{tot} = \left(1 - \frac{R_1}{R_1 + \frac{1}{\frac{1}{Z_{C_1}} + \frac{1}{R_2 + Z_{C_2}}}} \right) \cdot \frac{Z_{C_2}}{R_2 + Z_{C_2}} \,.$$

This is the total transfer function of the system for the general case. We have for both filters the same cut-off frequency, so $R_1 C_2 = R_2 C_2$, and from $R_2 \gg R_1$ we can derive that for all possible values of ω (check the limits $\omega \to 0$ and $\omega \to \infty$ yourself) the equality $R_2 + Z_{C_2} \gg Z_{C_1}$ holds. We can then simplify last expression:

$$H_{tot} \approx \left(1 - \frac{R_1}{R_1 + Z_{C_1}} \right) \cdot \frac{Z_{C_2}}{R_2 + Z_{C_2}}$$
$$= \frac{Z_{C_1}}{Z_{C_1} + R_1} \cdot \frac{Z_{C_2}}{Z_{C_2} + R_2} = H_1 \cdot H_2 \,.$$

And with this, we can see that the total transfer function may be written as the product of the transfer functions of the individual filters.

Wikipedia: **`https://en.wikipedia.org/wiki/Impedance_bridging`**

1.6 Impulse and step response functions

The transfer function directly shows what the output signal of a system looks like with an input signal of the form $e^{i\omega t}$. This is called the frequency response of the system. Sometimes it can be useful to use other response functions. Here we will discuss the impulse response and the step response function. These response functions are mathematically related to the frequency response. Moreover, they contain the same information, though in a different form.

1.6.1 Impulse response

The impulse response $h(t)$ is the output signal of a system when the input signal is a delta function $x(t) = \delta(t)$. The impulse response completely characterizes the behavior of a system. To see this one should realize that any input signal $x(t)$ can be written as the integral over infinitely many shifted delta functions:

$$x(t) = \int_{-\infty}^{\infty} x(\tau)\delta(t-\tau)\,\mathrm{d}\tau\,. \tag{1.15}$$

Since the output signal for a single delta function is given by the impulse response, the output signal for the input signal $x(t)$ will be the integral over infinitely many shifted versions of the impulse response:

$$y(t) = \int_{-\infty}^{\infty} x(\tau)h(t-\tau)\,\mathrm{d}\tau = x(t) \otimes h(t)\,, \tag{1.16}$$

where we used the convolution operator $\otimes$. The convolution of two functions $u(x)$ and $v(x)$ is defined as:

$$u(x) \otimes v(x) = \int_{-\infty}^{\infty} u(\tau)v(t-\tau)\,\mathrm{d}\tau\,. \tag{1.17}$$

The impulse response $h(t)$ is related to the transfer function $H(\omega)$ via a Fourier transform:

$$H(\omega) = \int_{-\infty}^{\infty} h(t)e^{-i\omega t}\,\mathrm{d}t\,. \tag{1.18}$$

You might have noticed the similarities between Equations 1.4a and 1.16, which gives the relation between the input and output signal in the frequency and time domain respectively. In Section 2.1 we will further explore this relation between the time and the frequency domain.

1.6.2 Step response function

The step response function is the output signal of a system when the input signal is a Heaviside step function.[7] Just like the frequency response and the impulse response, the step response also completely characterizes the behavior of a system. Again, to convince yourself of this fact you could rewrite any input signal as an integral over shifted step functions.

Just as the Dirac delta function is the derivative of the Heaviside step function, the impulse response $h(t)$ is related to step response function $s(t)$ by a

[7] The Heaviside step function is defined as $\theta(t) = 0$ for $t < 0$ and $\theta(t) = 1$ for $t \geq 0$.

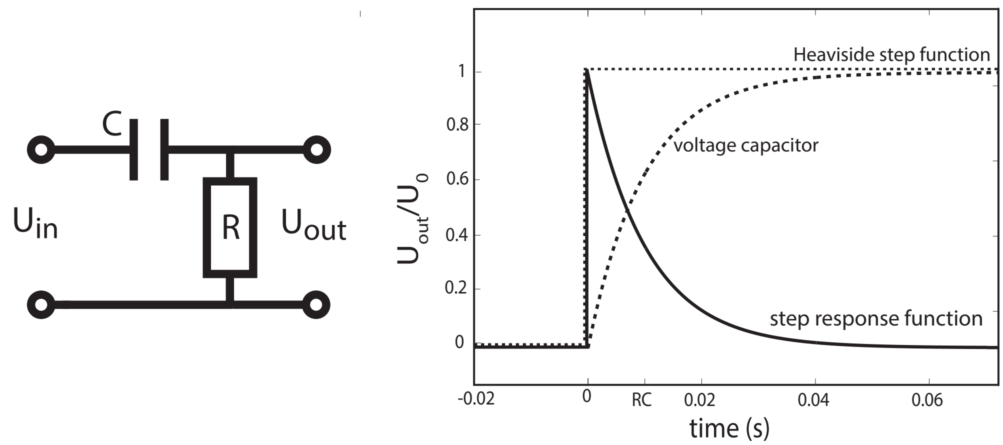


Figure 1.6: *On the left is an example of a high-pass filter. When U_{in} increases at $t = 0$ with a step from 0 to U_0, the voltage across the capacitor will increase exponentially to a new value with an exponential transient. On the other hand U_{out} will decrease exponentially to eventually become 0. This function is the step response function of the filter, shown on the right with a solid line with $RC = 0.01$ seconds, see Equation 1.20.*

derivative.

$$h(t) = \frac{\mathrm{d}s(t)}{\mathrm{d}t}\,. \tag{1.19a}$$

For the Fourier transform of the step function $S(\omega)$ we find:

$$H(\omega) = i\omega S(\omega)\,. \tag{1.19b}$$

It depends on the experiment which of the discussed response functions is most useful. As shown above, once you have found one of these response functions, the others can be easily calculated.

Example step response function of a high-pass filter

- **Question:** Determine the step response function of the high-pass filter in Figure 1.6.

- **Answer:** To find the step response function we first examine the differential equation. The current $i(t)$ through the resistor R must be equal to the current through the capacitor C, from this we find that the differential equation of the system is given by:

$$i(t) = \frac{U_{in}(t) - U_C(t)}{R} = C\frac{\mathrm{d}U_C(t)}{\mathrm{d}t}\,.$$

If we assume an input voltage $U_{in}(t) = 0$ for $t < 0$ and $U_{in}(t) = U_0$ for $t > 0$ we can solve the differential equation by making the substitution $U_{out}(t) = U_{in}(t) - U_C(t)$, and find that:

$$U_{out}(t) = U_0 e^{-\frac{t}{RC}} . \tag{1.20}$$

From this we know what the function $U_{out}(t)$ looks like when the input $U_{in}(t)$ is a step function.

Wikipedia: `http://en.wikipedia.org/wiki/Step_response`

1.7 System Equivalence

To conclude this chapter, we would like to stress that many different systems in physics are described by the exact same differential equations. As a consequence, this system equivalence can be used to solve, for example, the response of a mass-spring system with the same methods as the current to voltage response of a LC-resonator. By inspecting the differential equation describing the voltages in an LC-resonator and comparing it to the differential equation describing the position of the mass in a mass-spring system it becomes clear that the current in the first equation is equivalent to the velocity of the mass, the capacitor is equivalent to the spring etc.

Wikipedia provides a list of many equivalent systems and their parameters at `https://en.wikipedia.org/wiki/System_equivalence` which includes mechanical, electrical, thermal and fluid systems. For this reason, a transfer function doesn't always have units involving electrical voltages or currents. If one calculates the *motion* of mass spring system due to an applied *force* at a particular frequency, the transfer function $H(\omega) = \frac{X(\omega)}{F(\omega)}$ has units $[m/N]$.

Further reading

For more information on the mathematics behind the Fourier Transformations (see also Chapter 2) I recommend reading Chapter 12 from the mathematics book *Mathematical Methods for Physicists* [1]. You can also find more information on how to solve certain differential equations in this book.

The book by Karu [2] covers this chapter and Chapter 2 in its entirety and delves deeper in to Fourier Transformations, namely the more general Laplace Transformations.

An affordable and comprehensive book about electronic circuits, filters, Bode plots etc. is the book by Regtien [3].

Chapter 2

Signal processing

In physical experiments, we measure signals for a certain amount of time after which the signal is cut off, this is called windowing. Moreover, the measuring of physical phenomena takes place at discrete time intervals through sampling. This is necessary when the measurements have to be digitized before they can be processed by a computer.

However, sampling and windowing will cause artifacts in the information contained within the signal. For instance, sampling causes high-frequency changes between two sample points to not be recorded. Before we explore artifacts further in Section 2.3 we introduce a powerful mathematical tool that will simplify the interpretation of the processes that cause these artifacts, namely convolution.

2.1 Convolution

We have already encountered the convolution operator in Section 1.6, where we also already saw hints of an important connection between this convolution operator and the Fourier Transform. Here we will formalize that relation.

- **Convolution:** Let $u(x)$ and $v(x)$ be two functions. The convolution of $u(x)$ and $v(x)$, $y(x)$, is defined as:

$$y(x) = \int_{-\infty}^{\infty} u(s)v(x-s)\,\mathrm{d}s \stackrel{\text{def}}{=} u(x) \otimes v(x)\,. \tag{2.1}$$

  The convolution operation is commutative, $u \otimes v = v \otimes u$; distributive, $u \otimes (v_1 + v_2) = u \otimes v_1 + u \otimes v_2$; and associative, $u \otimes (v \otimes w) = (u \otimes v) \otimes w$.

- **Convolution theorem:** The Fourier Transform $Y(\omega)$ of the product of two signals $x_1(t)$ and $x_2(t)$ is equal to the convolution of their individual

Fourier Transforms:

$$y(t) = x_1(t)x_2(t)\,, \tag{2.2a}$$

$$Y(\omega) = \frac{1}{2\pi} X_1(\omega) \otimes X_2(\omega)\,. \tag{2.2b}$$

Note that the factor $1/2\pi$ depends on the definition (Equation 1.3b) of the Fourier Transform. We have a similar relationship for the reverse case: The convolution of functions in the time domain is given by the Fourier Transform of the product of the Fourier Transforms of the individual functions:

$$y(t) = x_1(t) \otimes x_2(t)\,, \tag{2.3a}$$

$$Y(\omega) = X_1(\omega)X_2(\omega)\,. \tag{2.3b}$$

- **Proof of the convolution theorem:** Using the Fourier Transform (Equation 1.3b):

$$\begin{aligned}
Y(\omega) &= \frac{1}{2\pi} X_1(\omega) \otimes X_2(\omega) \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} X_1(s) X_2(\omega - s)\, \mathrm{d}s \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x_1(t) e^{-ist}\, \mathrm{d}t \right) \left(\int_{-\infty}^{\infty} x_2(\tilde{t}) e^{-i(\omega - s)\tilde{t}}\, \mathrm{d}\tilde{t} \right) \mathrm{d}s \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1(t) x_2(\tilde{t}) e^{-i\omega\tilde{t}} \left(\int_{-\infty}^{\infty} e^{-is(t-\tilde{t})}\, \mathrm{d}s \right) \mathrm{d}t\mathrm{d}\tilde{t} \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1(t) x_2(\tilde{t}) e^{-i\omega\tilde{t}} \delta(t - \tilde{t})\, \mathrm{d}t\mathrm{d}\tilde{t} \\
&= \int_{-\infty}^{\infty} x_1(t) x_2(t) e^{-i\omega t}\, \mathrm{d}t\,.
\end{aligned}$$

And we do indeed see that this is the Fourier Transform of $y(t) = x_1(t)x_2(t)$.[1] The proof of the reverse case is obtained in a similar fashion.

An important example of the convolution theorem is the relation between input and output signals. In the time domain this relation is given by Equation 1.16 with $h(t)$ the impulse response function:

$$y(t) = h(t) \otimes x(t)\,, \tag{2.4a}$$

while in the frequency domain it is given by Equation 1.4a with $H(\omega)$ the transfer function and Fourier Transform of $h(t)$:

$$Y(\omega) = H(\omega)X(\omega)\,. \tag{2.4b}$$

[1] In the second from last line of the derivation we have used that $\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ix(a-b)}\,\mathrm{d}x = \delta(a-b)$.

The convolution of two signals is generally not very easy to see. In Karu's book [2] there is an explanation of how, by moving and mirroring the signals, you can determine the overlap of the signals and by extension the convolution, we will however not be looking into this any further. There is, however, one simple convolution that we will use and that is the convolution of a signal with the delta function, or a row of delta functions (think of a comb-like signal).

We will use the convolution with delta functions in both the time and frequency domain, as shown in Figures 2.2, 2.4 and 2.5. Why the convolution with a delta function is so important will be shown in the first example. In the second example we show how we can use convolution to explain how an (AM radio) signal is modulated.

Example convolution 1: Product of cosines

To make direct use of the previous proof we look at the product of two cosines with frequency ω_1 and ω_2:

$$\begin{aligned} x_1(t) &= \cos(\omega_1 t), \\ x_2(t) &= \cos(\omega_2 t), \\ y(t) = x_1(t)x_2(t) &= \cos(\omega_1 t)\cos(\omega_2 t) \\ &= \frac{\cos((\omega_1+\omega_2)t) + \cos((\omega_1-\omega_2)t)}{2}. \end{aligned}$$

The Fourier Transform of $y(t)$ is given by two delta functions per cosine term:[2]

$$\begin{aligned} Y(\omega) = \frac{\pi}{2}(\delta(\omega+\omega_1+\omega_2) + \delta(\omega-\omega_1-\omega_2) \\ + \delta(\omega+\omega_1-\omega_2) + \delta(\omega-\omega_1+\omega_2)). \quad (2.5) \end{aligned}$$

We can see the result in Figure 2.1. Now we would like to see if this result

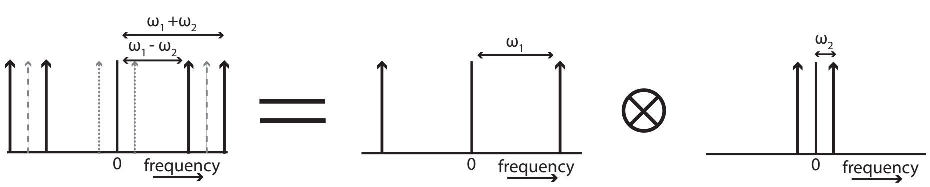


Figure 2.1: *On the left we have the Fourier Transform of the product of two cosines. We can use convolution to quickly determine the Fourier Transform of the product of two signals.*

[2]The Fourier Transform of $\cos(\omega_0 t)$ is $\pi(\delta(\omega+\omega_0)+\delta(\omega-\omega_0))$

could have been obtained using convolution:

$$\begin{aligned}
Y(\omega) &= \frac{1}{2\pi} X_1(\omega) \otimes X_2(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X_1(\tilde{\omega}) X_2(\omega - \tilde{\omega})\, \mathrm{d}\tilde{\omega} \\
&= \frac{\pi}{2} \int_{-\infty}^{\infty} \left(\delta(\tilde{\omega} + \omega_1) + \delta(\tilde{\omega} - \omega_1)\right) \left(\delta(\omega - \tilde{\omega} - \omega_2) + \delta(\omega - \tilde{\omega} + \omega_1)\right) \mathrm{d}\tilde{\omega}\,.
\end{aligned}$$

Now use the fact that when integrating the product of two delta functions you can substitute the value of $\tilde{\omega}$ that makes the argument of the second delta function 0, into the first delta function:

$$\int_{-\infty}^{\infty} \delta(\tilde{\omega} + \omega_1)\delta(\omega - \tilde{\omega} - \omega_2)\, \mathrm{d}\tilde{\omega} = \delta(\omega + \omega_1 - \omega_2)\,.$$

When we use this we get:

$$\begin{aligned}
Y(\omega) = \frac{\pi}{2}(\delta(\omega + \omega_1 + \omega_2) &+ \delta(\omega - \omega_1 - \omega_2) \\
&+ \delta(\omega + \omega_1 - \omega_2) + \delta(\omega - \omega_1 + \omega_2))\,. \qquad (2.6)
\end{aligned}$$

And we see that this is indeed equal to Equation 2.5. This example for the cosines extends to other convolutions with delta functions as we shall see in the next example.

Example convolution 2: Modulation of radio signals

- **Question:** Explain what the frequency spectrum looks like after amplitude modulation with a signal of 100 kHz is applied to a radio signal with a bandwidth of 5 kHz by using the convolution theorem.

- **Answer:** Amplitude modulation (AM) is a simple form of modulation wherein an original data signal $x(t)$ is multiplied with a carrier signal, a cosine with frequency f_d (angular frequency $\omega_d = 2\pi f_d$) and an amplitude A, to create a new signal $u(t)$. Multiplication in the time domain is a convolution of the signals in the frequency domain. A single cosine with a frequency f consists of two delta functions at $-f$ and f in the frequency domain. This gives:

$$u(t) = x(t) A \cos(\omega_d t)\,, \qquad (2.7a)$$

$$U(\omega) = \frac{1}{2\pi} X(\omega) \otimes A\left(\pi(\delta(\omega - \omega_d) + \delta(\omega + \omega_d))\right)\,. \qquad (2.7b)$$

For f_d = 100 kHz and $X(\omega)$ only having non-zero values between -5 kHz and 5 kHz the result is given in Figure 2.2. By using a different frequency of carrier wave for every radio station we can receive multiple radio stations because they will be distinguishable from each other in the frequency domain. The difference between the highest and lowest absolute value of the frequency of the signal is called the bandwidth.

