Academic year
2026-2027

Faculty of Applied Engineering

Systems Theory

Text book

Walter Daems

Bachelor of Science in de industriële wetenschappen: elektronica-ICT

1512FTISYS 5-Systems Theory

University of Antwerp

This document has been typeset using LaTeX and the `uantwerpendocs` package.

This text has been typeset using LuaLaTeX. It uses — when possible — a color pallet that takes into account a reduced color perception and that has a sufficient contrast for grayscale printout.
Calculations have been performed using Matlab/Octave and generic programming and script languages (C/C++/Perl/Python).
Graphics have been composed using PGF, TiKZ and InkScape.
All this material has been prepared on a GNU/Linux workstation.

Typesetting of this document was enabled by:

SYS-2026-1.1-TB

Contents

1 Introduction: what is Systems Theory? 1

1.1 The answer — by example . . . 1

1.2 It's all about models . . . 2

1.2.1 Models a.k.a. equations . . . 2

1.2.2 Models a.k.a. boxes . . . 3

1.3 The answer — in a structured way . . . 4

1.3.1 Definition . . . 4

1.3.2 Variants . . . 4

1.3.3 Classification . . . 6

1.4 Conclusion . . . 7

2 Signals 9

2.1 Definition of a signal . . . 9

2.2 Signal operations . . . 10

2.2.1 Operations . . . 10

2.3 Elementary signal properties . . . 11

2.3.1 Time limited vs. time unlimited signals . . . 11

2.3.2 Periodic vs. aperiodic signals . . . 11

2.3.3 Even/odd signals . . . 11

2.4 Example signals . . . 12

2.4.1 Physical signals . . . 12

2.4.2 Nonphysical signals . . . 16

2.5 Conclusion . . . 22

3 Systems **23**

3.1 Definition of a system . . . 23

3.2 Classification . . . 24

3.2.1 Time invariant vs. time variant systems . . . 24

3.2.2 Linear vs. nonlinear systems . . . 25

3.3 The engineering sweetspot: LTI systems . . . 27

3.3.1 Definition . . . 27

3.3.2 Important properties . . . 27

3.3.3 Linear differential equations . . . 29

3.3.4 Block diagrams . . . 29

3.3.5 Impulse response description . . . 33

3.3.6 Transformations . . . 37

3.4 Counting the inputs and the outputs . . . 38

3.4.1 SISO . . . 38

3.4.2 SIMO . . . 39

3.4.3 MISO . . . 39

3.4.4 MIMO . . . 39

3.5 Conclusion . . . 40

4 Convolution **41**

4.1 Why convolution? . . . 41

4.2 Impulse decomposition . . . 42

4.3 The convolution operation and its properties . . . 44

4.3.1 Definition . . . 44

4.3.2 Graphical example . . . 45

4.3.3 Properties . . . 46

4.4 Example . . . 48

4.5 Conclusion . . . 50

5 The Continous-time Fourier transform **51**

5.1 Introduction 51

5.2 Vectors and approximating them 52

5.2.1 Vectors in 2D 52

5.2.2 Signals as vectors 53

5.2.3 Scalar product for signals 55

5.2.4 The length or size of signals 56

5.2.5 Approximating signals 57

5.2.6 Improving the approximation 59

5.3 The Fourier series 61

5.3.1 Trigonometric series 61

5.3.2 Alternative trigonometric series 67

5.3.3 Exponential series 68

5.3.4 Conversion between exponential and trigonometric coefficients 73

5.3.5 Parseval's theorem 74

5.3.6 Spectra 74

5.4 The Fourier transform 75

5.4.1 Derivation 75

5.4.2 Definition 77

5.4.3 Examples 80

5.4.4 Properties 85

5.5 Description of LTI systems in the Fourier domain 88

5.5.1 The frequency transfer function 89

5.5.2 Eigenfunctions of LTI-systems 91

5.5.3 Determining a frequency transfer function 92

5.6 Conclusion 97

6 The Laplace transform **99**

6.1 From Fourier transform to Laplace transform 99

6.1.1 Forward... 99
6.1.2 ...and back 101
6.2 One-sided Laplace transform 102
6.3 Region of convergence 102
6.4 Properties 103
6.5 Practical ways to calculate the (inverse) Laplace transform 106
6.5.1 Forward 106
6.5.2 Inverse 107
6.6 Common transform pairs 109
6.7 Description of LTI-systems in the Laplace domain 109
6.7.1 The transfer function 110
6.7.2 Rational Laplace transforms - zeros and poles 112
6.7.3 Stability 113
6.8 Why do we need the Laplace transform? 116
6.9 Applications 118
6.9.1 Solving differential equations 118
6.9.2 Solving linear electronic networks 120
6.10 The extended Fourier family diagram 127

7 Analyzing LTI systems — time domain 129
7.1 Introduction 129
7.2 Basic analysis technique 130
7.2.1 Principle 130
7.2.2 Transfer function 130
7.3 First-order systems 132
7.3.1 Description 132
7.3.2 Impulse response 132
7.3.3 Step response 133
7.4 Second-order systems 134

7.4.1 Description . . . 134
7.4.2 Impulse response . . . 136
7.4.3 Step response . . . 137
7.5 The effect of zeros . . . 139
7.5.1 Description . . . 139
7.5.2 First-order systems with a zero . . . 141
7.5.3 Second-order systems with a zero . . . 141
7.5.4 Higher-order systems with zeros . . . 142
7.6 Examples . . . 143
7.6.1 An RC filter . . . 143
7.6.2 An RLC circuit . . . 148
7.6.3 Calculating switching transients in a spark plug voltage generator . . . 149
7.7 Conclusion . . . 154

8 Analyzing LTI systems — frequency domain 155

8.1 Introduction . . . 155
8.2 Composing frequency transfer functions . . . 157
8.2.1 Calculating the frequency transfer function . . . 157
8.3 Analysis in the Argand plane . . . 160
8.3.1 Some things to remember . . . 160
8.3.2 Recognizing the vectors in a factorized transfer function . . . 162
8.3.3 Using the vectors in a factorized transfer function . . . 162
8.3.4 First-order systems . . . 163
8.3.5 Second-order systems . . . 166
8.3.6 All-pass systems . . . 171
8.3.7 Minimum vs. non-minimum phase systems . . . 171
8.4 Bode plots — in theory . . . 174
8.4.1 Genesis . . . 174
8.4.2 The logarithmic magnitude axis — two worlds . . . 176

8.4.3 The logarithmic frequency axis — land of confusion . . . 176
8.4.4 The combination of both — sledgehammer! . . . 177
8.5 Bode plots — in practice . . . 180
8.5.1 The bricks . . . 180
8.5.2 Making the bricks 'standard size' . . . 181
8.5.3 Analyzing the standard-size bricks . . . 182
8.5.4 Resonance correction . . . 185
8.5.5 Procedure . . . 186
8.6 Examples . . . 188
8.6.1 Example 1 . . . 188
8.6.2 Example 2 . . . 190
8.7 Conclusion . . . 196

A The Decibel **197**
A.1 Definition . . . 197
A.1.1 Power ratio . . . 197
A.1.2 Absolute power quantity . . . 198
A.2 Use . . . 198
A.3 Examples . . . 199
A.4 Rules of thumb for calculating with dBs . . . 200
A.5 Conclusion . . . 200

B Representing impulse trains: the Sha-function **201**

Preface

The scenery

Systems theory is a topic that is taught in almost every engineering domain. It describes the foundations of modeling physical systems using the mathematics of calculus. Abstracting physical systems into continuous differential equations allows for analyzing, simulating and synthesizing systems.

Systems can be of various nature, such as mechanical systems, electromagnetic systems, hydraulic systems and chemical systems. It will be no surprise to the reader that this book will focus mostly on electrical (and electronic) systems as the course this book is intended for, is being taught to students in Electronics.

This book mostly concentrates on linear systems, as these are still the main workhorse for everyday engineering. Linear systems offer the advantage of being easy to analyze, test and predict.

The script

This book is the first in a series of four books on electronic systems and signal processing:

1. Systems Theory — Signals and Transformations
2. Digital Signal Processing — Signals and Transformations
3. Digital Image Processing
4. Digital Signal Processing — Signal Processing Systems

This first two volumes focus on signals and signal transformations (in the analog and in the digital domain). The third volume focuses on image processing. The fourth focuses on building (digital) signal processing systems.

On the language used in this textbook: I tried to write this book from the perspective of a tutor guiding his tutees. Therefore, the text lacks the formality of many scientific text books. I hope you like this style. Where appropriate, I left out some mathematical derivations in order not to clutter the overall picture. I tried to add as much hints as needed to allow you working your way through the (sometimes difficult) material on your own if you like. The "he-him-his" formulation that has been used is not to emphasize the lack of women in engineering. This wording (instead of the more elaborate he/she, him/her and his/hers) has been chosen to keep this text simple.

A lot of effort has been put into this edition.

- This edition has been equipped with a solution book containing the solutions to the exercises. Making exercises is thé way to make sure you understand the theory. Exercises marked with (*) are a bit harder than the standard non-marked exercises. Those marked with (**) are for the enthusiastic reader (to fill any rare rainy days).
- This edition has been equipped with a formula collection that brings the important equations and definitions within reach in a convenient survey. If you think equations are missing from this collection, please inform me about this and I will consider adding them.

Though I try to avoid any errors, human erring is of all times. Especially, given the fact that this is a first edition, expect to find a considerable amount of errors. Do not hesitate to check with me if you find any errors. Even when you think there is an error and there's not, you will gain my appreciation for taking your understanding of this material seriously.

Most of the material in this textbook is my original work or that of my predecessors. Some of it has been taken from other (free/open) sources. In case you notice a copyright infringement (or a reference that is not clear), please contact me. It is my firm desire to be 100% in line with the copyright legislation.
If there happens to be an infringement, I apologize for it. I will do all that is reasonably possible to overcome that issue as soon as it is brought to my attention.

The crew

Finally, I would like to thank many people who contributed to this text.
First, a special thanks to my editor, Paul Levrie, for helping me by reviewing this text and supporting me with references, his profound experience, joyful humor and music. Some of the chapters have been inspired on the nice material (in Dutch) composed by my former colleagues, P.E.M. Van den Wyngaert [vdW92] and M. Van Paemel [Van18]. Thanks to Nico Huebel and Jan Steckel for bringing interesting background material to my attention.
Thanks to Wouter Jansen, Thomas Verellen and Toon Stas, for their contributions in composing exercises and guiding the exercise series.
Finally, my deepest gratitude goes to my beloved wife and children for enduring me devoting my time to writing this text, instead of spending my time with them.

Any contribution to this work is welcome. You won't get any money for it. I can only offer you to be on my list of favorite people. You can contact me by e-mail to `walter.daems@uantwerpen.be`.

I hope you enjoy discovering digital signal processing!

Walter Daems
Summer 2026
Trolhattan, Västergötland (Sweden)

Symbol Table

Symbol	Meaning
Number sets	
$\mathbb{N}$	set of natural numbers (positive integer numbers)
$\mathbb{Z}$	set of integer numbers
$\mathbb{Q}$	set of rational numbers
$\mathbb{I}$	set of irrational numbers
$\mathbb{R}$	set of real numbers
$\mathbb{C}$	set of complex real numbers
$\mathbb{X}^+$	set $\mathbb{X}$ restricted to positive numbers (not for $\mathbb{C}$)
$\mathbb{X}^-$	set $\mathbb{X}$ restricted to negative numbers (not for $\mathbb{C}$)
$\mathbb{X}_0$	set $\mathbb{X}$ with 0 excluded
Transform symbols - Forward	
$C_k = \mathrm{FS}\left(x(t)\right)$	C_k are the harmonic numbers of the complex Fourier Series of $x(t)$
$X(\omega) = \mathcal{F}\left(x(t)\right)$	$X(\omega)$ is the Fourier Transform of $x(t)$
$X(s) = \mathcal{L}\left(x(t)\right)$	$X(s)$ is the Laplace Transform of $x(t)$
Transform symbols - Inverse	
$x(t) = \mathrm{FS}^{-1}\left(C_k\right)$	C_k are the harmonic numbers of the complex Fourier Series of $x(t)$
$x(t) = \mathcal{F}^{-1}\left(X(\omega)\right)$	$X(\omega)$ is the Fourier Transform of $x(t)$
$x(t) = \mathcal{L}^{-1}\left(X(s)\right)$	$X(s)$ is the Laplace Transform of $x(t)$
Mapping symbols	
$x(t) \xrightarrow{\mathrm{FS}} C_k$	C_k are the harmonic numbers of the complex Fourier Series of $x(t)$
$x(t) \xrightarrow{\mathcal{F}} X(\omega)$	$X(\omega)$ is the Fourier Transform of $x(t)$
$x(t) \xrightarrow{\mathcal{L}} X(s)$	$X(s)$ is the Laplace Transform of $x(t)$
DAC/ADC related acronyms	
S/N	Signal to Noise ratio

continued on next page

continued from previous page

Symbol	Meaning
THD	Total Harmonic Distortion
RMS	Root-Mean Square value
SNAD	Signal to Noise And Distortion ratio
SFDR	Spurious Free Dynamic Range
PFE	Partial Fraction Expansion

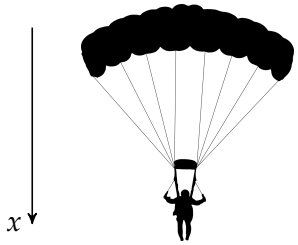
Chapter 1

Introduction: what is Systems Theory?

In this chapter, you will learn what systems theory is. We will start by giving an example that shows you the main gist of it. Then we will highlight the importance of models, to end with a structured overview of systems theory from the different perspectives. In the conclusion, we outline the rest of the course.

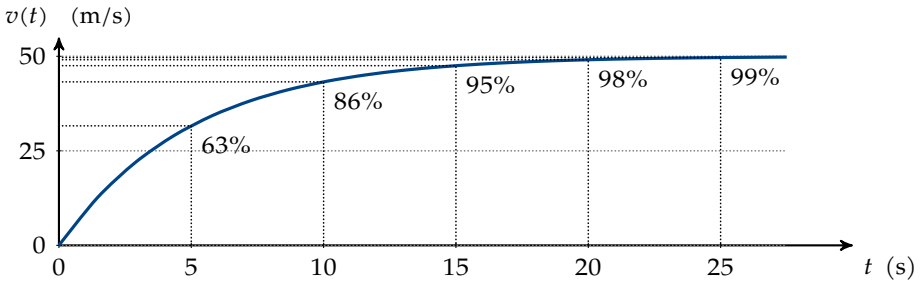
1.1 The answer — by example

Consider a skydiver that is on a plane, flying at a considerable altitude. Curious as we are, we feel like experimenting a little: we want to know what will happen if we ask her to jump off.



For the more caring amongst the readers: she's a trained skydiver with sufficient experience. She'll tell us if we ask her something that is not safe.

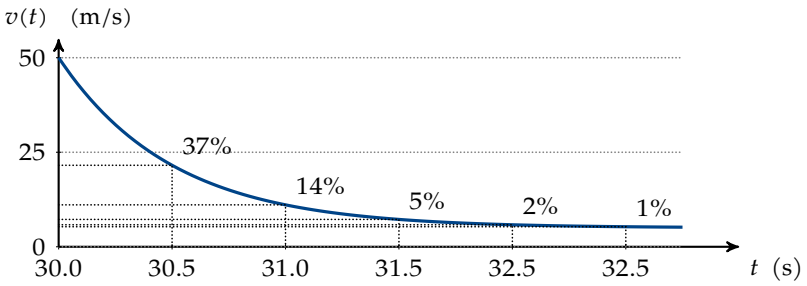
Not knowing systems theory, we need to revert to what we can do without it, and that is experiment and measure. To this end we gave the skydiver a velocity measuring device, such that, during her jump, she can record her velocity. The velocity is measured in the direction of her jump and oriented downwards (along the x-axis in the drawing above). We instruct her to jump and wait for thirty seconds, before opening her parachute. Safely back on the ground, she reports to us and shows us the graph that displays the first thirty seconds:



As we can see, the velocity increases according to an exponentially flattening curve to reach a constant velocity of 50 m/s after about 20 to 25 s. We call this velocity *terminal velocity*. We realize that it must have been breathtaking to travel through the air at 180 km/h. Our intuition helps us in explaining this curve, knowing that the gravity of the earth is pulling her down (explaining the acceleration, according to Newton's second law) and the wind drag is probably gradually slowing her acceleration down more and more when her speed increases (explaining the flattening of the curve).

Interesting. However, we ask ourselves whether it would have been possible to predict in advance the terminal velocity and the time it takes to reach that velocity.

Next, she shows us the graph of her velocity from the moment that she opened her parachute:



After only two and a half seconds her velocity is reduced to 18 km/h. That's a brutal deceleration. We're glad it wasn't us that were hanging in that parachute.

Again, could we have anticipated this fierce deceleration? Could we have predicted that she'd keep falling at a rate of 18 km/h? And wasn't that too fast to land on the ground safely? We realize that she is standing safely next to us and therefore we're good. However, we realize that we'd better study physics a bit more, before we further experiment with her.

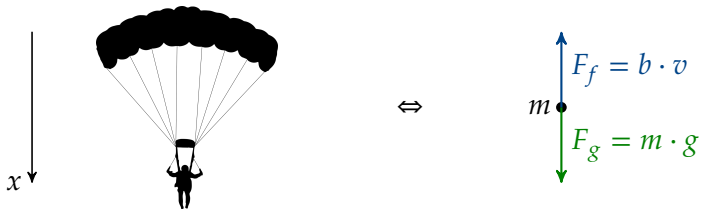
The study of the relationship of systems and their defining parameters and quantities in a specific context over time (like in this example the skydiver and her weight, her parachute and her velocity when jumping from a plane in the earth atmosphere) is called *systems theory*. This is what we'll be studying in this course.

1.2 It's all about models

1.2.1 Models a.k.a. equations

The central theme of systems theory are models, i.e. (mathematical) relationships describing the behavior of a system. In the case of the skydiver, this model is Newton's second law, together with equations that describe the forces acting on the skydiver, i.e. gravity and velocity-related drag of an object traveling through the air. This allows modeling the skydiver-system by creating a *free body diagram*, as you have learnt in your mechanics courses.

Assuming m is the mass of our skydiver, and v her velocity ($v = \frac{\mathrm{d}x}{\mathrm{d}t}$), we can discern two forces working on her: gravity (acting downwards) and drag (acting upwards).



Labeling her acceleration a ($a = \frac{dv}{dt}$) we can write down the following system of equations:

$$\begin{aligned} F_g - F_f &= m \cdot a \\ F_g &= m \cdot g \\ F_f &= b \cdot v \end{aligned}$$

Eliminating F_g and F_f from these equations, allows us to write the following differential equation:

$$\tau \frac{dv}{dt} + v = \tau g$$

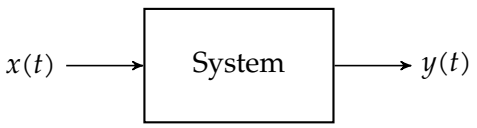
with $\tau = m/b$. This is the time constant that defines the reaction speed of the system (i.e. how long it takes to return to a stable situation, or as we often call it, a *stationary* situation). When the skydiver jumped from the plane with her parachute closed, b was small, leading to a large time constant, while after opening her parachute, the drag was considerably higher, leading to a small time constant.

Note that the equation above is a first-order linear differential equation in v. If you ever wondered why you had to spend such a significant amount of your life in studying differential equations and how to solve them, well... this is why. The obtained differential equation allows us to analyze this system using any initial condition we'd like. Of course, as long as x is such that the skydiver has not hit the ground. In that case the model we used becomes invalid. This also indicates that models are only valid in a certain *validity frame*.

Equations are often the description of some operation principle of a system and therefore model the system. Because models are very often equations and vice versa, we will be using the terms almost interchangeably.

1.2.2 Models a.k.a. boxes

Very often, we will depict the set of equations that describes the behavior of a system as a box. Quantities that we can control as a user of the system are depicted by variables with an arrow leading into the box and the resulting quantities that show the behavior of the system are depicted by arrows coming out of the box pointing to the observation we can make of the system. The former are called the *input signals* of the system, the latter the *output signals* of the system.



Based on how much we know about the equations describing the behavior of the system, we attribute a color to the box. We discern

a white box model a system about which we know all there is to know to calculate its behavior to a satisfactory level of detail;

a gray box model a system about which we know some things, but not everything; to get to know the system further, we need to study it using experimentation;

a black box model a system about which we know nothing; our only option is to experiment with the system. Automating this learning process is called *machine learning*.

1.3 The answer — in a structured way

1.3.1 Definition

Systems theory

Systems theory analyzes a well-defined set of processes (the system) that interact in a specific context and as such show some specific behavior (partly in observable quantities, partly in non-observable quantities) that may be subject to some influence we may exert on it (via input quantities). We denote the mentioned quantities as signals.

Depending on what parts in the definition above are given and what parts are unknown, we can distinguish multiple variants of systems theory:

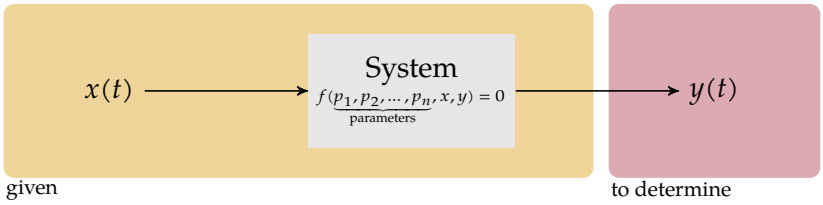
- system analysis
- system identification
- system synthesis
- control theory

Let's discuss these variants below.

1.3.2 Variants

System analysis
In system analysis, the system and its input signals are given. Using that information, we try to predict what the output signals will be, or — if possible — what the relationship between the input and the output signals is. Often the system can be described using some equations that may contain some specific numerical values, which we call the *parameters* of the system.

Graphically:

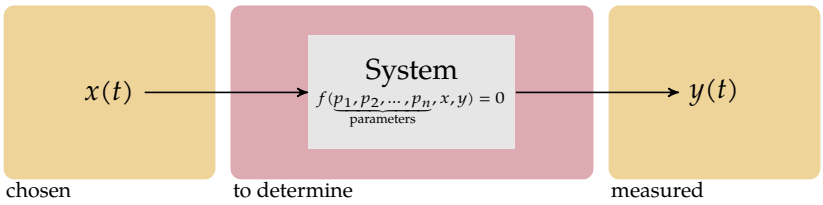


For example, we found a schematic of a music amplifier, and want to know how it works, what filtering it performs, what the input and output impedances are and what its maximal output power is.

System identification

System identification, treats a different problem. It starts from a situation in which we don't know the specific parameters that govern the system (or in a less favorable situation the equations that govern the system). In that case, the only thing we can do is to run some specific experiments, i.e. applying some input signals and observing the outputs. Our goal is to determine the missing parameters (or missing equations) from the $(x, y)-$pairs that we gathered in the experiments.

Graphically:

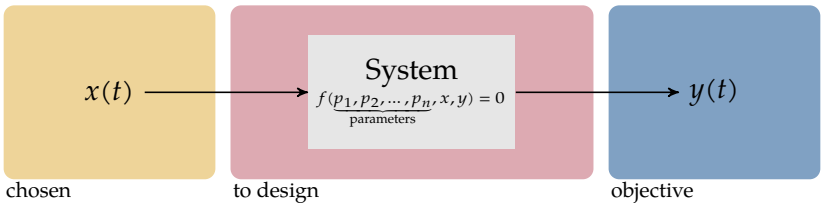


For example, we found a reluctance motor that we'd like to use in an autonomously guided vehicle and we need to determine its parameters to be able to use it in a safe way.

System synthesis

System synthesis, is the summum of engineering: we have a specific relationship between some input signals and output signals in mind, and we want to create a system that implements that relationship.

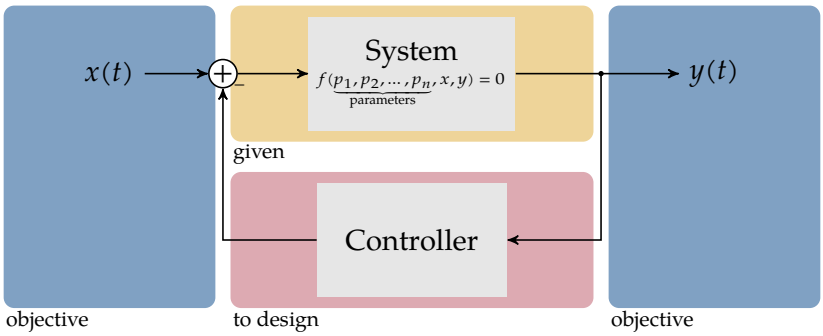
Graphically:



For example, we want to build a digital broadcasting receiver and we need to design a digital band-pass filter to allow it to tune in on the desired station (i.e. select the correct frequency band).

Control theory

And now for the nec plus ultra of engineering: control theory. Make a *controller* that measures the output of a system and provides feedback to its input, such that the system is easily set to the right behavior using the true input x.



For example, building a cruise control system that measures the speed of a vehicle and that adds the required extra throttle or braking to the input, such that a constant (desired) velocity is maintained.

1.3.3 Classification

Depending on the nature of the processes (their components) and their quantities that appear in systems, we can distinguish systems in many domains. A non-exhaustive overview is given in Table 1.1.

In addition, systems might even be crossing domains. E.g., you will rarely see a chemical plant in which only chemical processes are running. There will be electrical, electronic, hydraulic and thermal processes as well. And we probably need to study them together.
The fact that we will be able to abstract them all into mathematical models is actually quite a relief.

Domain	Examples of Components	Quantities
electrical	transfo/motors/generator	voltages, currents
electronic	r/c/l/tor	voltages, currents, fields
hydraulic	pipes, vessels, pumps	pressure, flow
thermal	heater, cooler, spreader, isolator	heat flow, temperature
optical	lamps, lenses, detectors	light flux
chemical	reactors, pipes, heat exchangers	conversion rate, energy
social	people	actions, mental state
financial	stock exchange, value adders, consumers	money, goods
political	governments, people	laws, wellfare, wellbeing
ecological	polluters, cleaners, nature	toxicity, diversity, sustainability

Table 1.1: Overview of some typical domains of systems

1.4 Conclusion

Systems theory is the theory of interaction of processes[1] with the objective of analyzing them, identifying their parameters, creating them or controlling them.