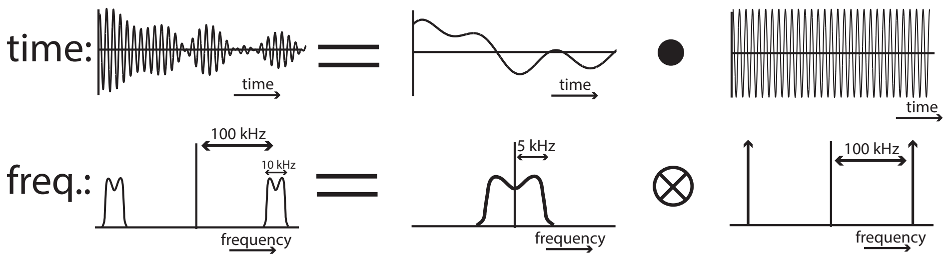


Figure 2.2: *Example of how, by using convolution, the modulation of a (radio) signal may be understood. Above is the direct multiplication of the signal with a rapidly oscillating cosine. Below is the (absolute value of) convolution of the Fourier Transforms of both signals. The amplitude of the Fourier Transform of a real signal is symmetric. We see that the frequency spectrum of the signal is shifted in the frequency domain. This modulation is used to allow multiple radio stations to be transmitted without interference occurring, low frequency signals can also only be picked up by large antenna.*

Amplitude modulation has a number of pros and cons when compared to other modulation techniques such as frequency modulation (FM). In FM the frequency of the carrier wave is modulated instead of the amplitude as in AM. One disadvantage of AM is that it is very sensitive to interference, an advantage is that multiple signals can be listened to simultaneously, while doing this with FM would result in a whistling tone. Furthermore FM works with a larger bandwidth, therefore fewer signals fit in a certain frequency range.

Wikipedia: **`http://en.wikipedia.org/wiki/Convolution`**

2.2 Fourier Transforms

> *The Fourier Transformation is a method to display information in a different way in order to gain a physical insight and facilitate calculations. Hence, it is important to obtain an intuition for the frequency domain.*

A Fourier Transform is a method to display a signal in a different way in order to gain a deeper physical or mathematical insight. The reason that the Fourier Transform is so successful at this can be understood from Chapter 1, where we saw that the complex exponentials are eigenfunctions of many electrical systems. As a result, the system can be fully characterized by the transfer function $H(\omega)$.

Sampling of a signal in order to manipulate it on a computer makes the signal discrete. It also makes the signal finite because you measure for a certain time period T. In this chapter we will examine what effect these two properties have on signal processing by looking at, among others, their effect on the Fourier Transform.

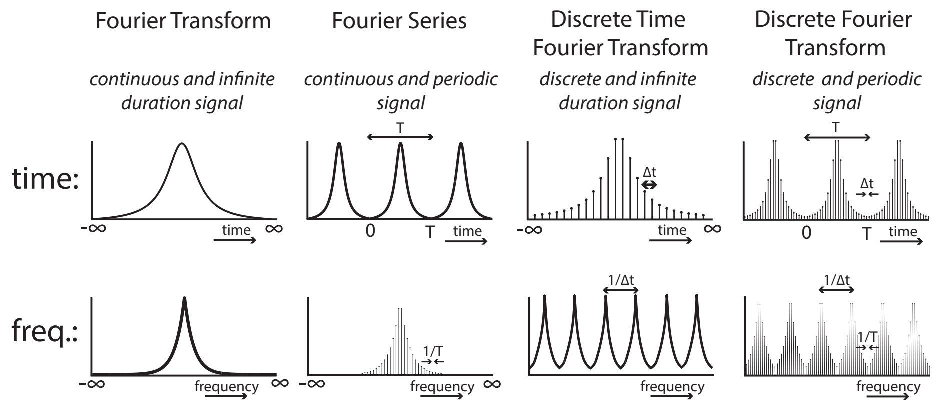


Figure 2.3: *Overview of the four forms of the Fourier transform. Above the signal is given in the time domain, the modulus of the (complex) Fourier coefficients in the frequency domain are given below.*

The Fourier Transform (FT) that we came across in Chapter 1 represent the most general form of a Fourier transformation. Nevertheless, the four different forms of a signal; discrete or continuous, periodic or non-periodic, all have their own specific Fourier transform. For example, with discrete signals we only have to use the finite number of points for a transformation rather than treating each point like a delta function within a continuous Fourier Transform.

The four Fourier transforms have their own name, and due to their special properties may be defined differently from the standard Fourier Transform that we have seen in Chapter 1. We will see that by making use of the Convolution theorem all the different Fourier transforms may be understood to be special cases of the standard Fourier Transform. In this reader we refer to a Fourier transform (without a capital T) as a general term that encompasses all four different transforms.

2.2.1 Overview of the Fourier transforms

Figure 2.3 gives an overview of the four Fourier transforms, we will briefly examine and give the definition of each of the transforms.[3] The time interval Δt and the period T of a signal are visible in this figure.

- **Time interval Δt:** The amount of time in seconds between two data points of a measurement.

- **Period T:** The total length in seconds of the measurement.

[3] The definitions sometimes differ from other sources in the pre-factors used, pluses and minuses, and whether angular frequency ω or frequency f is used.

I: Fourier Transform (FT)

$x(t)$ is continuous and aperiodic

All the information in a signal $x(t)$ is transferred, by means of the Fourier Transform, to a continuous function $X(\omega)$ in the frequency domain:

$$X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-i\omega t}\,\mathrm{d}t\,, \tag{2.8a}$$

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)e^{i\omega t}\,\mathrm{d}\omega\,. \tag{2.8b}$$

An important property of all Fourier transforms is that the sum of the square of the function in both domains has to be equal to each other, this is Parseval's theorem:

- **Parseval's theorem:** For the Fourier Transform (Equation 2.8b) the following applies for every signal $x(t)$:

$$\int_{-\infty}^{\infty} |x(t)|^2\,\mathrm{d}t = \frac{1}{2\pi}\int_{-\infty}^{\infty} |X(\omega)|^2\,\mathrm{d}\omega = \int_{-\infty}^{\infty} |X(f)|^2\,\mathrm{d}f\,. \tag{2.9}$$

- **Proof of Parseval's theorem:** We use the expression for the inverse Fourier Transform and the expression for a delta function as used earlier to prove Equation 2.1:

$$\begin{aligned}\frac{1}{2\pi}\int_{-\infty}^{\infty} |X(\omega)|^2\,\mathrm{d}\omega &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{d}t\int_{-\infty}^{\infty}\mathrm{d}\tilde{t}\int_{-\infty}^{\infty} x^*(t)x(\tilde{t})e^{i\omega(t-\tilde{t})}\,\mathrm{d}\omega\\ &= \int_{-\infty}^{\infty} x^*(t)\,\mathrm{d}t\int_{-\infty}^{\infty} x(\tilde{t})\delta(t-\tilde{t})\,\mathrm{d}\tilde{t}\\ &= \int_{-\infty}^{\infty} |x(t)|^2\,\mathrm{d}t\,.\end{aligned}$$

This theorem is useful when calculating the contribution of noise to your signal by only looking at the frequency spectrum of the noise. This theorem will mainly be used in Chapter 3.

Wikipedia: `http://en.wikipedia.org/wiki/Fourier_transform`

II: Fourier Series (FS)

$x(t)$ is continuous and periodic/finite

When a signal is periodic we see that the Fourier Transform of the signal only has non-zero Fourier coefficient at discrete values of $\omega_k = k \cdot \frac{2\pi}{T}$, with T the period of the signal. We can understand this by using the convolution from the previous section, see Figure 2.4.

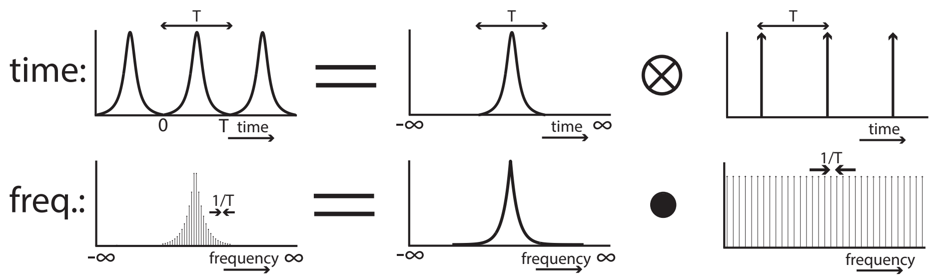


Figure 2.4: *The Fourier Series can be explained using the Fourier Transform and convolution. Above you can see how a periodic signal can be constructed by the convolution of a single period with a series of delta functions. The Fourier Transform is equal to the direct product of the individual Fourier transforms. We see that the frequency resolution is equal to* $1/T$

.

The Fourier Series $X[k]$, with k an integer, of a periodic signal $x(t)$ with period T is given by:

$$X[k] = \frac{1}{T}\int_0^T x(t)e^{\frac{-i2\pi kt}{T}}\,\mathrm{d}t\,, \tag{2.10a}$$

$$x(t) = \sum_{k=-\infty}^{\infty} X[k]e^{\frac{i2\pi kt}{T}}\,. \tag{2.10b}$$

We only have to consider the information of a single period when transforming the signal. This means that also a signal that has only been measured for a certain period T can be transformed to the Fourier domain using a Fourier Series. However, upon performing the inverse Fourier Series it becomes clear that the signal, which was previously only defined between $t = 0$ and $t = T$, is made continuous by repeating the signal over its measured period.

Still, a computer works with not only a discrete frequency spectrum but also with discretised time. The Fourier transforms for these signals will be covered in the next two sections where we look at the Discrete Fourier Transform and the Discrete Time Fourier Transform.

Wikipedia: `http://en.wikipedia.org/wiki/Fourier_series`

III: Discrete Time Fourier Transform (DTFT)

$x(t)$ is discrete and aperiodic

Suppose we make a discrete set of measurements with a time interval Δt between subsequent data points. Instead of a continuous signal $x(t)$ we are now dealing with a sampled signal $x_n = x[n\Delta t]$, a list of numbers wherein each number belongs to a specific discrete value of time $n\Delta t$. The DTFT is given by:

$$X(\tilde{\omega}) = \sum_{n=-\infty}^{\infty} x_n e^{-i\tilde{\omega}n}\,, \tag{2.11a}$$

$$x_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} X(\tilde{\omega})e^{i\tilde{\omega}n}\,\mathrm{d}\tilde{\omega}\,. \tag{2.11b}$$

Where $\tilde{\omega} = \omega\Delta t$ is the normalized (dimensionless) angular frequency with ω the physical frequency.

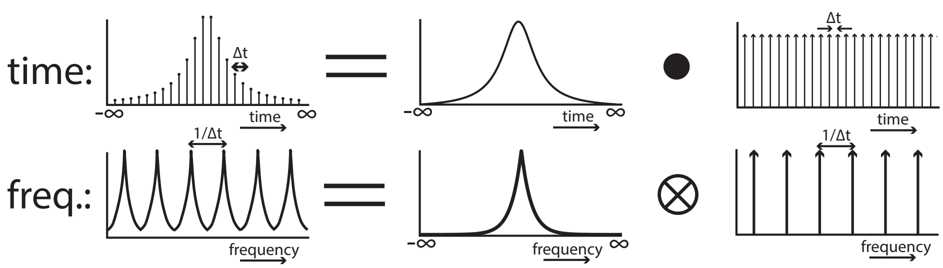


Figure 2.5: *The Discrete Time Fourier Transform causes a periodicity in the Fourier Transform. Because the Fourier Transform of the continuous signal can be broader in the frequency than the periodicity of the sampling signal, overlapping of the signals can occur. This causes Aliasing, see Section 2.3.1.*

Once again this form of transformation can be understood through the standard Fourier Transform and convolution as can be seen in Figure 2.5. We see that the Fourier Transform is continuous and periodic due to the sampling of the time. When sampling we have to look out for Aliasing, the loss of information due to sampling compared to the continuous signal, as we shall see in Section 2.3.1.

Wikipedia: `http://en.wikipedia.org/wiki/Discrete-time_Fourier_transform`

IV: Discrete Fourier Transform (DFT)

$x(t)$ is discrete and periodic/finite

The Discrete Fourier Transform is the transformation most commonly used to look at measurements in the frequency domain. By continuing a signal by repeating it periodically your frequency becomes discrete. We get a combination of the Fourier Series and the Discrete Time Fourier Transform. We will use the conventional definition, where we transform a list of numbers $x[n]$ with

$n = 0, 1, ..., N-1$ into $X[k]$ met $k = 0, 1, ..., N-1$:

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-i2\pi k \frac{n}{N}} , \tag{2.12a}$$

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] e^{i2\pi k \frac{n}{N}} . \tag{2.12b}$$

You can convert $X[k]$ into $X[\omega]$ by substituting $k = \omega T/2\pi$ where T is the time associated with $x[N-1]$. This ensures that you maintain a one-to-one relationship with the Fourier Transform.

The DFT is used in the very widely used algorithm Fast Fourier Transform (FFT). This algorithm is very important because it allows for the transformation of large quantities of experimental data.[4]

Wikipedia: `http://en.wikipedia.org/wiki/Discrete_Fourier_transform`

2.2.2 Frequency resolution

When a signal is measured for a period T and we Fourier transform this signal using the Discrete Fourier Transform, we can see from Figure 2.4 that we get discrete values for the frequency. In other words, a transform with only discrete values of the frequency is sufficient to fully describe the signal. Vice versa, this means that our resolution in the frequency domain is limited if we do not measure a signal for an infinite period of time. Clearly, the latter is impossible in practice. We define the so-called frequency resolution:

- **Frequency resolution:** The DFT of a signal that is measured for a period T consists of frequencies k/T with k discrete and $1/T$ the frequency resolution.

The frequency resolution is closely related to certain artifacts, such as Aliasing and Spectral Leakage, that can occur when we measure signals.

2.3 Measuring artifacts

When a signal is digitized two undesired phenomena can occur: Aliasing and Spectral Leakage. Both phenomena may be explained by working in the frequency domain, but take note: **the artifacts are not caused by the transformations themselves. They are due to the fact that we measure discrete points for a limited period of time.**

Digitizing consists of sampling and cutting off of the signal $x(t)$ which causes information to be lost. The digitized signal can no longer tell you what the

[4] The DFT is an algorithm that requires a number of calculations that scales with N^2. The number of calculations of a FFT scales with $N \log(N)$. This means that if for a million data points the FFT took 2 seconds, the DFT would take two days.

continuous signal did between two discrete points and you also do not know what happened before or after the sampling interval. The differences between the Fourier Transform of the true signal $x(t)$ and the DFT of the digitized signal x_n are called artifacts.

2.3.1 Sampling: Aliasing

- **Aliasing:** The phenomenon where the Fourier transform of a signal x_n contains signals that were not present in the original continuous signal $x(t)$, due to a sampling frequency $f_s = 1/\Delta t$ that is too low. The information of high frequency signals on the other hand, is lost.

The phenomenon of Aliasing may be intuitively understood by looking at how two different cosines signals are sampled, see Figure 2.6. We can fit different cosine signals through the same measurement points, this means that we have no way of knowing what the original continuous signal looked like. This (negative) effect is called Aliasing.

We can gain a further understanding of Aliasing by once again making use of Convolution as can be seen in Figure 2.7. Convolution with a Dirac comb leads to copies of the original spectrum shifted by an integer number of time f_s. [5]

We call the sampling frequency f_s and the maximum frequency f_{max}.[6] We can directly see from Figure 2.7 that we do not experience Aliasing when f_s is more than double f_{max}. This is the Nyquist sampling theorem:

- **Nyquist sampling theorem:** When a continuous function contains no frequencies higher than $f_s/2$, then this function is completely determined by that same function sampled with a frequency f_s.
- **Nyquist frequency:** The Nyquist frequency f_N is equal to half the sampling frequency f_s.

In order to avoid information loss the following must be true:

$$|f_{max}| < \frac{f_s}{2} \stackrel{\text{def}}{=} f_N \,, \tag{2.13}$$

with $|f_{max}|$ the maximum absolute frequency of your signal.

Preventing Aliasing

In order to prevent a signal being interpreted incorrectly as a result of sampling it is imperative that you know that your signal contains no signals of a higher frequency than the Nyquist frequency f_N:

[5] Note that negative frequency components can be shifted to positive frequencies and the other way around. For instance, a cosine with a frequency f_r when sampled with a frequency f_s, will show peaks in the spectrum at $f = \pm f_r + \pm n f_s, n \in \mathbb{N}$

[6] The Fourier Transform of a real signal is symmetric, this causes the minimum frequency to automatically be equal to the maximum frequency.

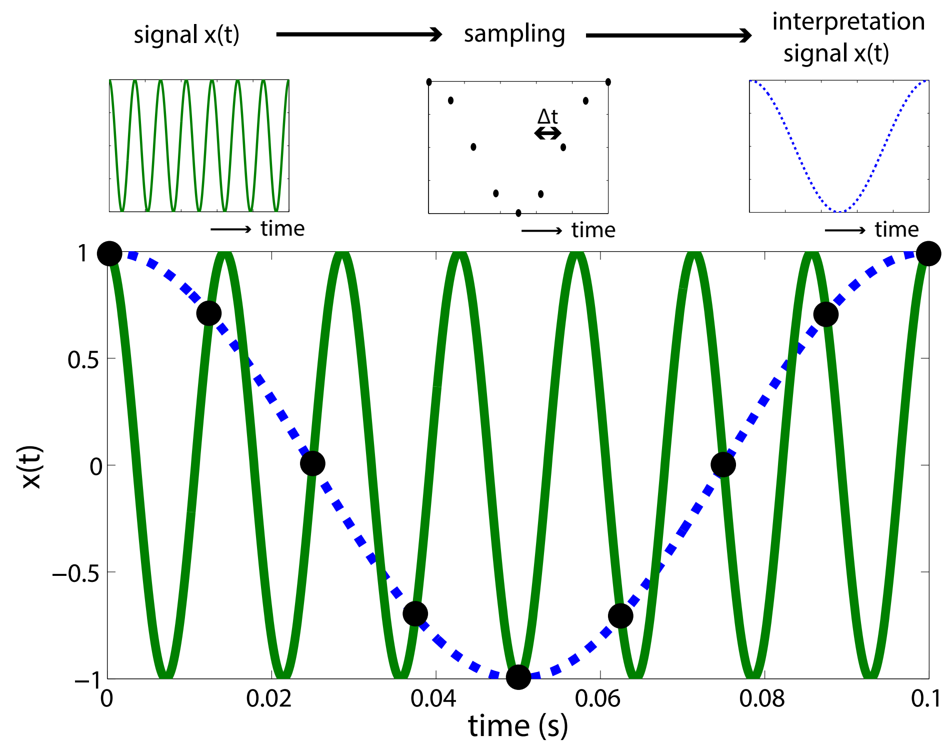


Figure 2.6: *Graphical reproduction of the phenomenon of Aliasing. A continuous signal of* 70 Hz *(solid line) is sampled with a frequency* $f_s = 80\,\mathrm{Hz}$. *We can see that a sampled signal of* 10 Hz *(dashed line) produces the exact same data points. We can thus never use our measurements to determine whether the original signal was* 10 Hz *or* 70 Hz.