Given what we have discussed in this chapter, it may be obvious that we first need to study the quantities that appear in systems (that we call signals). As soon as we've done that, we will be able to appreciate that systems can be seen as *signal processors*. In view of this, systems theory could also be labeled as *signal processing*.

[1] You shouldn't be surprised that there is also a domain called systems theory in the domain of psychology, studying the interaction of human processes to achieve a bigger goal. Though related in goal, their essence is quite different. Maybe the one of psychology is more difficult as we are more capable of keeping predictability in our processes.

Chapter 2

Signals

In this chapter, you will learn:

- what signals are,
- how they can be represented in a graph,
- that signal calculus is actually nothing more than function calculus,
- some important signal properties,
- some examples of important signals.

After having studied this chapter, you should be able to:

- describe what a signal is and make a correct drawing of a signal,
- understand and recognize some of their basic properties,
- describe and define a number of physical and nonphysical examples.

2.1 Definition of a signal

In the most practical sense, a signal is a relationship between two physical quantities. E.g., for the powerline entering your home (feeding your domestic appliances), this could be the AC voltage as a function of time, or the frequency (exhibiting some frequency distortion or noise around 50 Hz) as a function of time. It can also be quantities that seem less technical, e.g., the price of your most favorite beer during your latest visit to the US, or the daily windspeed (in knots) during your surfing holiday.

More theoretically sound, we can make the following definition:

Definition of a signal

A signal is a mathematical function, mapping one quantity x to another y.

$$x \mapsto y$$

This strict mathematical view allows defining some terminology that allows discussing signals in a more rigorous way. The signal has an *independent* variable, associated to a particular *domain* and a *dependent* variable, associated to a certain *range*.

Graphs

Very often, signals are plotted in a graph (as functions are). We usually put the independent variable in an orthogonal graph along a horizontal axis, the so-called x-axis (or abscissa) and the dependent variable along a vertical axis, the so-called y-axis (or ordinate). Consider, e.g., the graphical representation of the observed water level at the Niagara Intake Water Station, New York, USA, as a function of time (Figure 2.1). The careful reader will observe that the axes themselves are not present on the on the graph. Instead a system of bounding box plus tick marks is used to indicate the orientation, scale and values on the axes. Where mathematicians might shiver by the mere sight of this, it is quite common for scientists and engineers to do so.

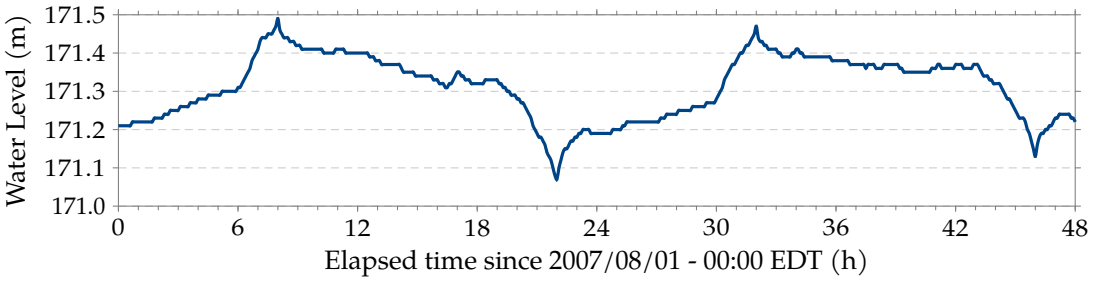


Figure 2.1: The observed waterlevel at the Niagara Intake Water Station, New York, USA, on August 1st and August 2nd, 2007 (*source: Water Level Stations Monitoring on Great Lakes Online* (*http://glakesonline.nos.noaa.gov/monitoring.html*))

Please, note the fact that in order for such a graph to become meaningful, a minimum amount of information is required. In addition to caption or title of the graph describing the relationship, the portrayed quantities, their units, and an indication of the scale, and, if relevant, one or more reference points for each axis, are minimal requirements for a graph to become a useful piece of information. Keep in mind that a graph without this minimal amount of information is *useless*. A grid may help, but very often tends to overload a graph.

2.2 Signal operations

2.2.1 Operations

If signals are just functions, then the whole myriad of possible operations on functions can be applied to any of these signals:

- unary operations
  - $\pm, |\cdot|$
  - $\exp(\cdot), \ln(\cdot)$
  - $\sin(\cdot), \cos(\cdot), \tan(\cdot)$
  - integration (running sum), differentation (finite difference)
  - substitution of the independent variable
  - ...
- binary operations

- multiplication, division
- addition, subtraction
- convolution, correlation
- ...

Of course any combination or any inverse of these functions is also a valid signal operation.

2.3 Elementary signal properties

2.3.1 Time limited vs. time unlimited signals

A signal $x(t)$ is called *time limited* if it only has nonzero values for a limited interval in time. Often we also denote this signal as a *finite-footprint* or *finite-support* signal. Mathematically:

$$\exists t_0, t_1 \in \operatorname{dom} x : t \notin [t_0, t_1] \Rightarrow x(t) = 0$$

If we choose t_0 and t_1 such that $t_1 - t_0$ is minimal, then we call the interval $[t_0, t_1]$, the footprint, the support or the time span of the signal.

A signal that is not time limited, is time unlimited.

2.3.2 Periodic vs. aperiodic signals

A signal $x(t)$ is periodic if and only if:

$$\exists T \in \mathbb{R}_0^+ : \forall t \in \operatorname{dom} x : x(t) = x(t+T)$$

The smallest value for T that exists is called the period of the signal. A signal that is not periodic is called aperiodic.

2.3.3 Even/odd signals

Though even and odd signals will rarely occur in real-life conditions, oddness and evenness are interesting theoretical properties a signal can exhibit. Later on, we will see how to exploit them to our benefit.

A continuous signal $x(t)$ is called *even* if and only if

$$\forall t \in \operatorname{dom} x : x(-t) = x(t)$$

A continuous signal $x(t)$ is called *odd* if and only if

$$\forall t \in \operatorname{dom} x : x(-t) = -x(t)$$

Obvious examples of signals that exhibit these properties are sinusoids.

2.4 Example signals

In the list of example signals that we will discuss below, we make a distinction between physical and nonphysical signals. Very often we will denote them as *waves*. The former are signals that we can meet in reality (i.e. in physics):

- step waves
- ramps
- sinusoidal waves
- square waves
- triangle (or triangular) waves

The latter are signals that we only can meet in the abstract world of mathematics:

- sinors
- phasors
- Dirac impulses

Though we will not encounter them in practice, they are as important as the physical signals, as they simplify our analysis of systems considerably.

2.4.1 Physical signals

Step waves

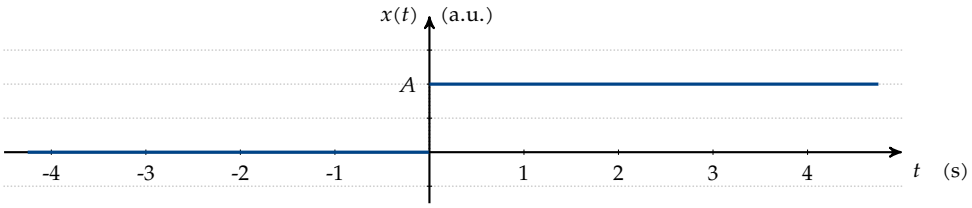
A step wave is a transition from one level to another level. It's most primitive form is the unit-step or *Heaviside* function denoted as $\mathrm{u}(t)$, making a step from 0 to 1:

$$\mathrm{u}(t) = \begin{cases} 0 & \text{if } t < 0 \\ 1 & \text{if } t \geq 0 \end{cases}$$

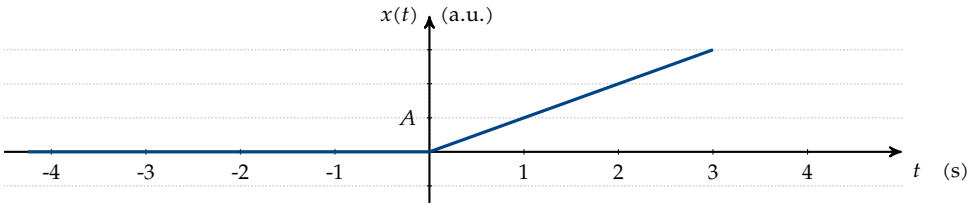
However, we can make any step, by providing the Heaviside function with an appropriate amplitude factor:

$$x(t) = A \cdot \mathrm{u}(t)$$

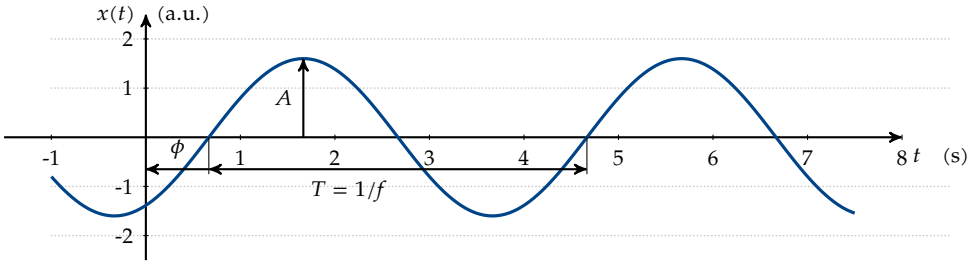
Its graphical representation can be found in Figure 2.2a. Note that we don't indicate on the graph that the definition states that $x(0) = A$ and not $x(0) = 0$ (as is usually done in mathematics books using a filled circle and an open circle), as in engineering, the difference is irrelevant.



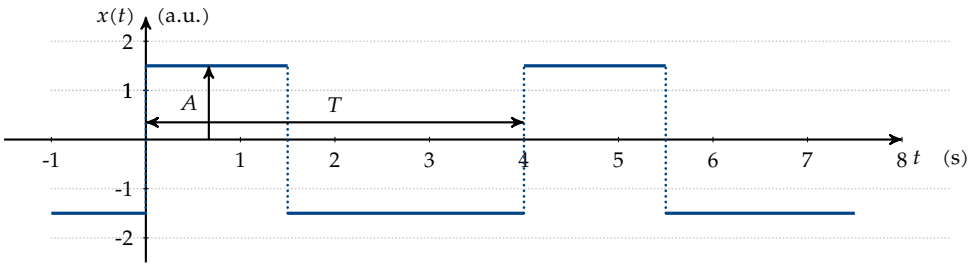
(a) Step function with amplitude A



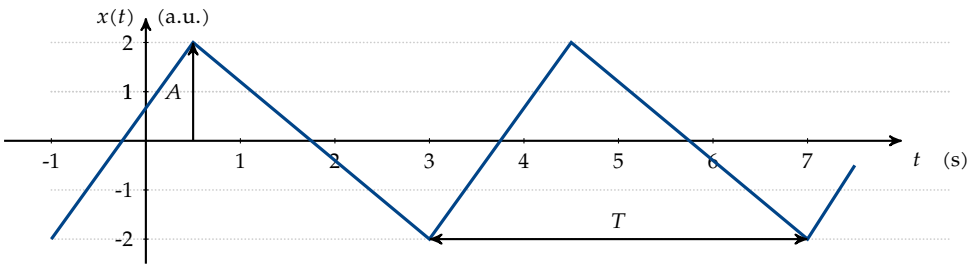
(b) First-order ramp function with gain A



(c) Sinusoidal wave with a frequency of 0.25 Hz, an amplitude of 1.6 and a phase offset of $-\pi/3$



(d) Square wave with an amplitude of 1.5, frequency 0.25 Hz and a duty cycle of 37.5 %



(e) Triangle wave with an amplitude of 2, a frequency of 0.25 Hz and a symmetry of 37.5 %

Figure 2.2: Graphical representation of some common physical waves

Ramps

A ramp is a gradual increase/decrease. It's most primitive form is the unit-ramp function:

$$\mathrm{r}_1(t) = \begin{cases} 0 & \text{if } t < 0 \\ t & \text{if } t \geq 0 \end{cases}$$

which can be generalized into an $n-th$ order unit ramp (with $n \in \mathbb{N}$) as follows:

$$\mathrm{r}_n(t) = \begin{cases} 0 & \text{if } t < 0 \\ t^n & \text{if } t \geq 0 \end{cases}$$

Note that a zeroth-order unit ramp is actually a unit step function.

Of course, we can make any ramp, by providing the ramp function with an appropriate amplitude factor:

$$x(t) = A \cdot \mathrm{r}_n(t)$$

Its graphical representation can be found in Figure 2.2b. Note that we don't indicate on the graph that the definition states that $x(0) = A$ and not $x(0) = 0$ (as is usually done in mathematics books using a filled circle and an open circle), as in engineering, the difference is irrelevant.