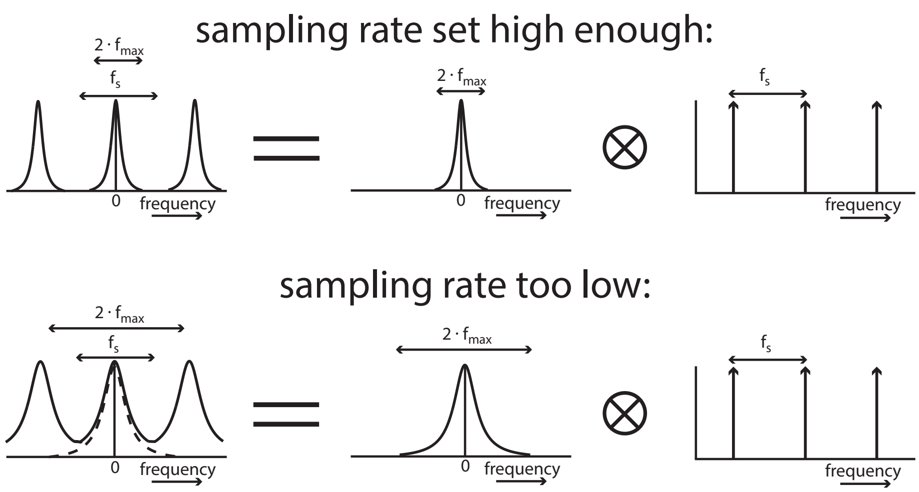


Figure 2.7: *Aliasing arises because we have sampled a signal with an insufficient sampling frequency. In the frequency domain we get a periodicity with a period equal to the sampling frequency, as can also be seen in Figure 2.5.* ***Above*** *is the situation of the convolution in Figure 2.5 in the frequency domain when a sufficiently high sampling frequency is used.* ***Below*** *is the situation when this is not the case. Notice that the low frequencies of one peak overlap with the high frequencies of another peak as a result of sampling. In the frequency domain we thus fail to obtain the desired spectrum (dashed line)*

- Use a *high enough order low-pass filter* with a bandwidth $B < f_N$ to filter your signal before you start sampling. This ensures that you adhere to Nyquist's sampling theorem.

- Make sure that you *oversample* enough. A low-pass filter is never perfect, it will still partially allow higher frequencies, $f > B$, to pass through. It is therefore wise that your sample frequency is several times higher than the cutoff frequency of the filter.[7]

Wikipedia: **`http://en.wikipedia.org/wiki/Aliasing`**

[7]The higher the order of the filter, the steeper the decline in signal pass through which increases the amount high frequency signals are blocked.

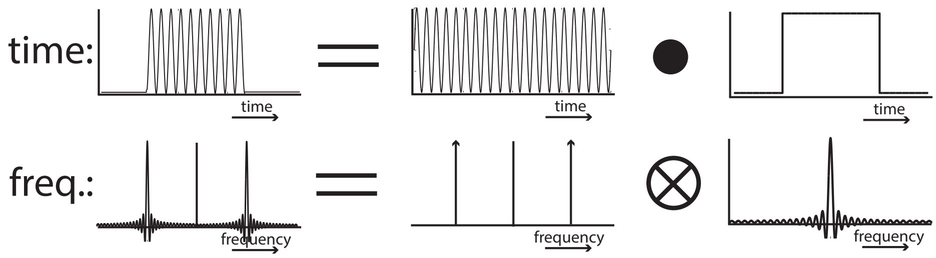


Figure 2.8: *The Fourier Transform of a measured signal will consist of the convolution of the FT of the full signal with the FT of the window. Because of this frequencies other than the ones in our original signal will appear in our spectrum, this is called Spectral Leakage. This effect also arises when working with Discrete Fourier Transforms.*

2.3.2 Windowing: Spectral Leakage

Figure 2.8 shows that a measured signal consists of the product of the original continuous signal and a rectangular window, the time that you are measuring. The Fourier Transform consists of the convolution of the Fourier Transform of your original signal and the Fourier Transform of the rectangular window. We see that as a result of this a perfect sinusoid will have non-zero values at points in the frequency domain other than the frequency of the sinusoid. This is called Spectral Leakage. The name comes from the fact that energy 'leaks' to surrounding frequencies, in Section 3.1 we will see that the square of the Fourier coefficients represents an energy.

Spectral Leakage may be reduced by measuring for a longer period of time. The width of the sinc function of the rectangular function in the Fourier domain is inversely proportional to the width of the window (i.e. the time period for which we measure). We can also choose to cut off our measurement with a different function. An example of a widely used function is called the Hanning window, as can be seen in Figure 2.9. The Hanning window is a cosine function that causes the window to have no sharp edges.

The choice of window is often a choice between good frequency resolution or good signal detection. The Fourier Transform of a broad/constant signal, as we have with a rectangular window, is a sharply peaked function in the frequency domain with large side lobes. The sharp peak gives good frequency resolution but the large side-lobes cause significant spectral leakage. The Fourier transform of a more narrow/peaked function, as we have with the Hanning window, is a broadly peaked function in the frequency domain with no (significant) side lobes. The broader peak gives a low frequency resoluti on but the lack of side lobes means that there is less spectral leakage.

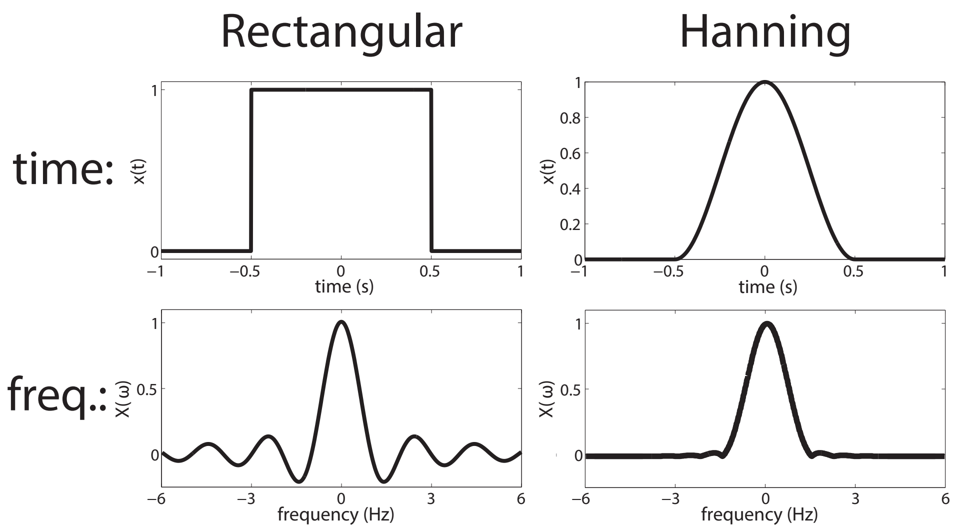


Figure 2.9: Rectangular window and a Hanning window with their Fourier transforms. The Hanning window has less prominent waves (side-lobes) in the frequency domain. The Rectangular window on the other hand has a better frequency resolution.

Spectral Leakage with the Discrete Fourier Transform

Because every measured signal is sampled and measured over a finite window the DFT is the most commonly used Fourier transform. Figure 2.8 shows how Spectral Leakage can be understood by using the Fourier Transform. DFT can cause new kinds of artifacts, as shown in Figure 2.10. When a window cuts off precisely a full period of a sine wave, the periodic continuation of our signal will make it look like the signal was never cut off at all. The DFT causes there to be no Spectral Leakage for certain frequencies, even though the FT gives every frequency the same amount of Spectral Leakage as we saw in Figure 2.8.

What is the difference between Figures 2.8 and 2.10? With a Fourier Transform you assume that your signal outside the window is equal to 0, and use this assumption when performing the Fourier Transform. You don't do this with the Fourier Series, or Discrete Fourier Transform, only the information of the measured signal is used, and when the inverse Fourier transform is performed the signal will be repeated periodically. Because the DFT is the Fourier transform that works with discrete time as well as discrete frequencies, it is the transform that we use in order to process signals in a computer.

Figure 2.10 can be derived using the convolution theorem. To obtain the figures on the left you multiply a periodic signal with a window and subsequently convolve this signal with a comb of delta functions for the periodic continuation. The Fourier Transforms (shown on the right-hand side of Figure 2.10) are then

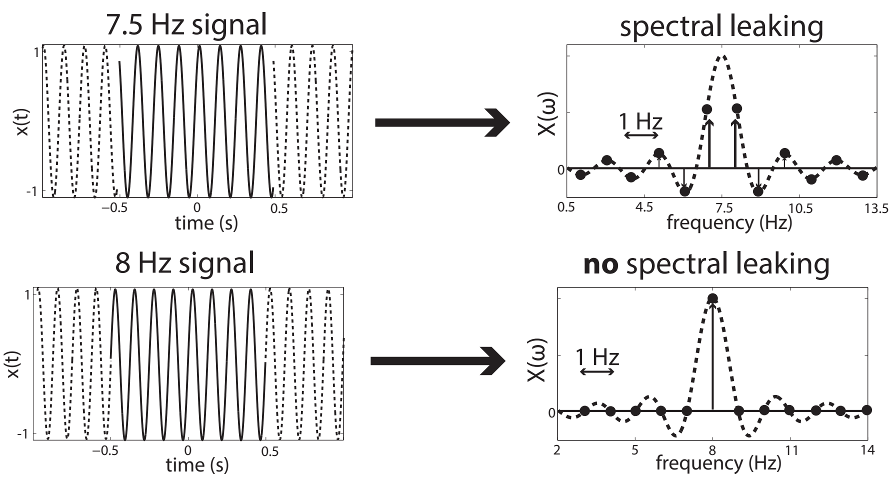


Figure 2.10: *When the window cuts of precisely a full period of your signal (bottom left), then the periodic continuation of the signal, the dotted line, will make it look as if the signal was never cut off. The discretization of the frequencies by a FS (continuous time) or by a DFT will fall together precisely with the zeros of the sinc function, as shown in the bottom right and there is no spectral leakage. In the top left we see that if the window does not precisely cut off a whole period spectral leakage does occur, as seen in the top right.*

the reverse, you first convolve two Fourier Transforms to subsequently multiply it by the Fourier Transform of a comb of delta function. It is useful to try and sketch this for yourself, because this will clarify convolution, DFT's as well as Spectral Leakage.

Wikipedia: `http://en.wikipedia.org/wiki/Window_function`

2.4 Data acquisition: Analog to Digital

Figure 2.11 shows the schematics of how a physical phenomenon is recorded on a computer. Firstly the signal is measured using a sensor, this sensor may be a device that amplifies the signal or manipulates it in some way in order to output a current or voltage. This current or voltage is still analog and must be sampled and digitized before it can be read by the computer at which point it can be processed by software such as Python or LabView. This entire process of data acquisition is also called a data acquisition system (DAS or DAQ).

We will first define the sensitivity of the sensor. This determines how much the output increases with a certain increase in the quantity we want to measure. A spring will, for example, elongate a certain amount with the application of a certain amount of force. When the change in length of the spring is measured using a laser and a photo-diode, the photo-diode will output a voltage with a sensitivity in the units V/m.

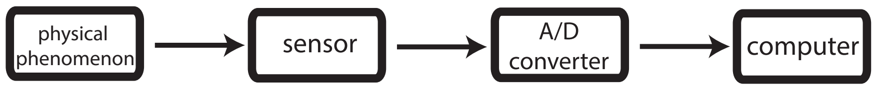


Figure 2.11: *A schematic representation of a data acquisition system (DAQ). A sensor converts a physical phenomenon into a signal that is then after being digitized by an analog-to-digital converter (A/D converter) sent to software such as LabView or Matlab on a computer.*

- **Sensitivity:** The relationship between the amount of output of a sensor/system and the amount of input.

It is often assumed that the sensitivity of a sensor is linear at every point, for instance the spring constant, but it may be that this is not the case. To avoid potential errors, you can perform a calibration whereby you map every output of the sensor to the correct input without the assumption of a constant sensitivity, or you can limit the operating range to the linear domain of the sensor.

We will now look at the process of digitizing an analog signal. We will take as an example an analog-to-digital converter that receives an input in Volts and converts this to a 12 bit digital signal. An analog-to-digital converter (A/D converter) must have a defined range of voltages that it is able to convert. Let's assume that in our example the input voltage is between 0 V and 10 V. The 12 bits allow for $2^{12} = 4096$ different possible output values (counts), but we want to reserve the highest one for signals of 10 V and above. As a result we must split our interval between 0 V and 10 V in to 4095 parts, which gives us a resolution of 2.442 mV. An input of 2 mV will give an output of 0, an input of 2.5 mV will give an output of 1.

- **Resolution of an analog-to-digital converter:** The resolution R of an A/D converter with N bits for an input range of V_{min} to V_{max} is given by:

$$R = \frac{V_{max} - V_{min}}{2^N - 1}\,. \tag{2.14}$$

- **Quantization errors:** Due to the finite number of bits the voltage range of the analog input is split into intervals equal to the resolution R. This causes quantization errors that come about due to the conversion of an analog signal into a digital signal.

- **Conversion errors:** Errors may arise during the conversion that come from, for example, noise from the electronic components of the A/D converter. The bit-noise of an A/D converter is usually greater than your smallest bit.

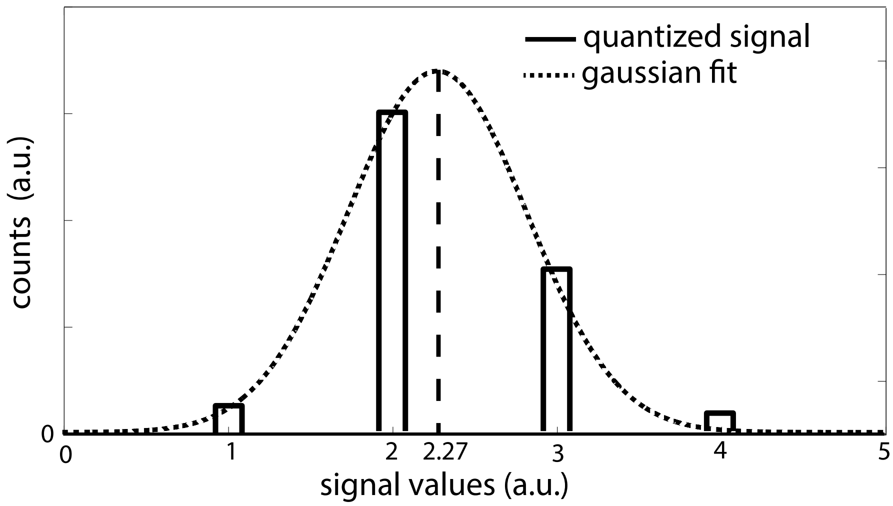


Figure 2.12: *You can use the bit-noise of an A/D converter to determine the analog values of your signal, which were rounded to discrete values by the A/D converter. The bit-noise causes a Gaussian distribution about the true value, by taking the mean value we can ascertain the true value of the signal. Do take note that you must over-sample, you have to collect data faster than the true value of the measurement changes.*

Minimizing quantization errors: Using bit-noise

The quantization errors that come about from using a finite number of bits can be decreased by using an A/D converter with more bits or by using a smaller voltage range. It is possible that even after taking these steps the quantization errors are too significant in your measurements. In that case you can use the bit-noise that is in your signal for a process called dithering. The bit-noise of an A/D converter is the noise that gets added to your signal during the conversion of the analog signal. The bit-noise of an A/D converter follows a Gaussian distribution.

Take for example a signal of which the digitized value should be 2.27, rounded to discrete bits your A/D converter would output a value of 2. By adding bit-noise the output value will most commonly still be 2 but may sometimes be 3 or even 1, though the former is more likely because it is closer to the true value of 2.27. Figure 2.12 shows how you can then use your measured values to ascertain the true value of 2.27: the mean value of your measured points gives you the true value of your measurement.

Some A/D converters generate so little noise that there is even a special function on the A/D converter to give it extra bit-noise so that the above method may be used. It is also vital that the signal is sufficiently oversampled because you want to ensure that enough data points are collected of a single measurement so you can take a good average.

Wikipedia: `http://en.wikipedia.org/wiki/Analog_digital_conversion`

Further reading

The theory in this chapter is covered by the same books as mentioned in Chapter 1.

Once again I refer to the affordable and comprehensive book by Regtien [3].

The book by Karu [2] is an easily readable book that uses a lot of graphics to explain the concepts surrounding convolution, Fourier transforms and measuring artifacts.

Mathematical Methods for Physicists [1] is a comprehensive mathematics book that covers the Fourier Transform and the Fourier Series.

Chapter 3

Noise

- **Noise:** A stochastic variation of your signal
- **Interference:** an undesired signal, but not random.

Experimental physicists will always have to deal with undesired signals in their measurements. They may be caused by a door slamming shut in the lab, interference from the mains electricity, or thermal fluctuations at 300 mK. We don't refer to the first two as noise but interference. Fluctuations that are random, and when measured over a long period of time average to 0, are considered noise.