Sinusoidal waves

A sinusoidal wave is constructed using a sine or cosine function, and is parameterized with an amplitude A, a pulsation (or anguluar frequency) ω and a delay ϕ:

$$x(t) = A \cdot \sin(\omega t + \phi) \qquad \text{with} \qquad \omega = 2\pi f = \frac{2\pi}{T}$$

Its graphical representation can be found in Figure **??**.

Square waves

A square wave is constructed using an amplitude, a duty cycle and a period. Note that:

- $f = 1/T$ (in Hz) with T the period
- duty cycle = high time / period

The *high time* is the duration during wich the signal holds the high value without interruption. The graphical representation of this signal can be found in Figure 2.2d.

Triangle waves

A triangle wave is constructed similarly to a square wave using an amplitude and a period, but has a symmetry instead of a duty cycle.
Note that:

- $f = 1/T$ (in Hz) with T the period
- symmetry = rise time / period

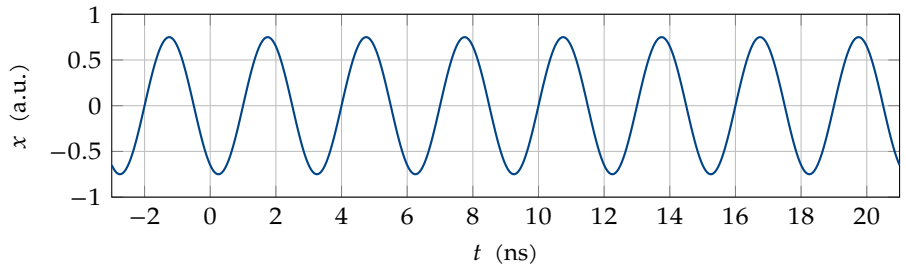
The *rise time* is the duration during wich the signal goes up in value without interruption. The graphical representation of this signal can be found in Figure 2.2e.

Exercises

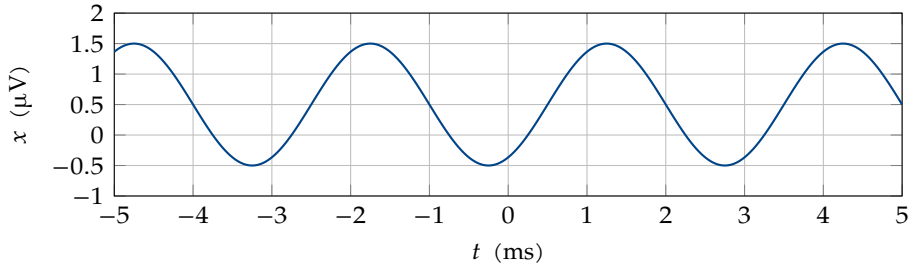
Exercise 2.4.1-1:
Draw a sine wave with amplitude 0.5 V and a frequency of 200 Hz.

- Write down its mathematical expression.
- Calculate the period of the sine wave.
- Calculate the pulsation (angular frequency) of the sine wave.

Exercise 2.4.1-2:
Write down a mathematical expression for the waveform graphed below:

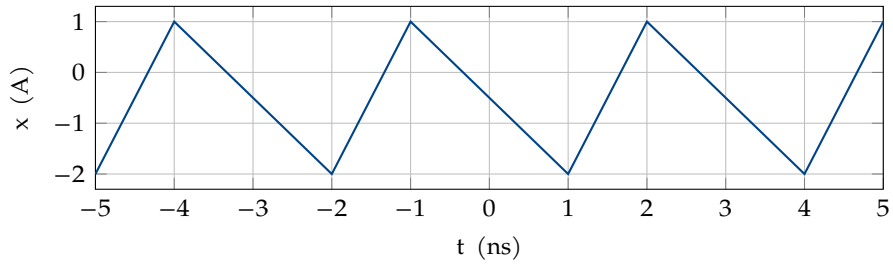


Exercise 2.4.1-3:
Determine the expression for the following signal:



Exercise 2.4.1-4:
Draw a square wave current with frequency of 150 Hz and a duty cycle of 25%. The low level is −2 A. The high level is 1 A.

Exercise 2.4.1-5:
Consider the triangle wave below. Determine its frequency and symmetry.



2.4.2 Nonphysical signals

Though physical signals are very practical, we need a set of signals to allow us to make the theoretical treatment of systems as crisp as possible. Nonphysical signals are signals that don't exist in nature. They only exist in our (mathematical) imagination:

- signals that are complex (i.e. have a real and imaginary part)
- signals that are infinitesimally short and have an infinitely large amplitude

The latter definition is a bit of a bodge. We will refine it later.

Sinors

What are they?

The sinor can be considered to be the atomic particle of the exponential Fourier theory, as we will see it later. It has an amplitude A and a pulsation ω:

$$x(t) = A \cdot \mathrm{e}^{\mathrm{j}\omega t} \qquad \text{with} \qquad \omega = 2\pi f = \frac{2\pi}{T}$$

Given Euler's relationship ($\mathrm{e}^{\mathrm{j}\alpha} = \cos\alpha + \mathrm{j}\sin\alpha$) we can rewrite this as:

$$x(t) = \underbrace{A\cos\omega t}_{\text{real part}} + \underbrace{jA\sin\omega t}_{\text{imaginary part}}$$

As we can see in Figure 2.3 the sinor is a rotating vector on a circle with radius A. It's real part corresponds to a cosine wave and its imaginary part corresponds to a sine wave. This complicated construction seems over the top, but it is essential in Fourier Theory and in telecommunications in which quadrature modulation has become one of the mainstream modulation techniques.

Interpretation

The most straightforward way to interpret a sinor, is to consider it together with a sinor with the same amplitude, but the opposite (negative) frequency. The result of this sum is a cosine as can be seen in Figure 2.4.

This effectively introduces the concept of negative frequencies in our *mathematical imagination*.[1]

[1] Try explaining to your non-engineering friends the concept of frequency, e.g. something happening 4 times per second; then try to explain them that something might also happen minus 4 times per second. Good luck!

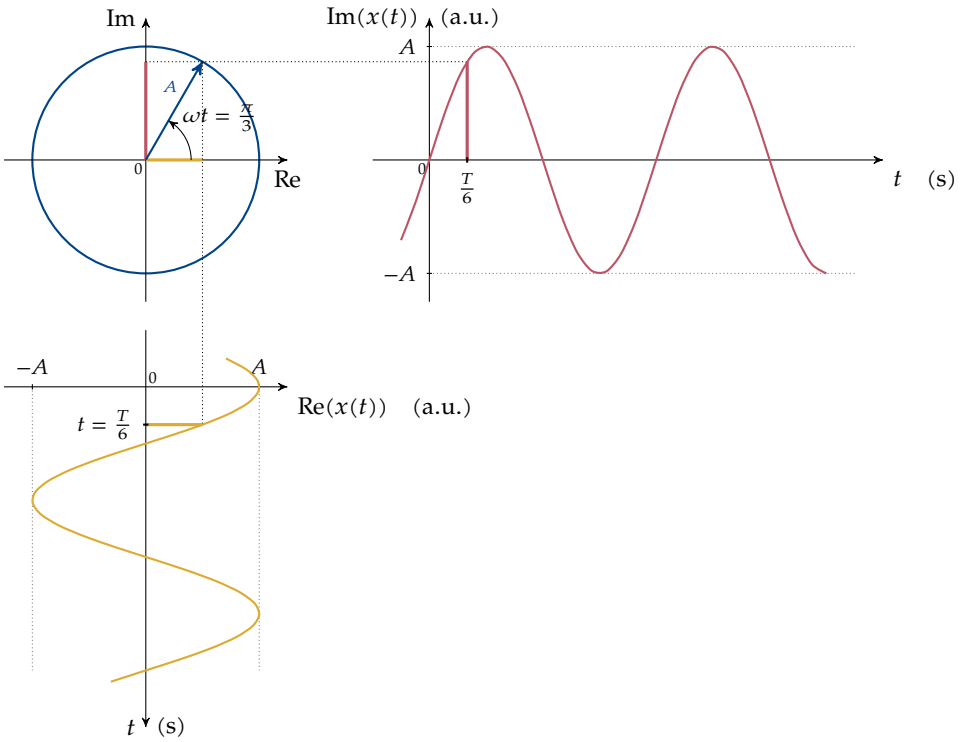


Figure 2.3: Sinor for $t = T/6$

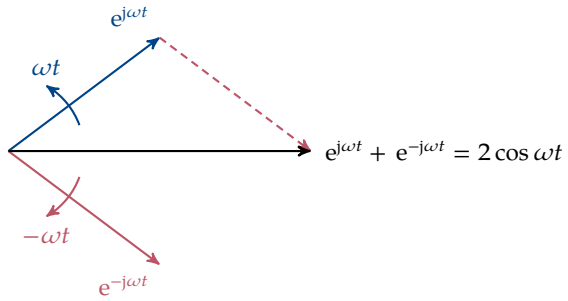


Figure 2.4: Sinors with opposite frequencies add up to a cosine of twice the amplitude.

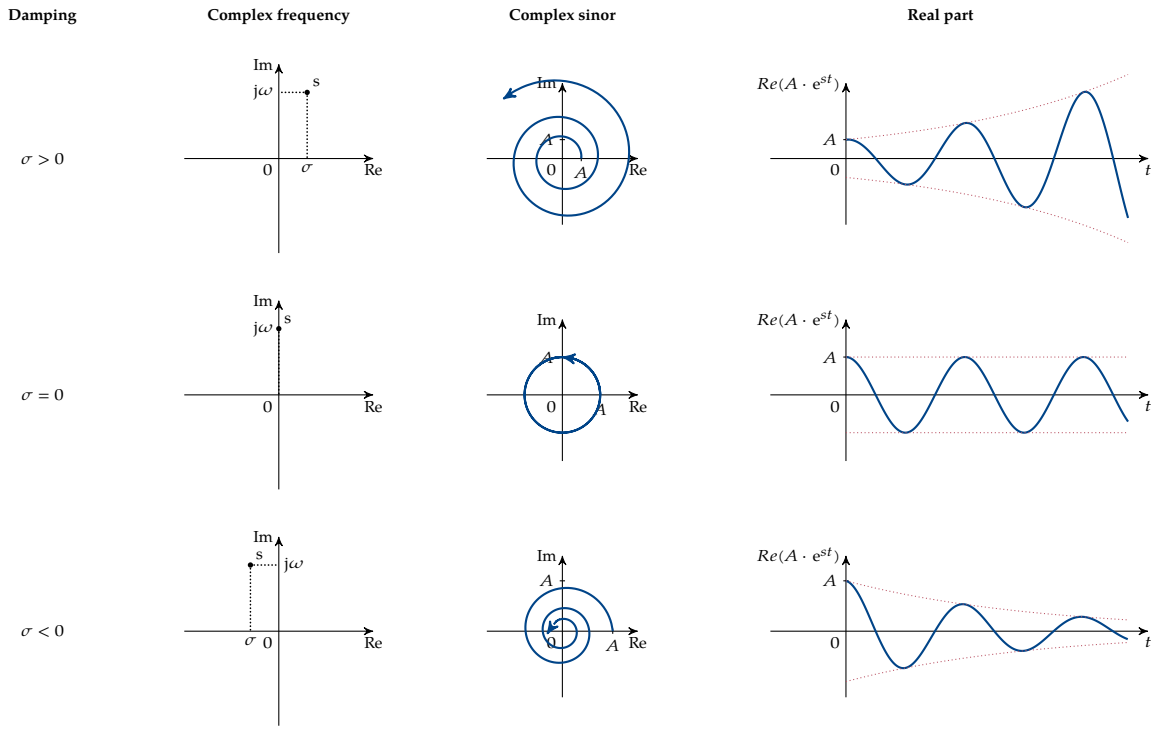


Figure 2.5: Generalized sinors for some typical values of s

Generalized sinors

We will see later, that a variant of the 'simple' sinor concept has some useful application: in the *generalized sinor* a damping factor is added:

$$\begin{aligned} x(t) &= A \cdot \underbrace{\mathrm{e}^{\sigma t}}_{\text{damping factor}} \cdot \mathrm{e}^{\mathrm{j}\omega t} \\ &= A \cdot \mathrm{e}^{(\sigma+\mathrm{j}\omega)t} \\ &\downarrow\ s = \sigma + \mathrm{j}\omega \\ &= A \cdot \mathrm{e}^{st} \end{aligned}$$

The variable s is a complex frequency variable. Figure 2.5 shows the relationship between the real part of the complex frequency and the behavior in the sinor domain and the real domain. Note that the value of the pulsation ω just controls the rotational speed the sinor revolves at. As one can see a value of $\mathrm{Re}(s) > 0$ results in an undamped (unstable) sinor and a value of $\mathrm{Re}(s) < 0$ results in a damped (stable) sinor.

Phasors

We rarely use phasors in systems theory. However, they are regularly used when studying AC power systems. The only reason we mention them here is to avoid confusion. They have little in common with sinors. The idea is that we represent sine waves exhibiting the same frequency and amplitude as vectors in which the angle under which we draw them corresponds to their phase difference. These vectors do not change their position over time (as opposed to sinors).

Purely as example, Figure 2.6 illustrates a phasor diagram for a 3N400 AC power supply.