Even though the mean value of the noise signal is 0, the fluctuations can still be characterized by an amplitude and frequency. There is statistical information hidden in the noise which can be well defined in the frequency domain. Because noise is stochastic we need to look at the Fourier transform of the average noise, to do this we will make use of the Wiener-Khinchin theorem. We will then focus on different types of noise, such as Shot noise and thermal noise (Johnson-Nyquist noise). Finally we will cover a few ways in which you, as an experimental physicist, can reduce noise.

3.1 Characterizing noise

We examine a pure noise signal $x(t)$ and calculate the standard deviation σ of the noise:

$$\sigma^2 = \langle x^2(t) \rangle - \langle x(t) \rangle^2 = \langle x^2(t) \rangle = \lim_{T \to \infty} \left(\frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} |x(t)|^2 \, \mathrm{d}t \right) . \tag{3.1}$$

We used that $\langle x(t) \rangle = 0$, the mean value of the noise is zero. The limit of infinite time is taken because we want to calculate the mean amplitude of the noise and we have centered the signal such that $t = 0$ is in the center of the measuring time T.

We can measure a signal $x(t)$ for a period T and measure σ^2. However, this value will not give you some useful information about the noise. You could be dealing with a very rapidly fluctuating signal, or rather a very slowly changing one. By characterising the different noise sources we can learn how filters can be utilised to remove noise from our measurement.

To characterize noise we will introduce the autocorrelation function in the time domain and the noise spectrum in the frequency domain. In this chapter we will show how both functions can be related to each other using the Wiener-Khinchin theorem.

The characterization of noise can be used to calculate what the margin of error of a measurement will be, and to predict whether or not a very weak signal will be detectable through the noise.

3.1.1 The noise spectrum

To characterize noise in the frequency domain we introduce the noise spectrum, the average noise as a function of the frequency. We would like to use directly the Fourier Transform instead of the Fourier Series, so we limit the signal $x(t)$ to be non-zero on the domain T so that we can remove the bounds in the integral and we can use Parseval's theorem (Equation 2.9) to relate the standard deviation with the Fourier Transform $X(f)$:

$$\sigma^2 = \lim_{T\to\infty}\left(\frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} |x(t)|^2 dt\right) = \lim_{T\to\infty}\left(\frac{1}{T}\int_{-\infty}^{\infty} |x(t)|^2 dt\right) \tag{3.2}$$

$$\overset{\text{Parseval}}{=} \lim_{T\to\infty}\left(\frac{1}{T}\int_{-\infty}^{\infty} |X(f)|^2 df\right) \tag{3.3}$$

The signal squared integrated over time is seen as the total 'energy' of a signal. For continuous signals this total energy is not bounded, so we would like to use a total power which explains the factor $\frac{1}{T}$ in above expression. We introduce the **Noise Spectral Density** $\tilde{S}(f)$ that equals the average noise power for a given frequency f:

$$\tilde{S}(f) \overset{\text{def}}{=} \lim_{T\to\infty} \frac{1}{T}|X(f)|^2 \,. \tag{3.4}$$

Using this definition we obtain:

$$\sigma^2 = \int_{-\infty}^{\infty} \tilde{S}(f)\,\mathrm{d}f \tag{3.5}$$

This expression for σ^2 is an important result, once we know the noise spectrum $S(f)$ of a type of noise we can calculate the amplitude of the noise signal σ.

Since the Fourier transform of a real signal is an even function, so we can simplify the integral to an integral over only positive frequencies:

$$\sigma^2 = 2\int_0^{\infty} \tilde{S}(f)\,\mathrm{d}f = \int_0^{\infty} S(f)\,\mathrm{d}f \,, \tag{3.6}$$

where we have defined the ***One-Sided*** **Noise Spectral Density**

$$S(f) \stackrel{\text{def}}{=} 2 \cdot \tilde{S}(f) \text{ for } f > 0\,. \tag{3.7}$$

Wikipedia: **`http://en.wikipedia.org/wiki/Power_spectral_density`**

3.1.2 The auto-correlation function and the Wiener-Khinchin theorem

The meaning of the noise spectrum density can also be clarified by using the inverse Fourier Transform of $\tilde{S}(f)$: the autocorrelation function $R_x(\tau)$.

The autocorrelation function $R_x(\tau)$, also written as $R_{xx}(\tau)$, is another useful way of characterizing noise because it describes the correlation between the signal at time t and time $t+\tau$. The auto-correlation function is a specific form of the general correlation function $R_{xy} = \langle x(t)y(t+\tau)\rangle$.

$$R_x(\tau) \stackrel{\text{def}}{=} \langle x(t)x(t+\tau)\rangle = \lim_{T\to\infty} \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} x(t)x(t+\tau)\,\mathrm{d}t\,. \tag{3.8}$$

We have so far assumed that the signal $x(t)$ is a real signal, it doesn't have to be but for us it always is.

The noise spectrum $\tilde{S}(f)$ and the autocorrelation function $R_x(\tau)$ are Fourier-coupled to each other through the Wiener-Khinchin theorem:

- **Wiener-Khinchin theorem:** The autocorrelation function $R_x(\tau)$ and the noise spectrum $\tilde{S}(f)$ are Fourier Coupled:

$$\tilde{S}(f) = \int_{-\infty}^{\infty} R_x(\tau)e^{-i2\pi f\tau}\,\mathrm{d}\tau\,, \tag{3.9a}$$

$$R_x(\tau) = \int_{-\infty}^{\infty} \tilde{S}(f)e^{i2\pi f\tau}\,\mathrm{d}f\,. \tag{3.9b}$$

  For a proof refer to appendix 3.B.

The autocorrelation can be used to directly find the standard deviation σ, by setting $\tau = 0$ in equations 3.8 and 3.9b we find:

$$\sigma^2 =< x(t)^2 >= R_x(0) = \int_{-\infty}^{\infty} \tilde{S}(f)\,\mathrm{d}f\,. \tag{3.10}$$

Wikipedia: **`http://en.wikipedia.org/wiki/Wiener-Khinchin_theorem`**

3.1.3 Noise transfer

When noise is present at the input of a circuit that noise will look different at the output of the circuit. In Chapter 1 we defined the transfer function

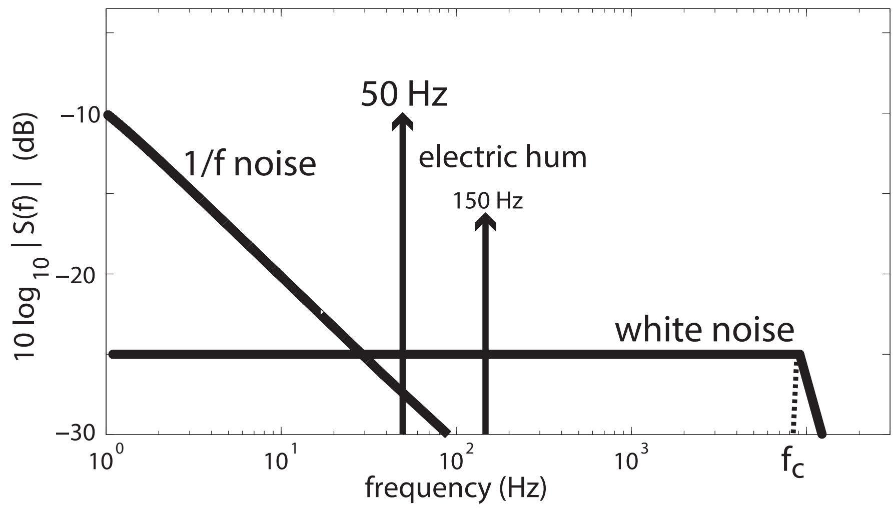


Figure 3.1: *Noise spectra with the absolute value of the noise spectrum $S(f)$ in* dB *on the vertical axis and the logarithm of the frequency on the horizontal axis. In this figure we have plotted the $1/f$ noise (pink noise) and white noise. Examples of white noise are thermal noise and shot noise. White noise has a constant value up to a certain cutoff frequency f_c. Also shown is the power spectrum of the mains electricity, the peaks at* 50 Hz, 150 Hz, 250 Hz, ... Hz *are correlated to each other. They are shown as arrows because they approach delta functions. The higher the frequency resolution with which you measure, the higher the peaks will be.*

$H(\omega)$ of a circuit. Let the input signal $X(\omega)$ be a noise signal and let $Y(\omega)$ be the noise signal at the output of the circuit. We can use our definition of the transfer function from Chapter 1 and Equation 3.4, which shows that the noise spectrum is equal to the square of the average of the absolute value of the Fourier Transform of a noise signal, to examine what happens to the noise signal as it passes through the circuit:

$$Y(\omega) = H(\omega)X(\omega)$$
$$|Y(\omega)|^2 = |H(\omega)|^2|X(\omega)|^2\,, \tag{3.11a}$$
$$S_{out}(\omega) = |H(\omega)|^2 S_{in}(\omega)\,. \tag{3.11b}$$

This result shows that we can use the filters from Chapter 1.5 to reduce noise. More ways to reduce the impact with noise are covered in Chapter 3.4.

3.1.4 Different types of noise

Noise comes in many different forms. The most common ones are thermal noise, shot noise and $1/f$ noise. The different forms of noise spectra are often characterized with colors. Thermal noise and shot noise will for instance have a white noise spectrum and $1/f$ noise is referred to as pink noise. Figure 3.1 shows these two spectra as well as typical inference from the mains electricity.

The physical mechanisms behind $1/f$ noise, also known as flicker noise for electronic systems, differ. Fluctuations at extremely low frequencies often arise as a result of drift, your system could for example be subject to temperature fluctuations. For slightly higher frequencies the $1/f$ noise can be caused by impurities in the conduction channels or by active components such as transistors in the circuit. The effects of $1/f$ noise at higher frequencies are overwhelmed by the white noise present in the system. We will now delve deeper into the mechanisms behind thermal noise and shot noise.

Wikipedia: **`http://en.wikipedia.org/wiki/Flicker_noise`**

Example: Calculating noise from the noise spectrum

- **Question:** Given a pink noise spectral density $S(f) = 1\,\mu V^2/f$, calculate the total noise in µV when we measure between a frequency of 1 Hz and 10 kHz.

- **Answer:** When we talk about the total noise we mean the standard deviation. Using Equation 3.5:

$$\sigma^2 = \int_0^\infty S(f)\,\mathrm{d}f = \int_1^{10^4} \frac{1\,\mu V^2}{f}\,\mathrm{d}f = 1\,\mu V^2 \left(\ln\left(\frac{10^4}{1}\right)\right) \approx 9\,\mu V^2$$

$$\sigma \approx 3\,\mu V\,.$$

We have hereby calculated the total noise, as a function of the standard deviation.

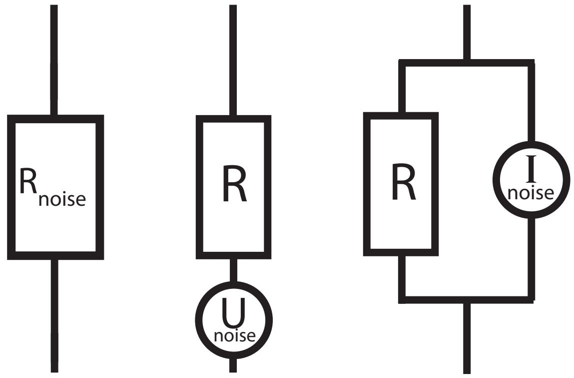


Figure 3.2: *We can replace a noisy resistor with an ideal resistor in series with a noisy voltage source, which is referred to as the Thevenin equivalent circuit, or place the ideal resistor parallel to a current source that supplies noise, this is referred to as the Norton equivalent circuit.*

3.2 Thermal noise

A standard resistor causes noise in the signal. To analyze the behavior of a non-ideal real-world resistor and calculate the noise spectrum, we can replace a noisy resistor in two ways by idealized electrical components. The first is that the noisy resistor is replaced by an ideal resistor placed in series to an ideal voltage source[1] that supplies the noise (referred to as the Thevenin equivalent circuit). The second way is that we replace the noisy resistor with an ideal resistor placed parallel to an ideal current source supplying the noise (referred to as the Norton equivalent circuit). Both ways have been sketched in Figure 3.2. The noise generated by a resistor is called thermal noise or Johnson noise, characterized by the noise spectrum for the voltage $S_V(f)$ or for the current $S_I(f)$. The name thermal noise can be understood when we see that the noise is directly proportional to the temperature of the resistor.

- **Thermal noise:** Thermal noise is constant for every frequency (white noise) up to a certain cutoff frequency f_c. The noise spectrum for a resistor R is given by:

$$S_V(f) = 4k_BTR \qquad \text{for } |f| < f_c\,, \tag{3.12}$$

  where k_B is the Boltzmann constant ($\approx 1.38 \cdot 10^{-23}$ JK^{-1}) and T the temperature.

- **Thermal noise in a circuit with bandwidth Δf:** The mean standard deviation of the thermal noise as the result of a resistance R and bandwidth Δf is given by:

$$\sigma_V^2 = 4k_BTR\Delta f\,. \tag{3.13}$$

[1]By an ideal source we mean that the internal resistance of the source is zero.

From equations 3.11b and 3.12 follows that the fluctuations in the current $S_I(f)$ caused by a resistance R is given by:

$$S_I(f) = \frac{4k_B T}{R}\,. \tag{3.14}$$

There are several different derivations for the thermal noise, we have *van der Ziel's method* and *Nyquist's theorem* to name a few. They make use of standard theories from thermal and statistical physics such as the equipartition theorem, Fluctuation-Dissipation theorem and the Fermi-Dirac distribution. In appendix 3.A you can find a derivation that uses the Wiener-Khinchin theorem, it is useful to read and understand this derivation.

Wikipedia: `http://en.wikipedia.org/wiki/Thermal_noise`

3.3 Shot noise

Shot noise is a source of noise that is present when discrete particles are measured. When photons or electrons are viewed as discrete particles the detector will sometimes record a particle and sometimes not, even though a constant flux of particles arrives at the detector (see Figure 3.3).

When the particles do not have a mutual interaction and are completely randomly distributed[2], then the number of particles entering a detector follows a Poisson distribution.[3]

- **Shot noise:** White noise that is created by the discrete character of the Poisson distributed events. With experiments this occurs when electrons and photons are involved. The noise spectrum for an average electric current I_0 is given by:

$$S_I(f) = 2eI_0 \qquad \text{for } |f| < f_c\,, \tag{3.15}$$

  where e is the elementary charge and f_c the cutoff frequency.

The cutoff frequency corresponds to the bandwidth of a detector, as is explained in the derivation below.

Wikipedia: `http://en.wikipedia.org/wiki/Shot_noise`

3.3.1 Derivation shot noise

The discrete character of the Poisson distributed events gives a special autocorrelation function. Because no event is correlated with the next or previous

[2] In reality electrons have the propensity to repel each other (anti-bunched), while photons coming from a filament lamp will tend to group together (bunched). This deviation from white noise is only really noticeable if you look very carefully in very short time periods (high frequencies).

[3] The Poisson distribution is the general result for random discrete processes with an average number of events in a certain time interval. For a process with a variable X and an average of λ events the probability $P(k,\lambda)$ of k events occurring is given by $P(k,\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$. Furthermore, there is the property that $\langle X\rangle = \lambda = \sigma^2$.

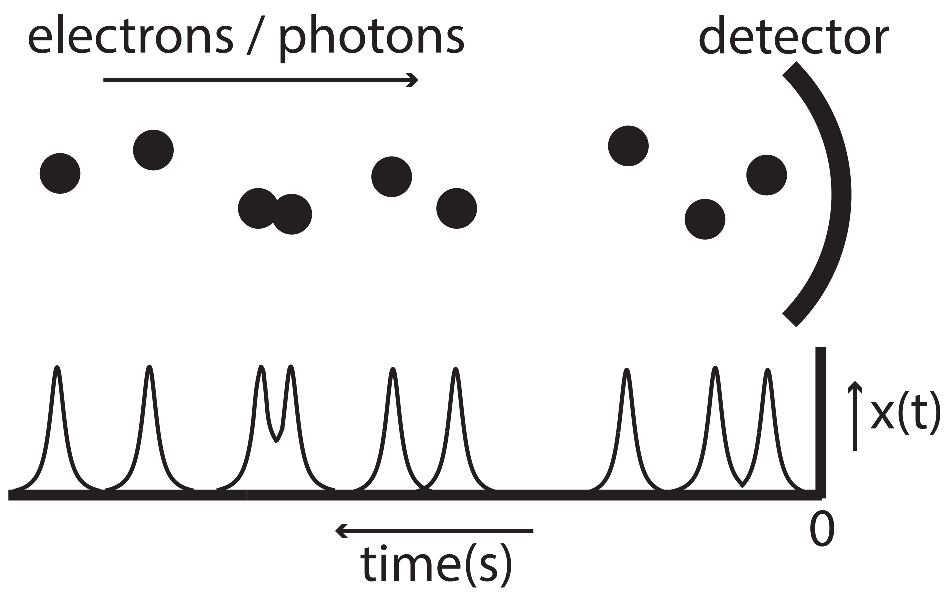


Figure 3.3: *The measuring of electrons and photons causes shot noise. Electrons and photons can be considered as point particles with a certain impact time on the detector (peak width). This width is the cutoff frequency f_c of the white noise spectrum of shot noise.*

event, the autocorrelation function of a Poisson distributed signal, such as a stream of electrons $I(t) = en(t)$ with an average I_0, is a delta function.[4]

$$R_I(\tau) = \langle I(t)I(t+\tau)\rangle = \lim_{T\to\infty} \frac{e^2}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} n(t)n(t+\tau)\,\mathrm{d}t = eI_0\delta(\tau)\,. \tag{3.16}$$

We can get the noise spectrum back by using the Wiener-Khinchin theorem (Equation 3.9b):

$$\tilde{S}_I(f) = eI_0\int_{-\infty}^{\infty} \delta(\tau)e^{-i2\pi f\tau}\,\mathrm{d}\tau = eI_0 \tag{3.17}$$

Before we continue we must realize that a physical detector takes a certain time Δt to make a measurement. We can see this as a window in the frequency domain with a width $1/\Delta t$, because of the symmetry about $f = 0$ the bandwidth

[4]The derivation is mathematically more correct when you use the expression $I(t) = e\sum_{i=1}^{\infty}\delta(t - t_i)$ with t_i the arrival times of discrete events. In that case this derivation is much easier to use in other shot noise equivalent derivations. By using the definition for the autocorrelation and properties of the Kronecker delta you end up with an expression that needs some elaboration for correlations between not correlated electrons averaged over all summations. Please contact the author if you would see a short version of this derivation and elaboration on this problem. Be aware that there are whole books written on this derivation.

is half of this, in other words $f_c = 1/2\Delta t$. Putting this in to Equation 3.15 we get the following expression for the noise spectrum:

$$S_I(f) = 2eI_0 \qquad \text{for } |f| < f_c\,.$$

3.4 Improving the signal-to-noise ratio

As we have seen in the previous sections, there are many sources of noise that can distort our desired signal. As a figure of merit, indicating how good our measurement is, we often use the signal to noise ratio (SNR).

- **Signal-to-Noise Ratio:** The signal-to-noise ration (SNR) for a signal with power P_{signal} without noise and a noise power of P_{noise} is given by:[5]

$$\text{SNR} = \frac{P_{signal}}{P_{noise}}\,. \tag{3.18}$$

Here the power is defined as:

- **Power of a signal:** For a signal $x(t)$ the power of the signal is given by $\langle x^2(t) \rangle$.

We have already come across the power of a noise signal. This is given by the square of the standard deviation σ of the noise.

$$P_{noise} = \sigma^2_{noise} = \int_0^\infty S(f)\,\mathrm{d}f\,. \tag{3.19}$$

Note that this is a mathematical definition of power that does not necessarily correspond to the physical concept of power, although they are often closely related through multiplication by a constant. For instance, the physical power dissipated in a resistor is given by $P_{phys} = I^2(t)R$, while the mathematical power of the current signal would be $P_{math} = I^2(t)$.

In order to better distinguish a signal from the noise the SNR must be as high as possible. In practice this means that you try to filter out the noise at frequencies where you do not have a signal. Doing this makes your P_{noise} as small as possible without also decreasing P_{signal}. Below is a list of a few important methods to reduce noise.

- **Using filters:** Equation 3.11b shows us that the noise spectrum $S(f)$ is multiplied with the square of the transfer function of the used filter, so the integral over the noise spectrum can be made smaller by using filters. We do have to pay attention to ensure that we do not also filter out our

[5] Sometimes the SNR as defined here is called the *Power SNR*. If that is done, the authors define the SNR without prefix as the ratio between the amplitudes of the noise and the signal. The thus obtained SNR is equal to the root of the Power SNR. In this reader by SNR we will always mean Power SNR

signal. If you know that your signal will be between 100 Hz and 1 kHz, you can make use of a band-pass filter. High-pass filters are often used to reduce the $1/f$ noise when the signal has a high frequency. Low-pass filters remove the high frequency noise sources, which smooths the signal.

- **Using averages:** Because the fluctuations are random, taking the average of a signal over a slightly longer time is an effective way of making your signal look smoother. We can also understand this from the frequency domain: In the derivation of both the thermal noise and shot noise we can see that when we measure the signal for a time period T, the cutoff frequency becomes $1/T$. The disadvantage is that many high frequency signals that you wish to measure are also averaged out. This method also does not work well for eliminating $1/f$ noise.

- **Repeated measurements:** We know from statistics that the standard deviation σ of a Gaussian distribution decreases with a factor $1/\sqrt{N}$ when you measure the signal N times. The to-be-measured signal must remain the same when repeating the measurements for this technique to work, this makes it problematic when the setup is sensitive to drift or changes in the environment.

- **Modulation of the signal:** By making use of techniques that we have seen in 2.1 a signal can be shifted in the frequency domain to an area where the noise is much weaker. This technique is useful when dealing with $1/f$ noise. This technique won't work for white noise because the noise signal is equally strong at all points. It even decreases the strength of your real signal because modulation causes your signal to not be measured occasionally (the sine of your modulation signal is zero twice every period).

- **Tackle the noise source:** Thermal noise in the current caused by a resistor can be reduced by lowering the temperature while mechanical noise of a cantilever can be addressed by decreasing the damping.

Further reading

The syllabus on noise written by Martin van Exter [4], which this reader makes use of, contains many experimental examples and information about noise and is often more comprehensive than what is handled in this reader.

A nice book on noise is the book by Wim C. van Etten [5]. This book is elaborate and uses neat formal derivation and is easily readable.

Appendix

3.A Thermal noise derived from colliding electrons

In this derivation we use the autocorrelation function $R(\tau)$ of colliding electrons to, using the Wiener–Khinchin theorem (Equation 3.9b), gain an expression for the noise spectrum $\tilde{S}(f)$

The current I as a result of a single electron with velocity v that has to cover a distance l is given by:

$$I_{el}(t) = \frac{ev(t)}{l}\,. \tag{3.20}$$

We consider an electron in equilibrium with length l, surface area A and electron density n. We must multiply our result for a single electron with the total number of electrons nAl.

$$\begin{aligned} R_I(\tau) &= nAl\langle I(t)I(t+\tau)\rangle \\ &= \lim_{T\to\infty} \frac{nAl}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} I_{el}(t)I_{el}(t+\tau)\,\mathrm{d}t \\ &= \lim_{T\to\infty} \frac{nAe^2}{Tl} \int_{-\frac{T}{2}}^{\frac{T}{2}} v(t)v(t+\tau)\,\mathrm{d}t\,. \end{aligned} \tag{3.21}$$

We now make use of what we know about the physics of an Ohmic resistor. The resistance arises because the electrons collide with impurities in the conduction channels. A collision causes the velocity of the electron to change in a random direction so that there is no correlation with the velocity before the collision. Because the number of collisions is proportional to the distance covered by the electron we have an exponential relationship (just as the amount of light let through a medium decreases exponentially with the thickness of the medium). With an average time between collisions τ_c we calculate the autocorrelation function:

$$R_I(\tau) = \lim_{T\to\infty} \frac{nAe^2}{Tl} \int_{-\frac{T}{2}}^{\frac{T}{2}} v(t)v(t+\tau)\,\mathrm{d}t = \frac{nAe^2}{l}\langle v^2\rangle e^{-\frac{|\tau|}{\tau_C}}\,. \tag{3.22}$$

Where $\langle v^2 \rangle$ is the mean of the square of the velocities of the electrons.[6] The equipartition theorem tells us that for a free particle in 1 dimension the mean kinetic energy is given by:

$$E_{kin} = \frac{1}{2} m \langle v^2 \rangle = \frac{1}{2} k_B T \,, \tag{3.23a}$$

$$\langle v^2 \rangle = \frac{k_B T}{m} \,, \tag{3.23b}$$

$$R_I(\tau) = \frac{nAe^2 k_B T}{ml} e^{-\frac{|\tau|}{\tau_c}} \,. \tag{3.23c}$$

We now use the Wiener–Khinchin relation (Equation 3.9b) to get an expression for the noise spectrum:

$$\begin{aligned} \tilde{S}_I(f) &= \int_{-\infty}^{\infty} R(\tau) e^{-i2\pi f \tau} \, \mathrm{d}\tau \\ &= \frac{nAe^2 k_B T}{ml} \int_{-\infty}^{\infty} e^{-\frac{|\tau|}{\tau_c}} e^{-i2\pi f \tau} \, \mathrm{d}\tau \,. \end{aligned} \tag{3.24}$$

We use the expression for the Fourier Transform of an exponential function which can be found in the back of the reader:

$$\tilde{S}_I(f) = \frac{nAe^2 k_B T}{ml} \left(\frac{2\tau_c}{1 + \tau_c^2 (2\pi f)^2} \right) \,, \tag{3.25a}$$

$$S_I(f) = \frac{4 k_B T}{R} \left(\frac{1}{1 + \tau_c^2 (2\pi f)^2} \right) \,. \tag{3.25b}$$

We define the cutoff frequency as $f_c \stackrel{\text{def}}{=} 1/(2\pi\tau_c)$ and make use of $R = 1/(\sigma A)$ with $\sigma = e^2 n \tau_c / m$ the conductivity. For low frequencies we get:

$$S_I(f) = \frac{4 k_B T}{R} \qquad \text{for } f < f_c \,.$$

Our result agrees with equation 3.14. We get a constant noise spectrum up to a certain cutoff frequency f_c. Because the mean collision time in a metal is in the order of nanoseconds we get a cutoff frequency in the order of THz. Most electronics circuits have an (implicit) low-pass filter with a much lower bandwidth $\Delta f \ll f_c$ so the thermal noise will be determined by the bandwidth of the circuit.

[6] The velocities are distributed following the Maxwell-Boltzmann distribution, this distribution comes up again in Statistical Physics.

3.B Derivation of the Wiener-Khinchin theorem

Wiener-Khinchin theorem: The autocorrelation function $R_x(\tau)$ and the noise spectrum $\tilde{S}(f)$ are Fourier coupled:

$$R_x(\tau) = \int_{-\infty}^{\infty} \tilde{S}(f) e^{i2\pi f\tau} \, \mathrm{d}f \,, \tag{3.26}$$

$$\tilde{S}(f) = \int_{-\infty}^{\infty} R_x(\tau) e^{-i2\pi f\tau} \, \mathrm{d}\tau \,, \tag{3.27}$$

where $\tilde{S}(f)$ and $R_x(\tau)$ are defined as:

$$\tilde{S}(f) \stackrel{\text{def}}{=} \lim_{T\to\infty} \frac{1}{T} |X[f]|^2 \,, \tag{3.28a}$$

$$R_x(\tau) \stackrel{\text{def}}{=} \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \bar{x}(t) x(t+\tau) \, \mathrm{d}t \,, \tag{3.28b}$$

with $\bar{x}(t)$ the complex conjugate of $x(t)$, for real signals they are equal to one another. We will derive equation 3.28b by start from the given definition for R_x and by taking the Fourier Transforms of both functions in the integral[7]:

$$R_x(\tau) \stackrel{\text{def}}{=} \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \bar{x}(t) x(t+\tau) \, \mathrm{d}t \,, \tag{3.29}$$

$$= \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \mathrm{d}t \left(\frac{1}{2\pi}\right)^2 \int_{-\infty}^{\infty} \bar{X}(\omega) e^{-i\omega t} d\omega \int_{-\infty}^{\infty} X(\tilde{\omega}) e^{i\tilde{\omega}(t+\tau)} d\tilde{\omega} \,, \tag{3.30}$$

$$= \left(\frac{1}{2\pi}\right)^2 \iint_{-\infty}^{\infty} \bar{X}(\omega) X(\tilde{\omega}) \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} e^{i(\omega-\tilde{\omega})t} dt. \tag{3.31}$$

We can use now the property of the delta function: $\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ix(a-b)} dx = \delta(a-b)$, so that we can remove one integral and replace $\tilde{\omega}$ with ω:

$$R_x(\tau) = \left(\frac{1}{2\pi}\right)^2 \iint_{-\infty}^{\infty} d\omega d\tilde{\omega} \left(\bar{X}(\omega) X(\tilde{\omega}) e^{i\tilde{\omega}\tau}\right) \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} e^{i(\omega-\tilde{\omega})t} dt \,, \tag{3.32}$$

$$= \frac{1}{2\pi} \lim_{T\to\infty} \frac{1}{T} \iint_{-\infty}^{\infty} d\omega d\tilde{\omega} \left(\bar{X}(\omega) X(\tilde{\omega}) e^{i\tilde{\omega}\tau} \delta(\tilde{\omega}-\omega)\right) \,, \tag{3.33}$$

$$= \frac{1}{2\pi} \lim_{T\to\infty} \frac{1}{T} \int_{-\infty}^{\infty} |\bar{X}(\omega)|^2 e^{i\omega\tau} d\omega \,, \tag{3.34}$$

$$= \int_{-\infty}^{\infty} \lim_{T\to\infty} \frac{1}{T} |\bar{X}(f)|^2 e^{i2\pi f\tau} df \,, \tag{3.35}$$

$$= \int_{-\infty}^{\infty} \tilde{S}(f) e^{i2\pi f\tau} \, \mathrm{d}f \,, \tag{3.36}$$

And with that we have derived the Wiener-Khinchin theorem.

[7] We shuffle integrals and limits like a real physicist. We invite a real mathematician to do everything properly and to send this to the author.

Chapter 4

Feedback

- **Feedback:** When the output of a system is returned to the input of a system.

In this chapter we will look at a few aspects of feedback in a system. Feedback is not only found in electronic systems, it appears in many other places. A thermostat keeps the temperature of a room constant and water is pumped out of certain parts of the Netherlands in order to maintain a regular groundwater level. In physics feedback is used in measurements with *Scanning Tunneling Microscopes, Atomic Force Microscopes* and other forms of *Scanning Probe Microscopy* where a cantilever or tip has to maintain a constant height above a sample in order to not damage the sample, but also to gain an accurate map of the surface.

In a system with feedback an important criterion is that the system must be stable. This means that the output signal does not keep growing over time. For instance, you want to avoid a situation where the pumps in the Netherlands start pumping sea water into the polders. The systems in Chapter 1 consisted of resistors, capacitors and inductors. These are called passive components. These kinds of systems are always stable. It is not possible that the output signal becomes significantly larger than the input signal. Up to now we have been able to Fourier transform our input signal $x(t)$ and make use of the transfer function $H(\omega)$ to calculate the output signal $y(t)$.

What happens however, when our system is unstable? If $y(t)$ continues to grow in time without there being a (permanent) input signal, then it has no Fourier transform. In other words, we need to find a way to tell if our system is stable and, as soon as we know for sure that it is, we can once again use the transfer function $H(\omega)$ to show the relationship between the input signal and output signal.

In this chapter we will look at a way in which we can see if a system that is described by a linear time-invariant differential equation is stable or unstable. To do this we will use the Laplace Transform. This is a transform that uses the complex frequency $s = \lambda + i\omega$ instead of the frequency ω that is used in the

Fourier Transform. The Laplace Transform is the most general method for solving a linear inhomogeneous differential equation in order to get the stationary solutions. The stationary solutions are the solutions for the output signal $y(t)$ when the input signal $x(t)$ is zero. We will see that on the basis of the Laplace Transform, which gives us the transfer function $H(s)$, we can determine whether a system is stable or unstable. $H(s)$ is very closely related to $H(\omega)$.

This chapters starts by introducing the Laplace Transform. We then derive the stability criterion with the help of a few relations and steps which are summarized in Figure 4.0.1. Finally we look at what a typical transfer function of a feedback system looks like and will cover an example of a feedback system and when it is stable or unstable.

4.1 The Laplace Transform

The Laplace Transform is a transformation operation similar to the the Fourier Transform, but there are important differences. The Laplace Transform can, aside from absolutely integrable signals that the Fourier Transform can also transform, transform diverging signals to the frequency domain with complex frequency s. A second difference with the Fourier Transform is that for the Laplace Transform we assume the function in the time domain to be *causal*, which means that for $t < 0$ the function is equal to 0.

- **Laplace Transform:** A signal $x(t)$ with $t > 0$ can be transformed to $X(s)$ with $s = \lambda + i\omega$ by means of a (one-sided) Laplace Transform:

$$X(s) = \int_0^\infty x(t)e^{-st}\,\mathrm{d}t\,. \tag{4.1}$$

  $X(s)$ is only defined for values of s for which this integral has a finite value, there is said to be a Region of Convergence (ROC). The information about the ROC in the function in the Laplace domain is necessary for the inverse Laplace Transform.

The Laplace Transform is a one-to-one transform, its inverse can be written as a complex contour integral called *Mellin's Inverse Formula*:

$$x(t) = \frac{1}{2\pi i}\lim_{T\to\infty}\int_{\lambda-iT}^{\lambda+iT} X(s)e^{st}\,\mathrm{d}s\,. \tag{4.2}$$

This is a path integral over a straight line in the complex plane with real value λ, where λ is to be chosen such that the integral is evaluated in the region of convergence. We see that when the complex axis is part of the region of convergence, $\lambda = 0$, gives the inverse Fourier transform. To actually do calculations with this inverse equation, knowledge of complex analysis is required, which will not be covered in this course. In practice one typically uses search tables to find the inverse function. We will only use the Laplace Transform to examine whether a linear time-invariant system is stable or not. As soon as we

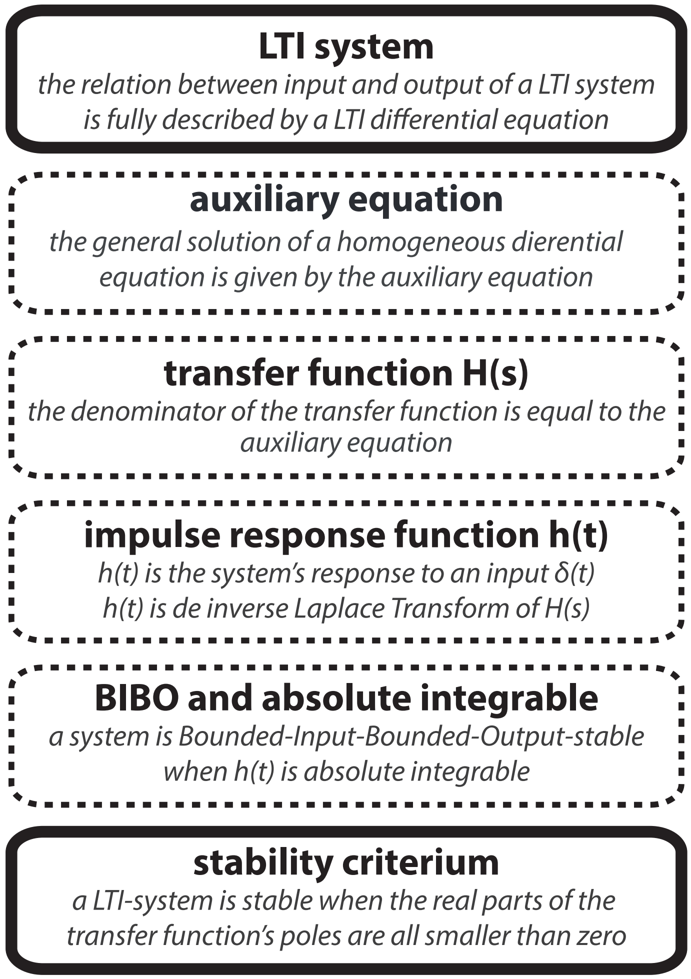


Figure 4.0.1: *This figure shows the steps that we will cover in this chapter in order to get to the stability criterion. A Linear Time-Invariant (LTI) system is described by a differential equation. The transfer function $H(s)$ contains the auxiliary equation in the denominator. The impulse response function $h(t)$ is the inverse Laplace Transform of $H(s)$. When $h(t)$ is absolutely integrable the output signal $y(t)$ is bound when the input signal $x(t)$ is also bound, we have a BIBO stable system. We use the connection between $h(t)$ and the transfer function $H(s)$ to get to our stability criterion. An LTI system is stable when the real parts of the transfer function's poles are less than zero.*

know whether a system is stable, we can use the Fourier Transform on signals to examine the effect of the system on the input signal.