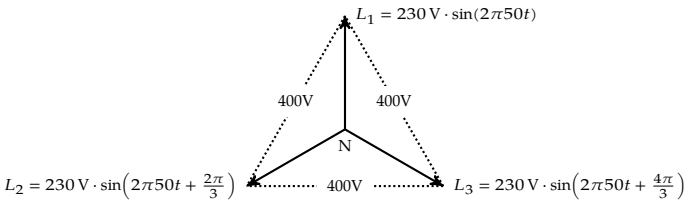


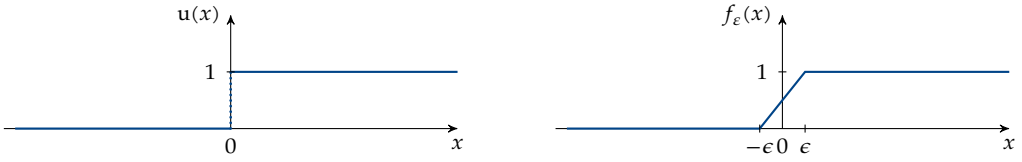
Figure 2.6: Phasor diagram of a 3N400 AC power system

Dirac impulses

The starting point is the unit step (or Heaviside function).

$$\mathrm{u}(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$$

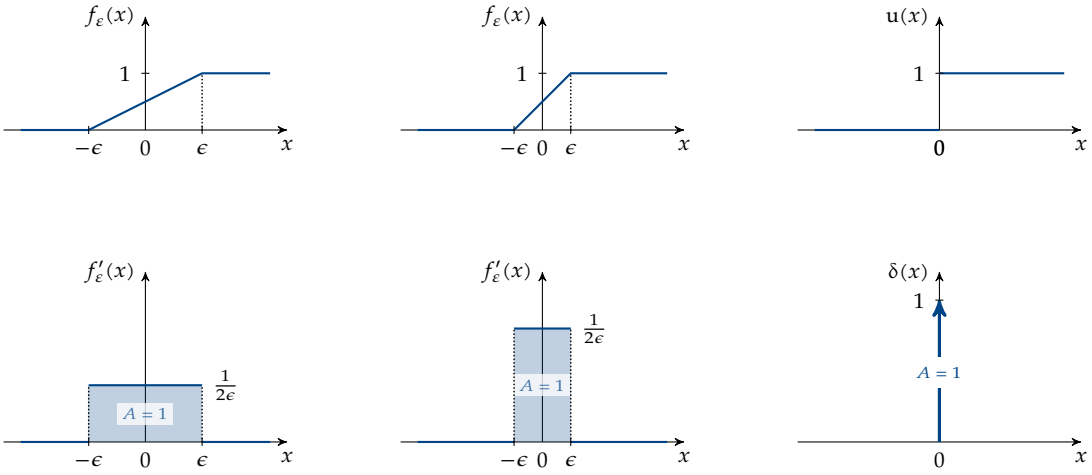
We use x as independent variable here, because we want to stress that it is a dimensionless variable, instead of t which has dimension time. We will see later that using t instead of x will complicate things a little bit. However, let's first focus on understanding this step function with x as independent variable. The problem is that the slope in the jump from 0 to 1 is infinite (as can been seen in Figure 2.7a), i.e. physically not possible. Let's therefore assume that the slope is finite as in Figure 2.7b.



(a) with infinite slope **(b)** with finite slope

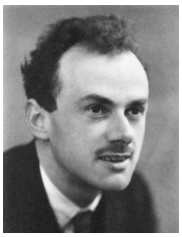
Figure 2.7: The Heaviside step function

Now consider this heaviside function and its derivative for increasing values of the slope $1/2\epsilon$:



It can be observed that when $\epsilon \to 0$, the rectangle formed by the graph of the derivative, becomes narrower and taller, keeping the same area $A = 1$. In the limiting case, the rectangle becomes infinitely narrow and infinitely tall. We represent it as an arrow, of which the length indicates the area that it covers.

(a) Oliver Heaviside (*1850 – †1925)



(b) Paul Dirac (*1902 – †1984)

(c) Laurent Schwartz (*1915 – †2002)

Figure 2.8: The impulse triumvirate

Mathematically this transition leads to the following definition:

Dirac impulse

The Dirac impulse $\delta(x)$ can be defined as:

$$\delta(x) = \begin{cases} 0 & \text{if } x \neq 0 \\ +\infty & \text{if } x = 0 \end{cases} \qquad \text{and} \int_{-\infty}^{+\infty} \delta(x)\ \mathrm{d}x = 1$$

Strictly speaking, the Dirac impulse is no function, as functions cannot have points in their domain that have values equal to infinity. In fact the Dirac impulse only makes sense when it appears in integrals. In that case the sifting property applies in case $f(t)$ is continuous around a:

$$\int_{-\infty}^{+\infty} f(t)\,\delta(t-a)\,\mathrm{d}t = f(a)$$

However, as stubborn engineers, we will keep dealing with the Dirac impulse as if it were a regular function.

The Dirac impulse is named in tribute of Paul Dirac, one of the founding father of *quantum physics*, who introduced the 'delta function' in physics. It must be said that the concept of a unit impulse was used much earlier by Oliver Heaviside. However, its foundation was disputed by fundamental mathematicians until Laurent Schwartz developed the theory of distributions, that makes the Dirac impulse a rock solid concept in mathematics, physics and engineering. It must be said that on more than one occasion, it takes an engineer, a physicist and a mathematician — not only to start off a good joke, but also — to make truly great breakthroughs. The contribution of this trio is of such a high level, that they deserve their photo in this course (see Figure 2.8).

Remark

Mathematical functions require their arguments to be dimensionless, i.e. in a sinewave $\sin(x)$, the Heaviside function $\mathrm{u}(x)$, or the Dirac impulse $\delta(x)$, we assume the x to be a real number without units. For a sine wave, this makes perfect sense. When we write:

$$v(t) = 230\,\mathrm{V} \cdot \sin(2\pi \cdot 50\,\mathrm{Hz} \cdot t)$$

we know that within the sine's argument the unit Hertz and the unit of the time variable, seconds, multiply to be 1 (without remaining unit). We are even so used to that, that we often write the above without the units:

$$v(t) = 230 \cdot \sin(2\pi 50 \cdot t)$$

However, this also means that if we write $\mathrm{u}(t)$ or $\delta(t)$, where t has the unit of time, we assume that in

fact this means:

$$\mathrm{u}\left(\frac{1}{\mathrm{s}}\cdot t\right) \qquad \text{and} \qquad \delta\left(\frac{1}{\mathrm{s}}\cdot t\right)$$

We will never write these units, but we must imply they are there. This becomes important, if we calculate a derivative of such a function, as we need to apply the chain rule. As an example, consider deriving the Heaviside function:

$$\frac{\mathrm{d}\,\mathrm{u}(t)}{\mathrm{d}t} = \delta(t)\cdot\frac{\mathrm{d}\frac{1}{\mathrm{s}\cdot t}}{\mathrm{d}t} = \delta(t)\cdot\frac{1}{\mathrm{s}}$$

This means that we have the following relationship between the Heaviside function and its derivative (first line) and the Dirac function and its primitive (second line):

$$\mathrm{u}(t) \xrightarrow{\frac{\mathrm{d}}{\mathrm{d}t}} \frac{1}{\mathrm{s}}\cdot\delta(t)$$
$$\delta(t) \xrightarrow{\int\cdot\,\mathrm{d}t} \mathrm{s}\cdot\mathrm{u}(t) + C$$

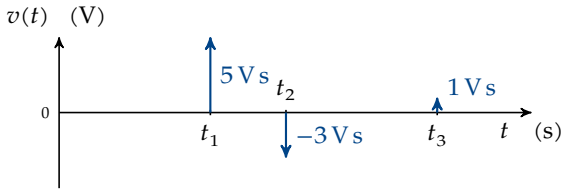
with C a constant. Specifically, this implies that the area under a Dirac impulse with a coefficient of 5 volt,

$$v(t) = 5\,\mathrm{V}\cdot\delta(t-t_0)$$

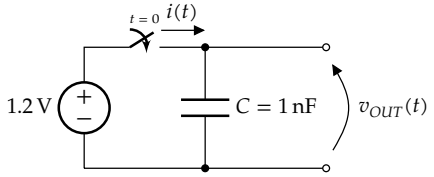
amounts to (try calculating it yourself!):

$$\int_{-\infty}^{+\infty} v(t)\,\mathrm{d}t = 5\,\mathrm{V\,s}$$

We often write the area contained in a Dirac impulse next to the arrow in a graph. E.g. for $v(t) = 5\,\mathrm{V}\cdot\delta(t-t_1) - 3\,\mathrm{V}\cdot\delta(t-t_2) + 1\,\mathrm{V}\cdot\delta(t-t_3)$:



Let's illustrate the combo of the Heaviside function and the Dirac impulse in an example. Consider the following electrical network of which we'd like to calculate the current $i(t)$:



The voltage applied to the capacitor C can be described as:

$$v_{OUT} = 1.2\,\mathrm{V}\cdot\mathrm{u}(t)$$

And therefore, we can calculate the current to be:

$$\begin{aligned}
i(t) &= C\frac{\mathrm{d}v_{OUT}}{\mathrm{d}t}\\
&= 1\times10^{-9}\,\mathrm{F}\cdot\frac{\mathrm{d}(1.2\,\mathrm{V}\cdot\mathrm{u}(t))}{\mathrm{d}t}\\
&\downarrow\quad \text{In fact: } \frac{\mathrm{d}\,\mathrm{u}(t)}{\mathrm{d}t} \equiv \frac{\mathrm{d}\,\mathrm{u}\left(\frac{t}{\mathrm{s}}\right)}{\mathrm{d}t} \overset{\text{chain rule}}{=} \delta(1/\mathrm{s}\cdot t)\cdot 1/\mathrm{s} \equiv \frac{\delta(t)}{\mathrm{s}}\\
&= 1\times10^{-9}\,\mathrm{F}\cdot 1.2\,\mathrm{V}\cdot\frac{\delta(t)}{\mathrm{s}}\\
&= 1.2\times10^{-9}\cdot\delta(t)\cdot\mathrm{F\,V/s} = 1.2\,\mathrm{nA}\cdot\delta(t)
\end{aligned}$$

Exercises

Exercise 2.4.2-1:
Draw the following sinors

- $5 \cdot \mathrm{e}^{(2+3\mathrm{j})t}$
- $3 \cdot \mathrm{e}^{(-2+\mathrm{j})t}$
- e^{t}
- $-2 \cdot \mathrm{e}^{-2\mathrm{j}t}$

in the following domains:

- in the complex frequency domain,
- in the sinor domain,
- in the time domain

and calculate their value for $t = 2$.

Exercise 2.4.2-2:
Draw the following sinor: $x(t) = -2 \cdot e^{1.3t} \cdot e^{-\mathrm{j}10t}$

- Draw the complex frequency argument in the complex plane.
- Draw the rotating vector trajectory as a function of time.
- Draw the real value of the sinor as a function of time.
- Is the sinor decaying or increasing over time?

2.5 Conclusion

We have defined a signal as a mapping of a quantity to another. It is convenient to see signals as functions of an independent variable. All function operations can therefore be seen as signal processing operations. In the category of physical signals, we have seen multiple periodic examples, as they are frequently used in testing. As for the nonphysical signals such as sinors and Dirac impulses, they will prove their good use in the next chapters.

Chapter 3

Systems

In this chapter, you will

- learn what systems are,
- learn which different classes of systems we can discern,
- get to know our love for linear time-invariant systems,
- learn how to describe these.

After having studied this chapter, you should be able to:

- describe what a system is,
- classify systems and understand the implications of that,
- define LTI systems, know their basic properties, how to represent them,
- convert one representation into another.

3.1 Definition of a system

We can use the following definition:

> **System**
>
> A system is a well-defined set of processes that interact in a specific context and as such show some specific behavior (partly in observable, partly in non-observable signals) that may be subject to some influence we may exert on it, via input signals.

Note that we have seen this definition earlier in Chapter 1, but then still used the description quantities for the concept of a signal.

Stated in a more simple way: a system is a 'block' that has a certain behavior that relates its input signals $x(t)$ to the output signals $y(t)$. Not all signals in the block may be observable. Not every detail about the block may be known.

We often use a simple mathematical notation to represent a system with some input signals $x(t)$ and some output signals $y(t)$:

$$x(t) \xrightarrow{H} y(t)$$

The capital letter H represents the system.

In the real world, such a system will be described by a (set of) differential equation(s). The coefficients that appear in these equations are the so called system or model parameters.

Remember the skydiver model that we used in Chapter 1. This model was a first-order differential equation with as parameters m, b:

$$\frac{m}{b}\frac{\mathrm{d}v}{\mathrm{d}t} + v = \frac{mg}{b}$$

We could consider g to be a model parameter, but let's not for now. Let's assume it is a context constant (related to the fact that 'earth' is the context). One could argue that g needs to be a signal that is dependent on the altitude and the location on earth, but let's keep the view of the world simple for now.

3.2 Classification

We can distinguish many types of systems. The system type will also determine how we will approach such a system from a mathematical point of view.