Wikipedia: **`http://en.wikipedia.org/wiki/Laplace_transform`**

4.1.1 Example: The Laplace Transform

- **Question:** Show that $X(s) = 2/(s-3)$ is the Laplace Transform of $x(t) = 2e^{3t}$.

- **Answer:** We use the definition of the Laplace Transform (Equation 4.1):

$$\begin{aligned} X(s) &= \int_0^\infty x(t)e^{-st}\,\mathrm{d}t \\ &= \int_0^\infty 2e^{3t}e^{-st}\,\mathrm{d}t = \left.\frac{2e^{(3-s)t}}{3-s}\right|_0^\infty \\ &= \frac{2}{s-3} \text{ provided that } \Re(s) > 3\,. \end{aligned}$$

  The region of convergence of this function is therefore given by $\Re(s) > 3$. This information may be necessary for the inverse Laplace Transform.

4.1.2 Laplace Transform vs. Fourier Transform

While the Fourier Transform and Laplace Transform are similar, there are important differences. Here we summarize some of the most important differences and similarities.

- The Laplace Transform allows for exponentially growing functions, while the Fourier Transform does not.

- The Laplace Transform assumes the function to be transformed to be 0 for $t < 0$, while the Fourier Transform does not.

- Just as the transfer function in the frequency domain $H(\omega)$ is the Fourier Transform of the impulse response $h(t)$, the transfer function in the Laplace s-domain $H(s)$ is the Laplace Transform of the impulse response $h(t)$.

- The convolution theorem holds for both Fourier and Laplace Transforms.

- Derivation and integration with respect to time in the time domain correspond to multiplication and division by $i\omega$ in the frequency domain and to multiplication and division by s in the s-domain.

- If you want to determine the transfer function in the Fourier or Laplace domain, then we must use the corresponding impedances for the passive components listed in table 4.1.1.

Component	**Impedance** $Z(\omega)$	**Impedance** $Z(s)$
Resistor	R	R
Capacitor	$\frac{1}{i\omega C}$	$\frac{1}{sC}$
Inductor	$i\omega L$	sL

Table 4.1.1: *Summary of the impedances of a resistor, capacitor and inductor in the frequency and s-domain.*

- In general the Fourier transform is typically used to analyze the response of a system to quasi-everlasting periodic input signals, while the Laplace transform is used to determine the stability of a system and to analyze the transient behavior when an input signal is turned on or off.

4.2 Stability of a system

A linear system is stable when the output signal is bound for every (bound) input signal. We call this the Bounded-Input-Bounded-Output principle, which is also called BIBO stability.

- **BIBO stability:** We have a Bounded-Input-Bounded-Output (BIBO) stable system when the output signal of the system is bound for every bound input signal.

Figure 4.0.1 shows how we can define a stability criterion that allows us to determine whether a system is stable or not from the differential equations that describe the system. We will ultimately see that the transfer function of the system gives us enough information to allow us to determine the stability of the system. Below we will show that a system can be shown to be BIBO stable if it obeys the stability criterion:

- **Stability criterion:** A system is BIBO stable when the real part of every pole s_i of the transfer function $H(s)$ is less than zero, in other words $\Re(s_i) < 0$.

We will arrive at this stability criterion in three steps. Firstly, in part 1 we look at the transfer function and its relationship to the differential equations. In part 2 we look at the impulse response function and its relationship to the transfer function and the BIBO stability. Part 1 and 2 then come together in part 3 where we formulate the stability criterion.

4.2.1 Part 1: The transfer function

The auxiliary equation

At the start of Chapter 1 we looked at the most general expression for a linear time-invariant system with output signal $y(t)$ and input signal $x(t)$:

$$a_0 x + a_1 \dot{x} + a_2 \ddot{x} + ... + a_n x^{(n)} = b_0 y + b_1 \dot{y} + b_2 \ddot{y} + ... + b_m y^{(m)} \,. \tag{4.3}$$

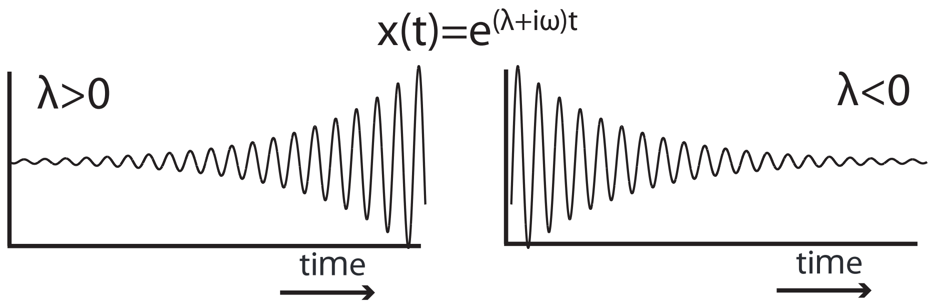


Figure 4.2.2: *Because a feedback system can amplify (left) or dampen (right) a signal we use the Laplace Transform to characterize a system. The Laplace Transform is used to find the general solution, characterized by $H(s)$, of a linear time-invariant differential equation. The image on the left is an example of a signal that gets amplified. When the signal is created spontaneously we are dealing with an unstable system.*

When the input signal is equal to 0 we get the following homogeneous linear differential equation:

$$0 = b_0 y + b_1 \dot{y} + b_2 \ddot{y} + ... + b_m y^{(m)} \,. \tag{4.4}$$

The stationary solutions for this differential equation can be found by using the Laplace Transform, which comes down to substituting in e^{st} and simplifying the expression until we are left with the auxiliary equation:

$$0 = b_0 + b_1 s + b_2 s^2 + ... + b_m s^m \,. \tag{4.5}$$

The solutions, s_i, of this equation give the most general solution for $y(t)$, namely $y(t) = A_1 e^{s_1 t} + A_2 e^{s_2 t} + ... A_m e^{s_m t}$.[1]

Each of the terms in the general solution as the form $e^{s_i t} = e^{\lambda_i t} e^{i\omega_i t}$ where λ and ω real numbers. When $\lambda > 0$ we have an exponentially increasing function, when $\lambda < 0$ we have an exponentially decreasing function, see Figure 4.2.2.

Relationship between the auxiliary equation and the transfer function

In Equation 1.2 we introduced the transfer function:

$$H(s) = \frac{a_0 + a_1 s + a_2 s^2 + ... + a_n s^n}{b_0 + b_1 s + b_2 s^2 + ... + b_m s^m} \,, \tag{4.6a}$$

$$Y(s) = H(s) X(s) \,. \tag{4.6b}$$

[1]We have assumed that all solutions, s_i, are different, when this is not the case we have to multiply the exponential solution by s_i^n with n an integer related to the degeneracy of s_i. We are mainly interested at this point in whether or not the system is stable, and for the instability the exponential function is more important than the polynomial.

Remember that the transfer function in the Fourier domain $H(\omega)$ is equal to the transfer function in the Laplace domain when you substitute $i\omega = s$. Do take note however, that this does not have to be the case for the transform of a function in general.

We assume that $n < m$, this is always the case for physically realizable systems. The poles of the transfer function are given by the equation $b_0 + b_1 s + b_2 s^2 + ... + b_m s^m = 0$, in other words the poles of the transfer function are equal to the solutions of the auxiliary equation of the homogeneous linear differential equation.

4.2.2 Part 2: The impulse response function and absolute integrability

In Section 1.6 we have seen that the impulse response function $h(t)$ is the response of a system to a delta function $\delta(t)$ as an input signal. It is also the inverse Laplace transform of the transfer function $H(s)$. We can use the impulse response function to determine whether or not a system is BIBO stable. We will show that the output signal is bound for every possible bound input signal provided that the impulse response function is absolutely integrable.

- **Assertion:** The output signal $y(t)$ is bound, $|y(t)| < \infty$, provided that the impulse response function $h(t)$ is absolutely integrable and the input signal is bound, that is $|x(t)| < x_{max}$. Absolutely integrable means that:

$$\int_0^\infty |h(t)| \,\mathrm{d}t < \infty \,. \tag{4.7}$$

- **Derivation:** We can write the output signal $y(t)$ as the convolution of the impulse response function $h(t)$ and the input signal $x(t)$ by using the convolution theorem. We use that $|x(t)| < x_{max}$ and $\int_0^\infty |h(t)| \,\mathrm{d}t < \infty$.

$$\begin{aligned} Y(s) &= H(s)X(s)\,, \\ y(t) &= h(t) \otimes x(t) = \int_0^\infty h(\tau)x(t-\tau)\,\mathrm{d}\tau \\ &\leq \int_0^\infty |h(\tau)||x(t-\tau)|\,\mathrm{d}\tau \\ &< x_{max} \int_0^\infty |h(\tau)|\,\mathrm{d}\tau < \infty\,. \end{aligned}$$

By now using the Laplace relationship between the impulse response function and the transfer function we can test the BIBO stability of our system by using only our knowledge of the transfer function of the system. This will be our general stability criterion.

4.2.3 Part 3: The stability criterion

We can combine the different steps to get a single stability criterion that only makes use of the transfer function $H(s)$.

We begin by rewriting the transfer function $H(s)$ by applying partial fraction decomposition with the poles of $H(s)$ given by s_i and the residues α_i, note that we have chosen the denominator to have a coefficient of 1 in the leading order.[2]:

$$\begin{aligned} H(s) &= \frac{a_0 + a_1 s + a_2 s^2 + ... + a_n s^n}{(s - s_1)(s - s_2)...(s - s_m)} \\ &= \frac{\alpha_1}{s - s_1} + \frac{\alpha_2}{s - s_2} + ... + \frac{\alpha_m}{s - s_m} \,. \end{aligned} \tag{4.8}$$

The inverse Laplace Transform of $H(s)$ is given by the impulse response function $h(t)$ which we can calculate using the Laplace Transform example:

$$h(t) = \left(\alpha_1 e^{s_1 t} + \alpha_2 e^{s_2 t} + ...\alpha_m e^{s_m t}\right) \cdot \theta(t) \,, \tag{4.9}$$

where $\theta(t)$ is the Heaviside step function. For the absolute integrability of an exponential function the following applies:

$$\int_0^\infty |e^{st}| \, \mathrm{d}t = \begin{cases} -\frac{1}{\Re(s)}, & \Re(s) < 0 \\ \infty, & \Re(s) \geq 0 \,. \end{cases} \tag{4.10}$$

The system is BIBO stable if $h(t)$ is absolutely integrable, because then, as we have seen, every possible output signal is also absolutely integrable. $h(t)$ is absolutely integrable when each of the exponential functions in Equation 4.9 is absolutely integrable. In other words, the real parts of all poles s_i must be less than zero. In this we have a powerful tool to determine the stability of the system, we will call this the stability criterion.

- **Stability criterion:** A system is BIBO stable when the real part of every pole s_i of the transfer function $H(s)$ is less than zero, in other words $\Re(s_i) < 0$.

Vice versa, if a system is unstable there must therefore exist a stationary solution that increases in time. We can use the pole that causes the instability to gain some knowledge about the frequency at which the system shows instability and the characteristic time at which the instability grows.

- **Properties of an unstable system:** For a system with a single pole $s_i = \lambda_i + i\omega_i$ with $\lambda_i > 0$, ω_i gives the frequency at which the instability occurs and λ_i is the characteristic time at which the instability grows. see Figure 4.2.2.

[2]Assuming that all s_i are unique the residue is given by $\alpha_i = \lim_{s \to s_i}(s - s_i)H(s)$. Note that we also have $n < m$.

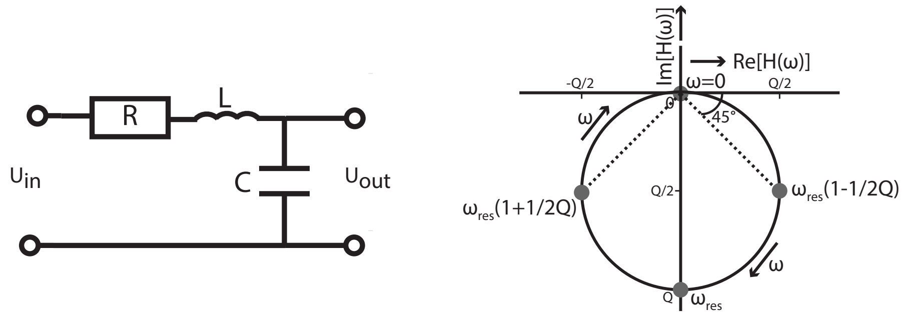


Figure 4.2.3: *On the left is a LCR circuit with a transfer function $H(s)$ given by the Equation 4.11c. This circuit is used as an analogy for a mechanical resonator. On the right the polar plot is shown for a mechanical resonator with high quality factor $Q \gg 1$. Note that at $\omega = 0$ the transfer function $H(\omega) = 1$.*

4.2.4 Example: An LCR-resonator in the Laplace domain

Figure 4.2.3 shows a series LCR circuit that is often used as an analogy for a mechanical resonator.

- **Question:** Give the transfer function $H(s)$ of this circuit. Furthermore, determine whether or not this system is stable.

- **Answer:** Using the complex impedances given in table 4.1.1:

$$U_{out}(s) = \frac{I(s)}{sC}\,, \tag{4.11a}$$

$$U_{in}(s) = I(s)\left(R + sL + \frac{1}{sC}\right)\,, \tag{4.11b}$$

$$H(s) = \frac{1}{1 + sRC + s^2LC} = \frac{1}{1 + \frac{s}{\omega_{res}Q} + \left(\frac{s}{\omega_{res}}\right)^2}\,. \tag{4.11c}$$

Where we define $Q \overset{\text{def}}{=} \frac{1}{R}\sqrt{\frac{L}{C}}$ and $\omega_{res} \overset{\text{def}}{=} \frac{1}{\sqrt{LC}}$. To determine whether or not this system is stable we look at the roots of the denominator, which are easily found using the quadratic formula:

$$0 = 1 + \frac{s}{\omega_{res}Q} + \left(\frac{s}{\omega_{res}}\right)^2$$

$$s_\pm = -\frac{\omega_{res}}{2Q} \pm \frac{i\omega_{res}}{2}\sqrt{4 - \frac{1}{Q^2}}$$

We can see that $\Re(s_i) = -\omega_{res}/(2Q) < 0$, so our system is stable.

This example shows another big advantage of examining systems in the Laplace domain rather than the Fourier domain. The poles that we found are in fact

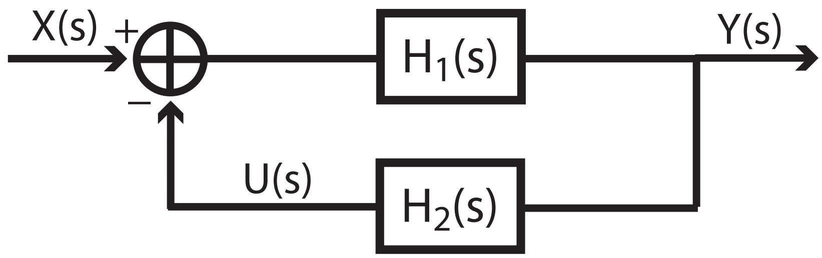


Figure 4.2.4: *A general system with negative feedback. The output signal is subtracted, via the transfer function $H_2(s)$, from the input signal $X(s)$. $U(s)$ is called the error signal.*

also the solutions to the auxiliary equation (Equation 4.5), i.e. they are the stationary solutions of the system. When you are not driving a system but for instance let go of a cantilever at amplitude A, then these will give you the solutions of your system, you only have to check the conformity to the boundary conditions. In the case of a resonator with a high quality factor, that is $Q \gg 1$ (there is little dissipation, R is very small), we see that $s_{\pm} \approx -\omega_{res}/(2Q) \pm i\omega_{res}$. In the time domain this gives the solution to the damped harmonic oscillator, namely $y(t) = Ae^{-\omega_{res}t/(2Q)} \cos(\omega_{res}t)$. We see that Q determines the damping of the system.

You could have obtained the above result in the Fourier domain too. You can use a Heaviside function as the input signal, Fourier Transform it, multiply it with $H(\omega)$ and then perform an inverse Fourier Transform. Apart from the fact that this is much more cumbersome it also provides less insight in to the system. A damped harmonic oscillator is a system that you come across frequently in nature and the damping behavior of the signal at a certain frequency is a behavior that can be given by one variable s in the Laplace domain. The Fourier domain cannot give this solution because it has a basis consisting of only infinitely ongoing sines and cosines.

4.3 Transfer function of feedback systems

Fortunately, if a system is unstable we can use feedback to make the system stable. A thin stick pointing straight up resting on your hand is a physically unstable system, but by giving feedback by actively moving your hand you can make the system stable. Moreover, feedback can also be used to regulate stable systems. The thermostat in your house is an example. Heating your house with constant power would after a while lead to a stable temperature. However, at which temperature the equilibrium would lie would depend strongly on the outside temperature and other factors. The feedback of the thermostat allows you to control the temperature in spite of varying outside temperatures.

However, feedback can also be the cause of instability. The feedback of sound from a speaker to a microphone can cause saturation of the amplifier, the all-too-familiar high pitched beep.

Figure 4.2.4 shows the general schematic notation of a feedback system. The output signal $y(t)$ is modified by a certain transfer function in to a signal $u(t)$ that is then either added to or subtracted from the input signal $x(t)$. The $+$ or $-$ in the drawing tells us whether we add $u(t)$ to the input signal (positive feedback) or subtract it (negative feedback).

- **Positive feedback:** When the feedback $u(t)$ is added to the signal $x(t)$.
- **Negative feedback:** When the feedback $u(t)$ is subtracted from the signal $x(t)$

To find the transfer function of a system with feedback, we try to find an expression for the output signal as a function of the input signal and itself. To this end we transform all the signals into their Laplace Transforms and make use of the linearity of our system. If H_2 does not load H_1 ($|Z_{in,2}| \gg |Z_{out,1}|$) we then have for the system in Figure 4.2.4 in the case of **positive feedback**:

$$\begin{aligned}
Y(s) &= H_1(s)(X(s)+U(s))\,, \\
U(s) &= H_2(s)Y(s)\,, && \text{(4.12a)} \\
Y(s) &= H_1(s)(X(s)+H_2(s)Y(s))\,. && \text{(4.12b)}
\end{aligned}$$

Rewriting gives for $H(s)$:

$$H(s) = \frac{Y(s)}{X(s)} = \frac{H_1(s)}{1-H_1(s)H_2(s)}\,. \tag{4.13}$$

For **negative feedback** we get after following the same procedure:

$$H(s) = \frac{H_1(s)}{1+H_1(s)H_2(s)}\,. \tag{4.14}$$

Now that we have the transfer function of the total system we can study the stability. The stability criterion tells us that a system with feedback is stable when the real part of the poles of the transfer function $H(s)$ is less than 0. This is equivalent to determining when the so-called loop transfer function $H_{loop}(s) = H_1(s)H_2(s)$ is equal to 1 or -1 for positive and negative feedback respectively.[3] Also, we often talk about the Open Loop transfer function $H_{open}(s)$, this is the transfer function without the feedback, in our case we would have $H_{open}(s) = H_1(s)$.

[3] The *Nyquist Stability Criterion* is a method that allows you to determine whether $\Re(s_i) < 0$ by using just a polar plot of $H_{loop}(s = i\omega)$. Because this method gives us little extra insight we will not be discussing this method in this reader.

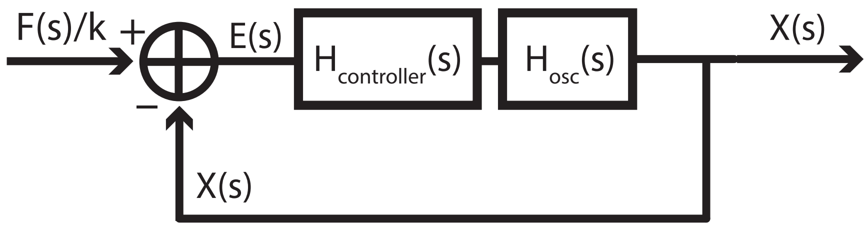


Figure 4.3.5: *A schematic drawing of a system where feedback is used to control the position of an oscillator using a force $F(s)$.*

4.3.1 Example of a feedback system

Consider a mass-spring system with transfer function $H_{osc}(s)$ which we would like to control by applying a force $F(s)$ to the system. One scheme to obtain this control is shown in Figure 4.3.5. In order to make the transfer function dimensionless we divide the input force by the stiffness k.

The controller has a transfer function $H_{controller}(s)$. In general, one takes a PID-controller, which consist of proportional, integral and derivative feedback. In order to solve the system analytically we only consider proportional control, so we have $H_{controller} = P$.

We assume that the quality factor of the oscillator $Q \gg 1$ and we only consider proportional gain $P \geq 0$.

- **Question 1:** Determine whether the system is stable or not.

- **Answer:** First we must determine the system's transfer function $H(s)$ of the system. We can then use the stability criterion to determine whether the system is stable or not.

  For the transfer function of the oscillator $H_{osc}(s)$ we use the result for the LCR resonator, Equation 4.11c:

$$H_{osc}(s) = \frac{1}{1 + \frac{s}{\omega_{res} Q} + \left(\frac{s}{\omega_{res}}\right)^2} . \tag{4.15}$$

  Using Equation 4.14 we obtain for $H(s)$:

$$\begin{aligned} H(s) &= \frac{P \cdot H_{osc}(s)}{1 + P \cdot H_{osc}(s)} \\ &= \frac{P}{1 + P + \frac{s}{\omega_{res} Q} + \left(\frac{s}{\omega_{res}}\right)^2} . \end{aligned} \tag{4.16}$$

To find the poles $s_\pm$ of $H(s)$ we use the quadratic formula:

$$\begin{aligned} 0 &= 1 + P + \frac{s}{\omega_{res} Q} + \left(\frac{s}{\omega_{res}}\right)^2 \\ s_\pm &= -\frac{\omega_{res}}{2Q} \pm \frac{\omega_{res}}{2}\sqrt{\frac{1}{Q^2} - 4(1+P)}\,. \end{aligned}$$

We see that for $P > 0$ and with $Q \gg 1$, the part inside the root is always negative, and therefore the root gives an imaginary part to the poles. We obtain for the real parts:

$$\Re\left(s_\pm\right) = -\frac{\omega_{res}}{2Q} < 0\,. \tag{4.17}$$

Since the real parts are smaller than zero, we have shown that we have a stable system.

- **Question 2:** Calculate the output signal $x(t)$ when the input signal is a delta function $\delta(t)$ for $P > 0$ and $Q \gg 1$.

- **Answer:** The function that we want to calculate is equal to the impulse response function $h(t)$. $h(t)$ is equal to the inverse Laplace Transform of the transfer function $H(s)$ or to the inverse Fourier Transform of $H(\omega)$ in the case of a stable system. We can also use the relationship between the impulse response function and the poles that we found, see Equation 4.9. The residuals $\alpha_\pm$ can be found directly by rewriting $H(s)$ using the found poles of the denominator. Be aware of the extra factor ω_{res} that is the coefficient of the leading order in the denominator.

$$\begin{aligned} \alpha_\pm &= \lim_{s\to s_\pm}(s - s_\pm)H(s) \\ &= \lim_{s\to s_\pm}(s - s_\pm)\frac{P\omega_{res}^2}{(s-s_+)(s-s_-)} \\ &= \frac{P\omega_{res}^2}{s_\pm - s_\mp} \\ &= \pm\frac{P\omega_{res}}{\sqrt{\frac{1}{Q^2} - 4(1+P)}}\,. \end{aligned}$$

We introduce the dimensionless positive real variable $C \equiv \sqrt{4(1+P) - \frac{1}{Q^2}}$. The impulse response function can now be calculated using Equation 4.9:

$$\begin{aligned} h(t) &= \left(\alpha_+ e^{s_+ t} + \alpha_- e^{s_- t}\right)\cdot\theta(t) \\ &= \frac{P\omega_{res}}{iC}\left(e^{\left(-\frac{\omega_{res}}{2Q} + i\frac{\omega_{res}C}{2}\right)t} - e^{\left(-\frac{\omega_{res}}{2Q} - i\frac{\omega_{res}C}{2}\right)t}\right)\cdot\theta(t) \\ &= \frac{2P\omega_{res}}{C}e^{-\frac{\omega_{res}}{2Q}t}\sin\left(\frac{\omega_{res}}{2}Ct\right)\cdot\theta(t)\,. \end{aligned} \tag{4.18}$$

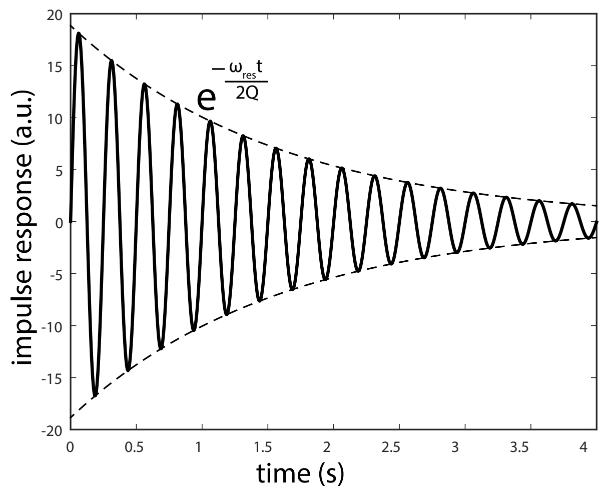


Figure 4.3.6: *The impulse response function (Equation 4.18) of the feedback system visible in Figure 4.3.5. We chose $Q = 10$, $\omega_{res} = 2\pi \cdot 2\,\mathrm{Hz}$ and $P = 3$.*

The result for $Q = 10$, $\omega_{res} = 2\pi{\cdot}2\,\mathrm{Hz}$ and $P = 3$ is plotted in Figure 4.3.6. In this plot you can clearly see the exponential decay. The proportional gain $P = 3$ has increased the resonance frequency of the system to $2\omega_{res}$.

Further reading

The book by Karu [2] is an easily readable book that uses a lot of graphics to explain the concepts surrounding feedback, stability and the Laplace Transform.

Chapter 5

OpAmp

The history of the Operational Amplifier (OpAmp) goes back nearly 100 years. The development of it arose from the need for a reliable amplifier of signals. The amplification was often not the problem, the amplification was however not constant for all frequencies, there were also problems with noise and non-linearity. By combining an amplifier with negative feedback, something that was considered strange in the early years, you can design a system that is largely frequency independent. In this chapter we will look at the golden rules that dictate how we use the OpAmp in a circuit. We will also discuss some applications of the OpAmp.

On the left in Figure 5.0.1 is an amplifier that does not utilize negative feedback. This is also called an Open-Loop (no feedback) amplifier. We have shown the OpAmp as a simple component in the diagrams but in truth, an OpAmp consists of many resistors and transistors and must be attached to a power source in order to function. This last point shows that the OpAmp is an example of an *active* component. Resistors, capacitors and inductors are examples of *passive* components, no energy is required to make these components work.

For the Open-Loop circuit the output voltage is given by:

$$U_{out} = G(U_+ - U_-)\,. \tag{5.1}$$

The amplification, also called the gain, G for an OpAmp is very large, often in the order of 10^6, but is not a constant value for all frequencies (we will cover this later). This is a problem if you want frequency independent amplification. As stated, we can use the OpAmp in the circuit in such a way that the performance is much more reliable. In order to do this we must make use of negative feedback. An example of how to connect an OpAmp to make use of negative feedback is shown on the right in Figure 5.0.1. The circuit drawn is an inverting amplifier. To determine the transfer function $H(\omega)$ of this circuit we must first cover some of the golden rules regarding the use of an OpAmp.

Wikipedia: **`http://en.wikipedia.org/wiki/Opamp`**

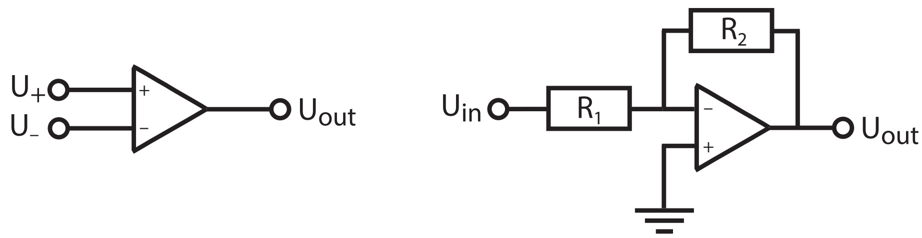


Figure 5.0.1: Left: *The OpAmp used without feedback. The output voltage U_{out} is given by the gain, G, times the difference between the input voltages: $U_{out} = G(U_+ - U_-)$* ***Right:*** *Here the OpAmp is used in combination with negative feedback. This circuit is called an inverting amplifier because the output voltage U_{out} is given by $U_{out}(\omega) = -\frac{R_2}{R_1} U_{in}(\omega)$.*

5.1 Golden rules of the ideal OpAmp

When an OpAmp is used together with feedback, the very large gain means that the voltage difference at the inputs of the OpAmp is essentially zero. Furthermore, the OpAmp is made in such a way that no current can run from the input to the output through the OpAmp itself. These two properties make it fairly easy to calculate the transfer function of an OpAmp circuit.

1. **Golden rule I:** For negative feedback, the output of an OpAmp will adjust its voltage and the current supplied in such a way that the voltage difference at the inputs is zero:

$$U_+ - U_- = 0\,. \tag{5.2}$$

2. **Golden rule II:** No current flows through the OpAmp, the input resistance of the OpAmp can be considered infinite.

3. **Stability:** Because the gain G is very large, the feedback has to be negative. This usually means that the feedback has to be connected to the negative input, or there has to be a phase difference in the feedback that makes it negative. When the feedback becomes positive, because of a phase shift for example, then the output voltage will go to the maximum possible voltage: the OpAmp becomes saturated.

We will take a further look at golden rule I later on in the chapter. It can be easily shown that golden rule I automatically holds because of golden rule II combined with the property $U_{out} = G(U_+ - U_-)$. We will then immediately see that the value of G does not have a big influence on the functioning of the OpAmp, as long as it remains large. This is the exact reason as to why the OpAmp is such a useful component in an electronic circuit.

5.2 Applications of the OpAmp

Using the golden rules of the OpAmp enables us to determine the transfer function of many of the applications of the OpAmp. We will give a few prime examples here in order to demonstrate the general methodology. We firstly take a look at the inverting amplifier, a way to get a reliable constant amplification of a signal without having to worry about a frequency dependent gain. We will also look at the differentiator and integrator, these make use of passive components in the feedback loop to get certain special properties in the transfer function of the circuit. This proves to be very useful in making Proportional-Integral-Differential (PID) controllers for example. Finally we will look at the application of the OpAmp as a converter of a (small) current to a voltage and how an OpAmp can be used to connect circuits together, as was discussed in Section 1.5.1.

5.2.1 The inverting amplifier

The circuit on the right in Figure 5.0.1 shows a system where an OpAmp is used in a circuit that provides an amplified output voltage. We use the golden rules of the OpAmp to calculate the transfer function $H(\omega)$:

$$\begin{aligned} U_+ = U_- &= 0\,, && \text{(golden rule I)} \\ I_{R_2}(\omega) = I_{R_1}(\omega) &= \frac{U_{in}(\omega)}{R_1}\,, && \text{(golden rule II)} \\ U_{out}(\omega) = -R_2 \cdot I_{R_2}(\omega) &= -\frac{R_2}{R_1} U_{in}(\omega)\,, \end{aligned}$$

$$H(\omega) = \frac{U_{out}(\omega)}{U_{in}(\omega)} = -\frac{R_2}{R_1}\,. \tag{5.3}$$

We can see that the ratio between R_2 and R_1 determines the amplification, or Voltage Gain. The minus in the transfer function is why this system is called and inverting amplifier.

Wikipedia: `http://en.wikipedia.org/wiki/Inverting_amplifier`

5.2.2 The integrator

In the previous chapter we discussed feedback systems. When the feedback signal is proportional to the incoming signal we call the feedback proportional feedback. When the feedback signal is the integral of the incoming signal we are said to have integral feedback. By making use of an OpAmp we can get a transfer function that gives the desired integral feedback. The easiest way to do this is called an integrator, which is pictured in Figure 5.2.2.

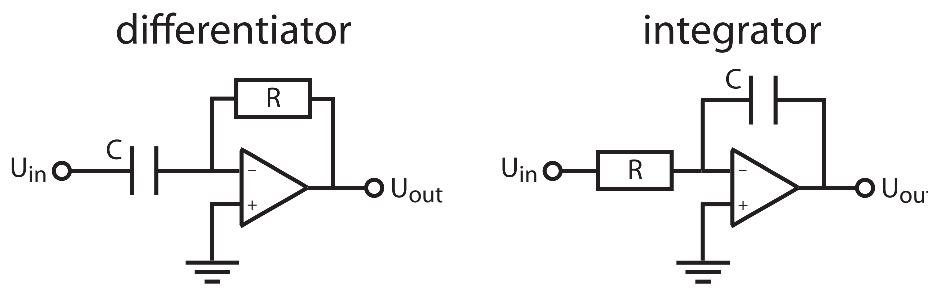


Figure 5.2.2: *Two common systems in which the OpAmp is incorporated. On the left we have the differentiator and on the right the integrator. The name comes from the link between the input and output voltages. These systems are incorporated in, for example, the Atomic Force Microscope, where the feedback is not only proportional, but also integrating and differentiating (PID systems).*

We are looking for the transfer function $H(\omega)$ of this system:

$$\begin{aligned} U_+ &= U_- = 0\,, && \text{(golden rule I)} \\ I_C(\omega) &= I_R(\omega) = \frac{U_{in}(\omega)}{R}\,, && \text{(golden rule II)} \\ U_{out}(\omega) &= -\frac{I_C(\omega)}{i\omega C} = -\frac{U_{in}(\omega)}{i\omega RC}\,, \\ H(\omega) &= \frac{U_{out}(\omega)}{U_{in}(\omega)} = -\frac{1}{i\omega RC}\,. \end{aligned} \tag{5.4}$$

We see that the transfer function contains a factor $\frac{1}{i\omega}$ in the frequency domain, this is equal to an integral over time in the time domain (see the impedance of the capacitor in Section 1.3). This explains why this system is called an integrator.