3.2.1 Time invariant vs. time variant systems

Sloppy definition

If the model parameters are not a function of time, we label the system as time invariant. If the parameters are a function of time, the system is time variant.

If we take a look back at the example of the skydiver we used earlier, then we can conclude that this model was time variant. Indeed, at first the parachute was not opened, leading to a low drag coefficient b. That parameter increased abruptly as soon our skydiver opened her parachute.

The definition of time invariance given above may seem crisp, but it isn't. The problem lies in the fact that we may not know every detail of the system. How can we check the constantness of parameter, if we even don't know the part of the model in which the parameter appears?

To this end, we need a definition that is independent of the (partially unknown) model.

> Time invariant system
>
> A time invariant system is a system that independently of time shows the same response to an excitation.

Crisp definition

Two new words were used in this definition:

excitation this is a set of input signals we present to the system

response these are the output signals that we can observe from the system

Mathematically we can write this definition down in a very concise manner. Assuming the system is defined as:

$$x(t) \xrightarrow{H} y(t)$$

This means that exciting the system with a specific input signal $x(t)$ will result in a specific output signal $y(t)$.

Conclusion: a system is time invariant if and only if

$$\forall \tau \in \mathbb{R}, \forall x(t) \in \mathbb{H}: \quad x(t) \xrightarrow{H} y(t) \quad \Rightarrow \quad x(t-\tau) \xrightarrow{H} y(t-\tau)$$

For now, you can think of $\mathbb{H}$ as the set of all possible signals. We call that set the *Hilbert space*.

3.2.2 Linear vs. nonlinear systems

Sloppy definition

A system is linear if its model is linear in the signal variables.

If we take a look back at our skydiver model, the equation was linear in the signal variable v.

$$m\frac{\mathrm{d}v}{\mathrm{d}t} + bv = mg$$

If we would have used a different drag model (e.g. drag proportional to the square of the variable v, as in

$$m\frac{\mathrm{d}v}{\mathrm{d}t} + bv^2 = mg$$

then the model would no longer be linear and therefore the system is no longer linear.

Again, the same problem arises with this simple definition: how can we use a definition that relies on knowing the entire model of the system? To solve this issue, we need a definition that is independent of the model, and for which we can only rely on the input and output signals.

Linear system

A system is linear if and only if a linear combination of specific excitations results in the same linear combination of the specific responses.

Crisp definition

The rigorous mathematical definition is again very compact. A system is linear if and only if

$$\forall a, b \in \mathbb{R}, \forall x_1(t), x_2(t) \in \mathbb{H}:$$
$$x_1(t) \xrightarrow{H} y_1(t) \text{ and } x_2(t) \xrightarrow{H} y_2(t) \quad \Rightarrow \quad ax_1(t) + bx_2(t) \xrightarrow{H} ay_1(t) + by_2(t)$$

We can decompose this property into two subproperties that are a consequence of this:

- **homogeneity**, i.e. a scaled input results in a scaled output:

$$ax_1(t) \xrightarrow{H} ay_1(t)$$
$$bx_2(t) \xrightarrow{H} by_2(t)$$

  for any real a and b.

- **superposition**, i.e. a sum of two inputs results in the sum of the two corresponding outputs:

$$x_1(t) + x_2(t) \xrightarrow{H} y_1(t) + y_2(t)$$

Remarks

- We will not make rigorous exercises on determining time invariance or linearity in this course. We will keep this for a later course (on Electronic Systems) that is part of your master's program.
- That said, please don't study the crisp definitions by heart. For now it is sufficient that you know the sloppy definitions and realize that they are sloppy, but nonetheless a good start.

Note on linearity

Nonlinear systems can be further subdivided into easy and not-so-easy systems. To put things in perspective, we compare three different cases:

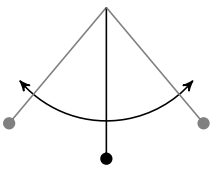
1. a pendulum in which the deviation from the resting position is small: this is linear
2. a pendulum in which the deviation from the resting position is considerable: this is nonlinear
3. a double pendulum: this is a special case of a nonlinear system that we call a *chaotic system*

Our goal is not to study the three cases in depth, only to compare some basic aspects of them, to help you appreciate linear systems.

Pendulum small deviation
as prototype of
a linear system:

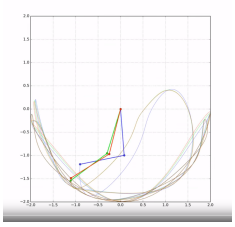
- easy to analyze (uniform theory)
- easy to predict
- engineering sweetspot

Pendulum large deviation
as prototype of
a nonlinear system:



- not so easy to analyze (no uniform theory)
- easy to predict evolution (in particular cases)
- still useful in engineering

Double pendulum
as prototype of
a chaotic system:



- special case of nonlinear system (with bifurcations[1]) = theoretical mess
- hard to predict evolution
- engineering nightmare

[1] A bifurcation (containing to the words 'bi' and 'fork') is behavior in which based on a small difference the state of the system takes a totally different route (two different teeth of the fork). Click the image above or view `https://www.youtube.com/embed/pEjZd-AvPco` to see an illustration of this.

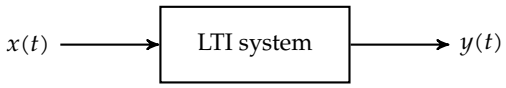
The take-home message for the special case of nonlinear systems with bifurcations, is that small deviations on the initial conditions can lead to huge differences in the behavior of the system. These systems are extremely difficult to predict. We want to keep away from them in mainstream engineering.

3.3 The engineering sweetspot: LTI systems

3.3.1 Definition

The types of systems we like very much in (systems) engineering are *linear time-invariant* (*LTI*) *systems*. The obvious reasons are: they keep their behavior unaltered over time and they are linear which will prove to be a great advantage when analyzing, designing, testing and using these systems.

Let's assume we have an LTI system as depicted below:



Mathematically we can write this down as:

$$x(t) \xrightarrow{H} y(t)$$

Let's assume we run two arbitrary experiments with this system, feeding a particular $x_1(t)$ and $x_2(t)$, resulting in an output $y_1(t)$ and $y_2(t)$ respectively.

The system is linear if for any real values of a_1 and a_2 it obeys:

$$a_1 x_1(t) + a_2 x_2(t) \xrightarrow{H} a_1 y_1(t) + a_2 y_2(t)$$

It is time-invariant if for any real value of τ and any $x_1(t) \xrightarrow{H} y_1(t)$, the following holds:

$$x_1(t-\tau) \xrightarrow{H} y_1(t-\tau)$$

We will be using that property to our benefit in the analysis of these systems.

3.3.2 Important properties

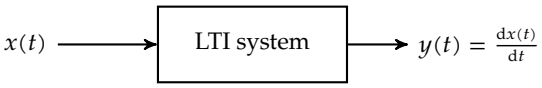
The two important properties that we want to discuss for now are related to differentiation:

- Differentiation is linear and time-invariant.
- LTI systems maintain differentiation from input to output.

Let's discuss these below.

Differentiation is linear and time-invariant

The property assumes that we take a system that is a differentiator, i.e.



or mathematically:

$$x(t) \xrightarrow{H} y(t) = \frac{\mathrm{d}x(t)}{\mathrm{d}t}$$

What we claim is that this system is linear and time invariant.

Proof

Obviously this system is linear, because assuming $x_1(t) \xrightarrow{H} y_1(t)$ and $x_2(t) \xrightarrow{H} y_2(t)$ and any real a and b, we can prove the following:

$$ax_1(t) + bx_2(t) \xrightarrow{H} y(t) = \frac{\mathrm{d}(ax_1(t) + bx_2(t))}{\mathrm{d}t} = a\frac{\mathrm{d}x_1(t)}{\mathrm{d}t} + b\frac{\mathrm{d}x_2(t)}{\mathrm{d}t} = ay_1(t) + by_2(t)$$

It is also time-invariant, as for any real τ we can prove that:

$$x(t-\tau) \xrightarrow{H} \frac{\mathrm{d}x(t-\tau)}{\mathrm{d}t} = \frac{\mathrm{d}x(u)}{\mathrm{d}u} \cdot \frac{\mathrm{d}u}{\mathrm{d}t} = \frac{\mathrm{d}x(u)}{\mathrm{d}u} \cdot 1 = \frac{\mathrm{d}x(u)}{\mathrm{d}u} = y(u) = y(t-\tau)$$

in which we used the substitution $u = x - \tau$ and the chain rule. ■

Don't focus on the proofs. Just retain the main message: "differentiation is LTI".

LTI systems maintain differentiation from input to output

This means that if we have the following system with a specific input x and output y

$$x(t) \xrightarrow{H} y(t)$$

then exciting the system with the derivative of the input, will result in a response equal to the derivative of the output:

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} \xrightarrow{H} \frac{\mathrm{d}y(t)}{\mathrm{d}t}$$

Proof

The proof is straightforward. The definition of differentiation states that:

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = \lim_{\Delta t \to 0} \frac{x(t+\Delta t) - x(t)}{\Delta t}$$
$$\frac{\mathrm{d}y(t)}{\mathrm{d}t} = \lim_{\Delta t \to 0} \frac{y(t+\Delta t) - y(t)}{\Delta t}$$

the expression behind the limit operation is a linear combination

$$\frac{1}{\Delta t} \cdot x(t+\Delta t) + \frac{-1}{\Delta t} \cdot x(t)$$

We know that because of time invariance, the following must hold for any real Δt:

$$x(t+\Delta t) \quad \xrightarrow{H} \quad y(t+\Delta t)$$

and therefore, because of linearity also the following must hold:

$$\frac{1}{\Delta t}\cdot x(t+\Delta t)+\frac{-1}{\Delta t}\cdot x(t) \quad \xrightarrow{H} \quad \frac{1}{\Delta t}\cdot y(t+\Delta t)+\frac{-1}{\Delta t}\cdot y(t)$$

Because this holds for any real Δt, it also must hold for $\Delta t \to 0$:

$$\begin{aligned} & \lim_{\Delta t\to 0}\left(\frac{1}{\Delta t}\cdot x(t+\Delta t)+\frac{-1}{\Delta t}\cdot x(t)\right) \quad \xrightarrow{H} \quad \lim_{\Delta t\to 0}\left(\frac{1}{\Delta t}\cdot y(t+\Delta t)+\frac{-1}{\Delta t}\cdot y(t)\right) \\ \Leftrightarrow \quad & \lim_{\Delta t\to 0}\frac{x(t+\Delta t)-x(t)}{\Delta t} \quad \xrightarrow{H} \quad \lim_{\Delta t\to 0}\frac{y(t+\Delta t)-y(t)}{\Delta t} \\ \Leftrightarrow \quad & \frac{\mathrm{d}x(t)}{\mathrm{d}t} \quad \xrightarrow{H} \quad \frac{\mathrm{d}y(t)}{\mathrm{d}t} \end{aligned}$$

Actually, this reasoning only holds if both $\frac{\mathrm{d}x(t)}{\mathrm{d}t}$ and $\frac{\mathrm{d}y(t)}{\mathrm{d}t}$ exist and are finite. We assume this is the case. ■

Again, don't focus on the proofs, they are there for your satisfaction and understanding. The main message is: "LTI systems maintain differentiation".

There are multiple ways to represent an LTI system. We'll discuss three of them right now:

- linear differential equations
- block diagrams
- impulse response description

3.3.3 Linear differential equations

In view of the fact that we know that differentiation is an LTI operation, it will not surprise you that linear differential equations with constant coefficients describe LTI systems, as all of the composing operations are linear and time invariant. In addition all parameters (the coefficients) are time invariant.

A linear differential equation with constant coefficients is an expression of the form:

$$a_n\frac{\mathrm{d}^n y}{\mathrm{d}t^n}+a_{n-1}\frac{\mathrm{d}^{n-1}y}{\mathrm{d}t^{n-1}}+\cdots+a_2\frac{\mathrm{d}^2 y}{\mathrm{d}t^2}+a_1\frac{\mathrm{d}y}{\mathrm{d}t}+a_0 y=b_0 x \tag{3.1}$$

with $a_i \in \mathbb{R}$ and $n \in \mathbb{N}$. We call n the order of the differential equation. Note that we didn't write the dependence on 't' of $x(t)$ and $y(t)$ to keep the equation as simple as can be. It will not surprise you that mathematicians like this simplicity and will almost always write a differential equation like this.

Any pair $x(t)$, $y(t)$ that fulfills this equation is called a *solution* of this equation, and therefore comprises a valid input-output pair for the system described by the differential equation. Therefore in fact, the following line

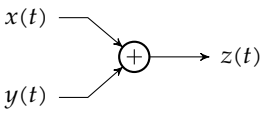
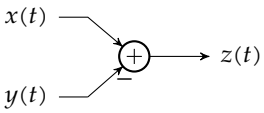
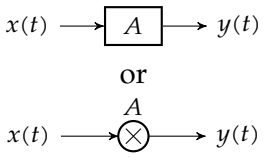
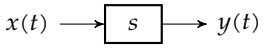
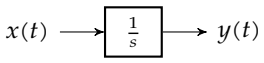
$$x(t) \xrightarrow{H} y(t)$$

means: "the input-output pair $(x(t), y(t))$ is a solution of the differential equation (3.1)".