Wikipedia: `http://en.wikipedia.org/wiki/Integrator`

5.2.3 The Differentiator

Aside from proportional and integral feedback we also have differential feedback. An example of a system that can be used for differential feedback is the differentiator, see Figure 5.2.2. We again examine the transfer function to see if we really are dealing with a system that can provide differential feedback:

$$\begin{aligned} U_+ &= U_- = 0\,, && \text{(golden rule I)} \\ I_C(\omega) &= I_R(\omega) = \frac{U_{in}(\omega)}{Z_C} = U_{in}(\omega)\cdot i\omega C\,, && \text{(golden rule II)} \\ U_{out}(\omega) &= -R\cdot I_R(\omega) = -i\omega RC\cdot U_{in}(\omega)\,, \\ H(\omega) &= \frac{U_{out}(\omega)}{U_{in}(\omega)} = -i\omega RC\,. \end{aligned} \tag{5.5}$$

We see that the transfer function contains a factor $i\omega$ in the frequency domain, which is equal to a time derivative in the time domain. This explains why this system is called a differentiator.

Differential feedback can be useful when you need a very quick reaction to a sudden change in the input signal. For example, when the tip of a Scanning Tunneling Microscope encounters an atom on the surface of your sample there will be a sudden large increase in your signal. If we were to use only proportional feedback, the tip will only go up once the signal has risen by a certain amount, at which time it may be too late. By using differential feedback, feedback is given as soon as the incoming signal starts to rise. This allows the system to respond before the tip and the surface come in contact.

Wikipedia: `http://en.wikipedia.org/wiki/Differentiator`

5.2.4 The buffer amplifier

We determine the transfer function of the left-hand circuit in Figure 5.2.3:

$$U_{out} = U_- = U_+ = U_{in}\,, \qquad \text{(golden rule I)} \tag{5.6}$$

$$H(\omega) = \frac{U_{out}(\omega)}{U_{in}(\omega)} = 1\,. \tag{5.7}$$

We see that the output voltage is equal to the input voltage.

The buffer amplifier is a very simple circuit that can be used when we want to connect two systems for which $|Z_{in,2}| \not\gg |Z_{out,1}|$. In this case we cannot simply multiply the individual transfer function as we saw in Section 1.5.1. Because the input impedance of an OpAmp may be regarded as being infinite, a current will never flow through the output of a system that is connected to the OpAmp, which means that the transfer function is well defined. Additionally, the voltage at the output of the OpAmp is held constant by OpAmp, which means that when a buffer amplifier is connected between two circuits with transfer functions H_1 and H_2, the total transfer function of the system is given by the product of the two transfer functions, $H_1 \cdot H_2$.

5.2.5 The I/V converter

An I/V converter is used to convert a current into a voltage, often with an extra amplification too. A simple I/V converter consists of an OpAmp and a resistor R, as show in Figure 5.2.3.

We make use of the properties of an OpAmp to determine the transfer function $H(\omega)$ of the circuit:

$$\begin{aligned}
U_+ &= U_- = 0\,, && \text{(golden rule I)}\\
I_R(\omega) &= I_{in}(\omega)\,, && \text{(golden rule II)}\\
U_{out}(\omega) &= -I_R(\omega)R = -I_{in}(\omega)R\,, \\
H(\omega) &= \frac{U_{out}(\omega)}{I_{in}(\omega)} = -R\,.
\end{aligned} \tag{5.8}$$

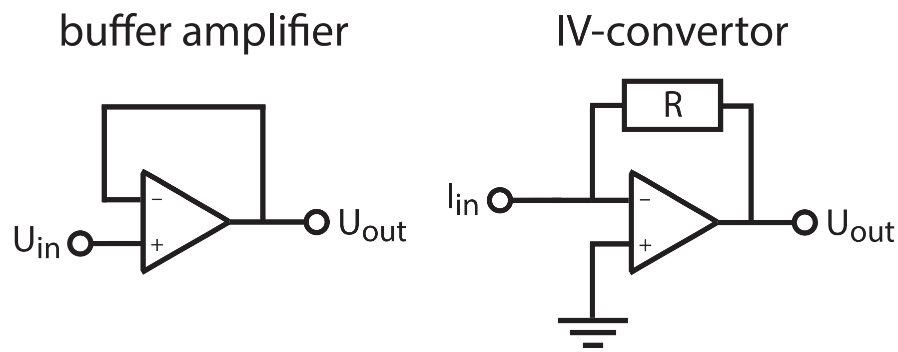


Figure 5.2.3: Left: *The buffer amplifier which is used in a circuit to prevent problems with impedances.* ***Right:*** *The IV-converter converts a current into a voltage. The amplification in this case is R.*

To get a certain amplification of a signal all you need to do is find the right resistor R. We can even easily switch between different values of R to get just the right value of the output voltage to, for instance, get the maximum resolution for an A/D converter. Making a good I/V converter is, however, not as easy as it sounds. The resistance of resistors is not fully frequency independent, the resistance may be different for different frequencies and as a result the amplification is not constant. The OpAmp can also decrease the Signal-to-Noise ratio by adding or amplifying noise.

Wikipedia: `http://en.wikipedia.org/wiki/Current-to-voltage_converter`

5.3 Non-ideal OpAmp

The systems we have come across so far were solved by using the golden rules of the ideal OpAmp. In reality we have to work with OpAmps that are not ideal in several ways.

5.3.1 The finite gain of an OpAmp

As we have discussed earlier, the use of negative feedback ensures the reliable performance of an OpAmp. To gain a further understanding of this, as well as a derivation of golden rule I ($U_+ - U_- = 0$), we will reexamine the inverting amplifier from Figure 5.0.1. Instead of using golden rule I we will now use the property that $U_{out} = G(U_+ - U_-)$:

$$\begin{aligned} U_{out} &= G(U_+ - U_-) = -GU_- \,, \\ I_{R_2}(\omega) &= I_{R_1}(\omega) = \frac{U_{in}(\omega) - U_-}{R_1} \,, && \text{(golden rule II)} \\ U_{out}(\omega) &= U_- - I_{R_2}R_2 = -\frac{U_{out}}{G} - \left(\frac{U_{in} + U_{out}/G}{R_1}\right) R_2 \,, \end{aligned}$$

$$H(\omega) = \frac{U_{out}(\omega)}{U_{in}(\omega)} = -\frac{R_2}{R_1} \cdot \left(\frac{G}{G + 1 + \frac{R_2}{R_1}}\right) . \tag{5.9}$$

We see that if $G \gg 1$ and $G \gg R_2/R_1$ we get the answer for the ideal OpAmp (Equation 5.3) where we used golden rule I instead of the property $U_{out} = G(U_+ - U_-)$.

The condition that $G \gg \frac{R_2}{R_1}$ is logical if you consider that the amplification of the entire system can never be larger than the Open-Loop amplification G.

As long as the conditions $G \gg 1$ and $G \gg \frac{R_2}{R_1}$ are satisfied, G is free to fluctuate at different frequencies, temperatures, levels of wear etc. without affecting the amplification of your system. This shows the importance of negative feedback when making a reliable and constant amplifier and the reason why the OpAmp is such an important component in an electronic circuit.

5.3.2 Frequency-dependent gain

The Gain of an OpAmp is dependent on the internal circuitry of an OpAmp and as a result may be frequency dependent, as is shown in Figure 5.3.4. It shows a simple frequency-dependence of the gain that is comparable to the transfer function of a first order low-pass filter.

With the design and use of OpAmps we refer to the Gain-Bandwidth Product (GBP). The GBP is the product of the Gain with the bandwidth over which the gain is at least that value:

- **Gain-Bandwidth Product GBP:** The product of the gain g with the bandwidth Δf at which the OpAmp has a gain greater than or equal to g is the GBP. Complicated OpAmps may have a different GBP at different gains, in other words, GBP = GBP(g)

- **Unity-Gain Bandwidth Product:** This is the GBP measured at a gain of 1. Many OpAmps have a wide range at which the GBP is equal to the Unity-GBP.

For the OpAmps with a nearly constant GBP, it is easy to calculate the bandwidth for each gain, since the minimum gain you achieve is inversely proportional to the bandwidth. For example, if you demand a gain of at least 10 with a GBP of 1 MHz, you have an available bandwidth of 0.1 MHz. If you demand a gain

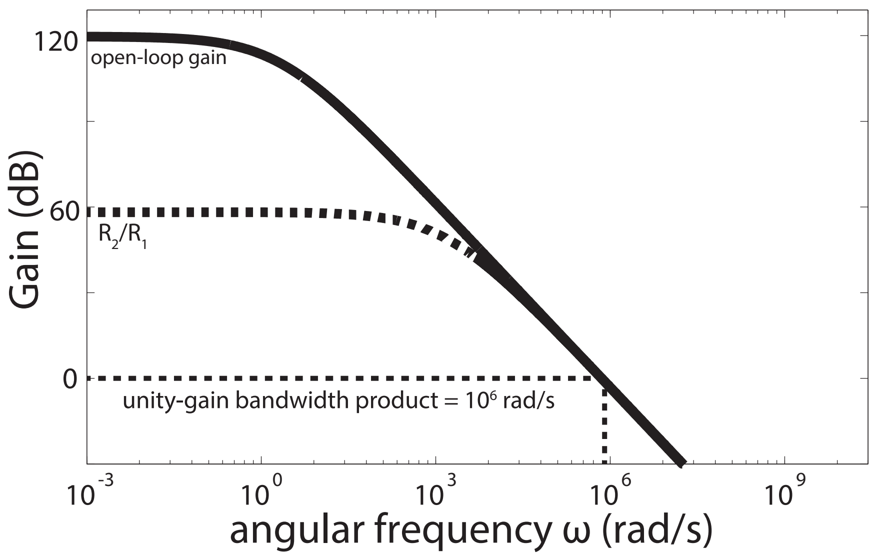


Figure 5.3.4: *Plotted is the frequency-dependence of the open-loop gain* $G(\omega) = 10^6/(1+i\omega)$ *and the transfer function of the inverting amplifier in the case that the gain is not finite and* $R_2/R_1 = 1000$ *(Equation 5.9). We can see that the frequency-dependence only plays a role when the amplification of your closed-loop circuit (which in the case of the inverting amplifier is equal to* R_2/R_1*) becomes as big as the open-loop gain of the OpAmp. Demanding a high gain of your OpAmp will result in a lower bandwidth at which this gain is obtained. Also plotted is the Unity-Gain-Bandwidth Product. Note that all axes are logarithmic and that for the gain we used a conversion* $20\log_{10}|H(\omega)|$.

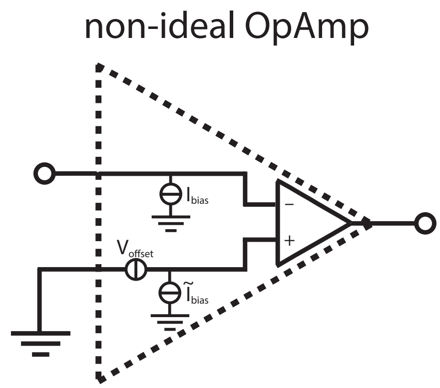


Figure 5.3.5: *A realistic OpAmp suffers from, among others, an offset voltage V_{offset} and current leakage I_{bias}. You can take these effects in to consideration in your system by using the above circuitry.*

of at least 1000 to design an amplifier, the available bandwidth is reduced to 1 kHz.

In Figure 5.3.4 the frequency-dependent gain $G(\omega) = G_0(1 + i\omega)$ has been plotted together with the effect that it has on the amplification of the inverting amplifier (Equation 5.9). We can see that the frequency-dependence of the gain only becomes important when the open-loop gain decreases to $G = R_2/R_1 = 1000$. Even though the open-loop gain drops already significantly at a frequency of a few Hz, the amplification of the inverting amplifier remains constant. This is exactly the purpose of using an OpAmp in your circuit, having a stable constant amplification for all frequencies in a certain bandwidth.

Wikipedia: **`http://en.wikipedia.org/wiki/Gain-bandwidth_product`**

5.3.3 Further issues with the non-ideal OpAmp:

- **Noise:** The components in the OpAmp cause noise. Some of this noise can even be amplified by the OpAmp. You can add noise to your model by for example, connecting a few resistors in series or parallel to the ideal OpAmp, comparable to how the model in Figure 5.3.5 accounts for offset voltages and current leakage.

- **Finite gain:** We have assumed that the gain is so large that we can assume that $U_+ = U_-$. In reality this is not the case. This causes there to be a small difference between the inverting and non-inverting input voltages.

- **Current not zero:** A (small) current passes through the OpAmp, this means that the impedance is not infinite as it is for an ideal OpAmp.

- **Offset voltage:** There is a small offset voltage at the input of an OpAmp, it is there even if there is no input signal. You can take this offset voltage and the offset current in to consideration in your model by adding extra components to the ideal OpAmp, as shown in Figure 5.3.5.

- **Phase shift:** At high frequencies the OpAmp can cause there to be a phase difference between the input and output signals. This phase difference can cause the system to become unstable, at high frequencies your negative feedback can become positive feedback for example.

Further reading

The book by Regtien [3] has comprehensive chapters on transistor and the properties of OpAmps and has many examples of circuits with OpAmps.

Bibliography

[1] K. F. Riley, M. P. Hobson and S. J. Bence, Mathematical Methods for Physics and Engineering, Cambridge University Press. New York, 2006

[2] Zoher Z. Karu, Signals and Systems, Zizi Press, 2002

[3] P. P. L. Regtien, Electronic Instrumentation, VSSD, 2005

[4] Martin van Exter, Noise and Signal Processing, 2003

[5] Wim C. van Etten, Introduction to Random Signals and Noise , John Wiley and Sons, 2006

Formulas

Angular frequency	$\omega = 2\pi f$
Impedance capacitor	$Z_C = \frac{1}{i\omega C}$
Impedance inductor	$Z_L = i\omega L$
RC low-pass filter	$H(\omega) = \frac{1}{1+i\omega RC}$
RC high-pass filter	$H(\omega) = \frac{1}{1+\frac{1}{i\omega RC}}$
Convolution	$v(x) \otimes u(x) \stackrel{\mathrm{def}}{=} \int_{-\infty}^{\infty} v(s)u(x-s)ds$
Fourier Transform	$X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-i\omega t}dt$
Inverse Fourier Transform	$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)e^{i\omega t}d\omega$
Fourier Series	$X[k] = \frac{1}{T}\int_0^T x(t)e^{\frac{-i2\pi kt}{T}}dt$
Inverse Fourier Series	$x(t) = \sum_{n=-\infty}^{\infty} X[k]e^{\frac{i2\pi kt}{T}}$
Discrete Time Fourier Transform	$X(\tilde{\omega}) = \sum_{n=-\infty}^{\infty} x_n e^{-i\tilde{\omega}n}$
Inverse Discrete Time Fourier Transform	$x_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} X(\tilde{\omega})e^{in\tilde{\omega}}d\tilde{\omega}$
Discrete Fourier Transform	$X[k] = \sum_{n=0}^{N-1} x[n]e^{-i2\pi k\frac{n}{N}}$
Inverse Discrete Fourier Transform	$x[n] = \frac{1}{N}\sum_{k=0}^{N-1} X[k]e^{i2\pi k\frac{n}{N}}$
Standard deviation	$\sigma^2 = \langle x^2(t)\rangle - < x(t) >^2$
Parseval's theorem	$\int_{-\infty}^{\infty} \lvert x(t)\rvert^2 dt = \int_{-\infty}^{\infty} \lvert X(f)\rvert^2 df$
Wiener-Khinchin theorem	$\tilde{S}(f) = \int_{-\infty}^{\infty} R(\tau)e^{-i2\pi f\tau}d\tau$
	$R_x(\tau) = \int_{-\infty}^{\infty} \tilde{S}(f)e^{i2\pi f\tau}df$
Noise transfer	$S_{out}(\omega) = \lvert H(\omega)\rvert^2 S_{in}(\omega)$
Laplace Transform	$X(s) = \int_0^{\infty} x(t)e^{-st}dt$
Delta function	$\delta(a-b) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ix(a-b)}dx$

Fourier Transforms

Function	**Time domain**	**Frequency domain**
Cosine	$x(t) = \cos(\alpha t)$	$X(\omega) = \pi(\delta(\omega - \alpha) + \delta(\omega + \alpha))$
Sine	$x(t) = \sin(\alpha t)$	$X(\omega) = \frac{\pi}{i}(\delta(\omega - \alpha) - \delta(\omega + \alpha))$
Constant	$x(t) = \alpha$	$X(\omega) = 2\pi\alpha\delta(\omega)$
Delta function	$x(t) = \alpha\delta(t - \tau)$	$X(\omega) = \alpha e^{-i\omega\tau}$
Exponential	$x(t) = e^{-\alpha\|t\|}\ \alpha > 0$	$X(\omega) = \frac{2\alpha}{\alpha^2+\omega^2}$
Gaussian	$x(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-t^2}{2\sigma^2}}$	$X(\omega) = e^{\frac{-\sigma^2\omega^2}{2}}$
Rectangular window	$x(t) = 1 \text{ for } \|t\| < \frac{T}{2}$	$X(\omega) = \frac{2\sin\left(\frac{T\omega}{2}\right)}{\omega}$

Dictionary

English:	**Dutch:**
Amplifier	Versterker
Angular frequency	Hoekfrequentie
Auxiliary equation	Karakteristieke vergelijking
Band-pass filter	Banddoorlaatfilter
Band-stop filter	Bandsperfilter
Bandwidth	Bandbreedte
Bode Magnitude Plot	Bode-amplitudeplot
Bode Phase plot	Bode-faseplot
Bounded	Begrensd
Capacitor	Condensator
Carrier wave/signal	Draaggolf
Current	Stroom
Cutoff frequency	Karakteristieke (afsnij-) frequentie
Feedback	Terugkoppeling
Gain	Versterking
High-pass filter	Hoogdoorlaatfilter
Impedance	Impedantie
Inductance	Zelfinductie
Inductor	Spoel
Load	Belasting
Low-pass filter	Laagdoorlaatfilter
Noise	Ruis
Quadratic formula	ABC-formule
Resistor	Weerstand
Saturated	Verzadigd
Sensitivity	Gevoeligheid
Shot noise	Hagelruis
Step response function	Stapresponsfunctie
Time-invariant	Tijdsinvariant
Transfer function	Overdrachtsfunctie
Voltage	Spanning
Window	Afkapvenster