3.3.4 Block diagrams

Symbols used in a block diagram

Most often, we will use a graphical representation of the system, using the following symbols for the basic operations:

Operation	Expression	Symbol
Addition	$z(t) = x(t) + y(t)$	
Subtraction	$z(t) = x(t) - y(t)$	
Scaling	$y(t) = A \cdot x(t)$	
Differentiation	$y(t) = \frac{\mathrm{d}x(t)}{\mathrm{d}t}$	
Integration	$y(t) = \int_0^t x(\tau)\mathrm{d}\tau$	

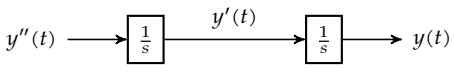
Note that though the table mentions differentiation, we will try to avoid that in a real world system.[2] The fact that the Laplace transform maps differentiation to multiplication with s, and integration to division by s explains the symbols for derivation and integration.

Converting differential equations to block diagrams

If we start from a differential equation, our goal is to end up with an *all-integrator block diagram*, i.e. a block diagram that contains no differentiation blocks, only integration blocks. To this end, you first need to solve the differential equation for the highest derivative of y:

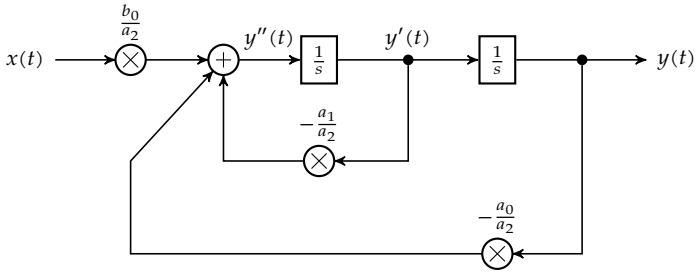
$$\begin{aligned} & a_2 y'' + a_1 y' + a_0 y = b_0 x \\ & \qquad\qquad \downarrow \text{ solve for highest derivative of } y \\ \Leftrightarrow \quad & y'' = \frac{b_0}{a_2} x - \frac{a_1}{a_2} y' - \frac{a_0}{a_2} y \end{aligned} \tag{3.2}$$

The next element of the solution is the knowledge that we can build an integrator chain that joins y with its two derivatives:



[2] The output of a differentiation will be very large when the input changes suddenly, which is the case when noise is present in the system. This will lead to huge values that may cause the system to go outside of its zone of linear operation.

Combining (3.2) with this delay chain, allows us to compose the following block diagram:

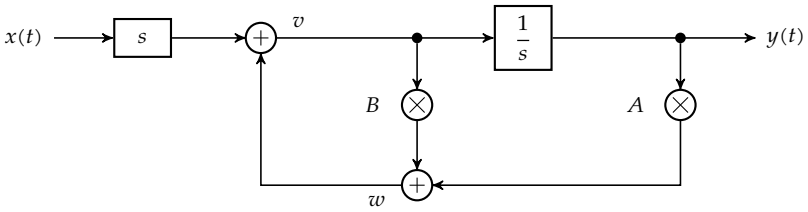


This works for any linear differential equation with constant coefficients.

Converting block diagrams to differential equations

The other way around is just a matter of writing down the equations for every block in the block diagram and eliminating any (unnecessary) intermediate variables.

A *first example*, starting from an arbitrary block diagram:



Note that this block diagram still contains a derivative, which we would like to avoid in practical systems. However, it is a valid system.

In the above diagram, we already labeled the intermediate signals. This allows to write down very conveniently the following equations:

$$\begin{cases} v = \dfrac{\mathrm{d}x}{\mathrm{d}t} + w \\ w = Bv + Ay \\ y = \displaystyle\int_0^t v \,\mathrm{d}t \end{cases}$$

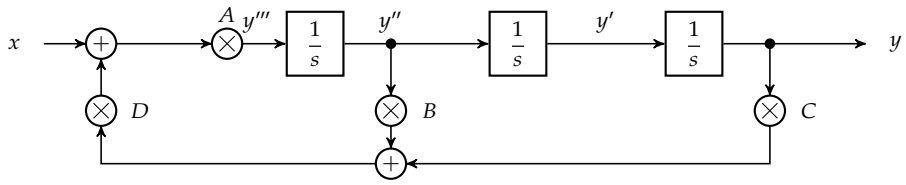
Out of these equations, we can eliminate w:

$$\begin{cases} v = \dfrac{\mathrm{d}x}{\mathrm{d}t} + Bv + Ay & (3.3) \\ y = \displaystyle\int_0^t v \,\mathrm{d}t & (3.4) \end{cases}$$

As (3.4) implies that $v = \frac{\mathrm{d}y}{\mathrm{d}t}$, we can substitute v by this value in (3.3), which leads to:

$$(1 - B)\frac{\mathrm{d}y}{\mathrm{d}t} - Ay = \frac{\mathrm{d}x}{\mathrm{d}t}$$

A *second example*, starting from an all integrator schematic:



Knowing the procedure to convert a differential equation to a block diagram, we can write with a single glimpse of the eye:

$$y''' = A\left(x + D\left(By'' + Cy\right)\right)$$
$$\Leftrightarrow \quad \frac{1}{A}y''' - DBy'' - DCy = x$$

This is a third-order linear differential equation with constant coefficients.

Exercises

Exercise 3.3.4-1:
Draw a block diagram for the system described by the following differential equation:

$$5\frac{\mathrm{d}^3y}{\mathrm{d}t^3} - 2\frac{\mathrm{d}y}{\mathrm{d}t} = y - x$$

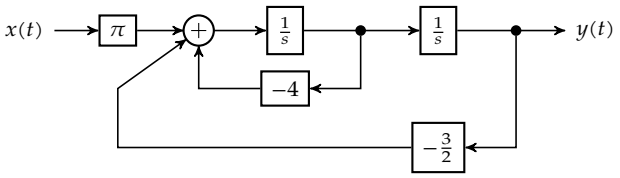
Exercise 3.3.4-2:
Draw a block diagram for the system described by the following differential equation:

$$3\frac{\mathrm{d}^2y}{\mathrm{d}t^2} + 9\frac{\mathrm{d}y}{\mathrm{d}t} - y = 0$$

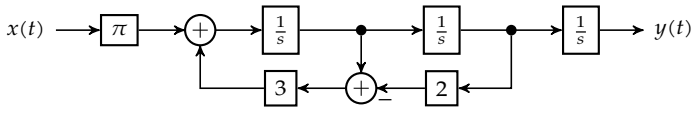
Exercise 3.3.4-3:
Draw a block diagram for the system described by the following differential equation:

$$\frac{\mathrm{d}^2y}{\mathrm{d}t^2} + \frac{\mathrm{d}y}{\mathrm{d}t} = x$$

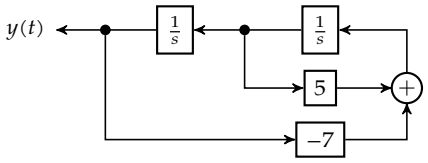
Exercise 3.3.4-4:
Compose the differential equation that describes the block diagram below:



Exercise 3.3.4-5:
Compose the differential equation that describes the block diagram below:



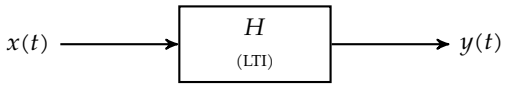
Exercise 3.3.4-6:
Compose the differential equation that describes the block diagram below:



3.3.5 Impulse response description

3.3.5.1 Definition

Consider the LTI system below:



In symbols: $x(t) \xrightarrow{H} y(t)$. Let's assume we submit a Dirac impulse $\delta(t)$ to it. We will call the corresponding output signal the *impulse response* and denote it as $h(t)$. Mathematically:

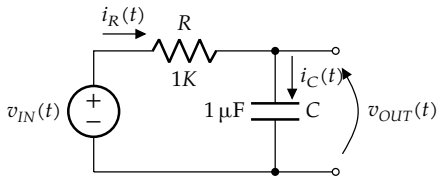
$$\delta(t) \xrightarrow{H} h(t)$$

Though the experiment of exciting the system with a Dirac impulse and recording its output can be conducted for every possible system (not only LTI systems), it only makes sense to do for LTI systems, as for these systems the impulse response contains the entire heart and soul of the system. We will devote the entire Chapter 4 to it.

To be specific, this means that the impulse response carries all required information to fully define the system (and to calculate its output, given an input). Indeed, for an LTI system the impulse response is as good as the original differential equation describing the system. How do you obtain/calculate the impulse response? By 'kicking' the system with an impulse at the input and observing what comes out of it.

3.3.5.2 Converting differential equations to an impulse response

Let's see how we can determine the impulse response of a system. We'll do this using an example.
Consider the electrical circuit below, and let's try to calculate its impulse response, assuming v_{IN} is the input and v_{OUT} is the output.



To prepare ourselves, we first write Kirchhoff's current law (KCL) on the output node:

$$i_C = i_R$$

Then we use the branch equations

$$i_C = C\frac{\mathrm{d}v_{OUT}}{\mathrm{d}t}$$
$$i_R = \frac{v_{IN} - v_{OUT}}{R}$$

to eliminate i_C and i_R. This leads to:

$$\begin{aligned} & C\frac{\mathrm{d}v_{OUT}}{\mathrm{d}t} + \frac{1}{R}v_{OUT} = \frac{1}{R}v_{IN} \\ \Leftrightarrow \quad & \tau\frac{\mathrm{d}v_{OUT}}{\mathrm{d}t} + v_{OUT} = v_{IN} \quad \text{with} \quad \tau = RC = 1\,\mathrm{ms} \end{aligned} \tag{3.5}$$

We now have two options to calculate the impulse response of this system:

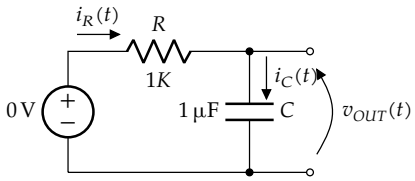
- **Direct method**: solve the differential equation with $v_{IN}(t) = \delta(t)$
- **Indirect method**: solve the differential equation with $v_{IN}(t) = \mathrm{u}(t)$ and then differentiate the output signal and multiply with the unit 'second'.

Note that the former can only be done in theory, i.e. through calculations, while the latter can be setup in an experiment, yielding the response to a step input that we can record during the experiment. A numerical differentiation of the recording can give us the impulse response.

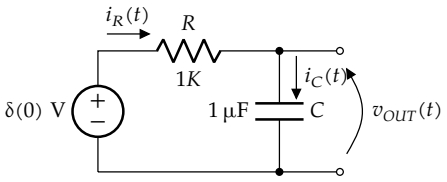
Direct method

The direct method applies a Dirac impulse directly to the circuit. We distinguish three phases in the Dirac impulse:

1. $t < 0 \Rightarrow v_{IN}(t) = 0$:

   This has been the situation since $t = -\infty$, therefore, we may safely assume that any charge that could have been on the capacitor, has leaked away by now (the input source is a short circuit after all).
   Conclusion $i(t) = 0$ and $v_{OUT}(t) = 0$.
2. $t = 0$: the input becomes infinite for an infinitely short period



We assume that the energy the source can deliver with its Dirac impulse is limited (as the area below the Dirac impulse is limited). Therefore, the voltage that can be present on the capacitor is limited, i.e. the full strength of the infinite Dirac impulse is over the resistor. Therefore, the current at $t = 0$ amounts to:

$$i(0) = \frac{v_{IN}(0)}{R} = \frac{\delta(0)\ \text{V}}{1\,\text{k}\Omega} = \delta(0)\ \text{mA}$$

The capacitor integrates this current as charge:

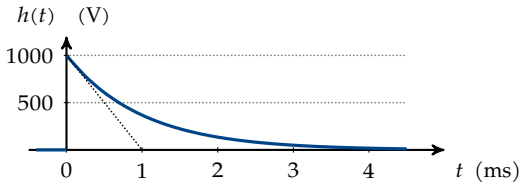
$$q = \int_{0^-}^{0^+} \delta(t)\ \text{mA}\,\text{d}t = 1\,\text{mA}\,\text{s} = 1\,\text{mC}$$

This charge corresponds to a voltage of:

$$v_{OUT}(0) = \frac{q}{C} = \frac{1\,\text{mC}}{1\,\mu\text{F}} = 1\,000\,\text{V}$$

3. $t > 0 \Rightarrow v_{IN}(t) = 0$
   The input voltage source again acts as a short circuit, allowing the charge on C to leak away through R. This process can be described using the following differential equation:
   $$\tau\frac{\text{d}v_{OUT}}{\text{d}t} + v_{OUT} = 0\,\text{V} \text{ with } v_{OUT}(0) = 1\,000\,\text{V}$$
   The solution of this initial value problem can be determined to be:
   $$v_{OUT}(t) = 1\,000\,\text{V} \cdot \text{e}^{-\frac{t}{\tau}}$$

This leads to the following conclusion for the impulse response:



Impulse response:
$h(t) = 1\,000\,\text{V}\ \text{e}^{-\frac{t}{\tau}}$

For the curious reader: try to make a graph of the current $i(t)$!

Indirect method

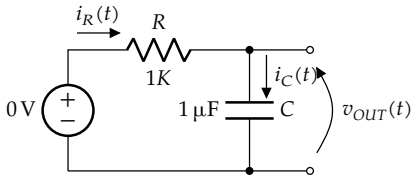
The indirect method is based on the property that LTI systems preserve derivatives from input to output.

$$\begin{aligned}
\text{u}(t) &\xrightarrow{H} v(t) \\
\frac{\text{d}\,\text{u}(t)}{\text{d}t} &\xrightarrow{H} \frac{\text{d}v(t)}{\text{d}t} \\
\frac{1}{\text{s}}\,\delta(t) &\xrightarrow{H} \frac{\text{d}v(t)}{\text{d}t} \\
\delta(t) &\xrightarrow{H} \text{s} \cdot \frac{\text{d}v(t)}{\text{d}t}
\end{aligned}$$

So, we will first calculate the step response $v(t)$, then differentiate it and multiply it with unit s to obtain the impulse response.

A step input consists of two constant values:

1. $t < 0 \Rightarrow v_{IN}(t) = 0$:



   This has been the situation since $t = -\infty$, therefore, we may safely assume that any charge that could have been on the capacitor, has leaked away by now (the input source is a short circuit after all). Conclusion $i(t) = 0$ and $v_{OUT}(t) = 0$.

2. $t \geq 0 \Rightarrow v_{IN}(t) = 1\,\mathrm{V}$
   This means that the behavior of the output voltage can now be described by (3.5) with $v_{IN} = 1\,\mathrm{V}$:
   $$\tau \frac{\mathrm{d}v_{OUT}}{\mathrm{d}t} + v_{OUT} = 1\,\mathrm{V} \text{ with } v_{OUT}(0) = 0\,\mathrm{V}$$
   Solving this initial value problem, yields:
   $$v_{OUT}(t) = \left(1 - \mathrm{e}^{-\frac{t}{\tau}}\right) \cdot \mathrm{V}$$
   And therefore:
   $$h(t) = \frac{\mathrm{d}v_{OUT}(t)}{\mathrm{d}t}\mathrm{s} = \frac{1}{\tau}\,\mathrm{e}^{-\frac{t}{\tau}} \cdot \mathrm{V\,s}$$

This is the same result as the one we obtained using the direct method.

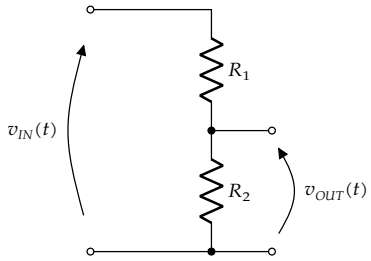
Now, let's take a step back. These calculations are good fun, but admittedly, rather complicated in order to assess something as simple as an impulse response. Especially, if your circuit contains more than a resistor and a capacitor (e.g., a full blown operational amplifier containing 10 to 20 transistors), this kind of reasoning becomes rather involved.

Luckily, we have a number of new tricks on our sleeve, i.e. transforming the problem to another descriptive domain, such that its solution becomes way more easy. The Laplace transform is such a trick.

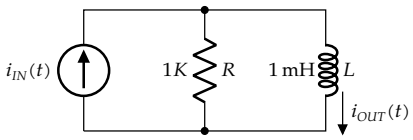
Exercises

Exercise 3.3.5.2-1:
Consider a system $x(t) \xrightarrow{H} y(t)$.
We know its step response, i.e. $\mathrm{u}(t) \xrightarrow{H} \sin(20t)$.
Determine its impulse response.

Exercise 3.3.5.2-2:
Determine the impulse response of the following circuit from input v_{IN} to output v_{OUT}.



Exercise 3.3.5.2-3:
Determine the impulse response of the following circuit from input i_{IN} to output i_{OUT}.

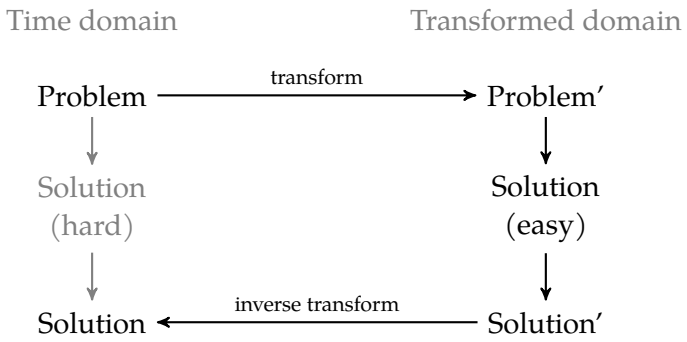


3.3.5.3 Converting the impulse response into differential equations

This isn't such an easy topic. It is a subject of the scientific subdomain of *system identification*, that we will not cover in this book. However, it is possible, under the assumption the system is linear, to determine the set of differential equations that correspond with the impulse response of a system. Both time-domain and frequency-domain techniques can be used to this end.

3.3.6 Transformations

The basis idea of applying transformations to simplify the problem is depicted in the following schematic drawing:



Instead of solving the problem directly (which is hard, as we have seen in the previous section), we will first transform the problem to a different descriptive domain (e.g., the complex frequency domain), then solve it in that domain, to inversely transform the obtained solution back to the original domain. Though this route involves a forward and a backward transform, it will prove to be much simpler than the original direct route.

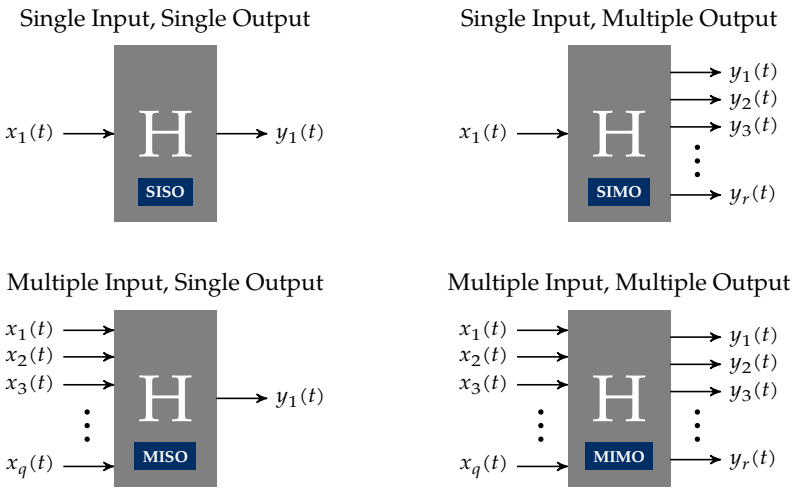
3.4 Counting the inputs and the outputs

Systems are often classified according to the number of inputs and outputs they have. Based on this criterion, we get the following system classes:

- SISO: single input, single output system
- SIMO: single input, multiple output system
- MISO: multiple input, single output system
- MIMO: multiple input, multiple output system

The relevance of these classes is that we can study any MIMO system by analyzing the behavior of the outputs one by one. Therefore, we only need to study MISO systems. The same holds for any SIMO system. We can study that by focusing on one output at the time. Therefore studying SISO systems is as far as we go.

We will also see that if a system is LTI, that you can study any MISO system as a sum of SISO systems. Therefore, for LTI systems we can focus on the techniques to analyze SISO systems.

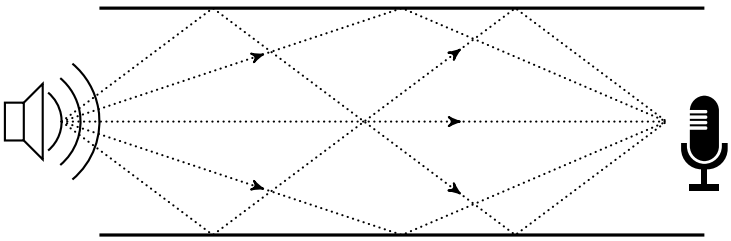


We will give some simple examples to allow you to further appreciate the nature of these four categories.

3.4.1 SISO

An example of such a system is a corridor with a speaker that emits sound waves and a single microphone that records them. The corridor itself can be seen as a SISO system, with a single input (the speaker) and a single output (the microphone).

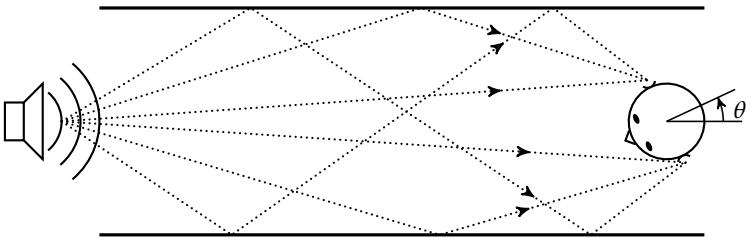
Important parameters of the system are the width of the corridor, the distance between speaker and microphone, their positions in the corridor, the medium through which sounds propagates, the reflectivity of the walls, the emitted power and its emission pattern and the sensitivity of the microphone and its directivity.



3.4.2 SIMO

The same corridor in which we replace the single microphone by a person with binaural hearing capabilities turns it into a system with a single input (the speaker) and two outputs (the two ears of the person).

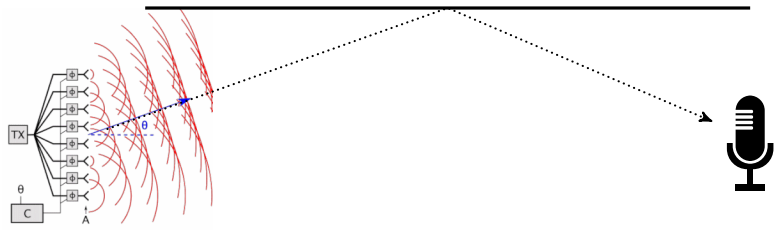
Additional important parameters are the azimuth angle of the head and the shape of the head and torso of the person.



3.4.3 MISO

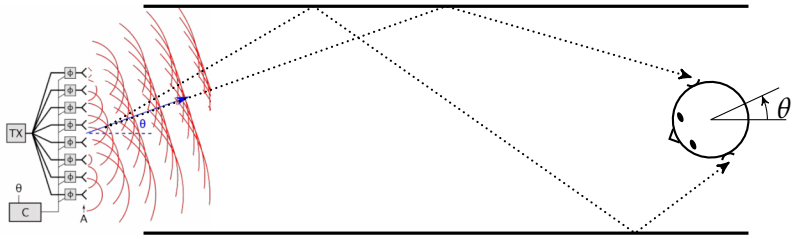
Now assume that we don't have a single speaker, but a linear array of speakers and a single receiving microphone. The corridor now becomes a MISO system in which the effect of beam-steering can be modeled.

An important parameter in this case is the delay between the speakers that will determine the beam-steering angle.



3.4.4 MIMO

Combining this idea of a linear array with the binaural listener, transforms the corridor into a MIMO system. The angle of the head is in this case an additional key parameter.



3.5 Conclusion

A system is a well-defined set of processes that we can influence with a number of input signals and observe through a number of output signals. It can be described using a mathematical model (often a set of differential equations). If the system is LTI, we can model it as a set of linear differential equations with constant coefficients or using its impulse response. We often draw a block diagram for such a system as it shows the flow of information (signals) in the model itself.

Chapter 4

Convolution

In this chapter, you will learn:

- how the convolution operation is based on an idea by Oliver Heaviside, the impulse decomposition,
- why it is related to LTI systems,
- how the operation is defined, and what its properties are, and
- how to use the operation in practice.

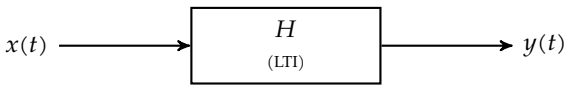
After having studied this chapter, you should be able to:

- calculate the output of an LTI system based on the convolution operation

4.1 Why convolution?

For any system, we face the typical engineering problem of calculating or predicting the output given a certain input. For most physical systems, this involves solving or simulating the model exciting it with the desired input signals. The model is often a set of differential equations.

In many cases we solve this model using transformations (as we will see later, when discussing the Fourier and the Laplace transform). In the case of LTI systems, we have an additional option, and that is to use its impulse response description. Consider as an example, the SISO LTI system H, as depicted in the following diagram:



In the previous chapter, we have seen that we can excite a system with a Dirac impulse, to have it produce its impulse response. This notion of the response to a Dirac impulse will lead to the possibility of calculating the output of the LTI system as the *convolution* of the impulse response with the input:

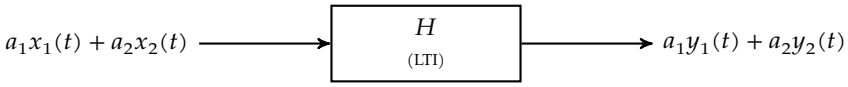
$$y(t) = h(t) \star x(t)$$

It is time to study this *convolution operation* in more detail. Note that in this book, we use the five-pointed filled star ($\star$) to denote this operation. We do this to avoid the confusion that may arise of using a traditional asterisk ($*$) as is done in many books.

4.2 Impulse decomposition

The convolution operation is based on the idea of impulse decomposition, an idea of Oliver Heaviside, one of the engineering champions of linear systems theory (see Figure 2.8a).

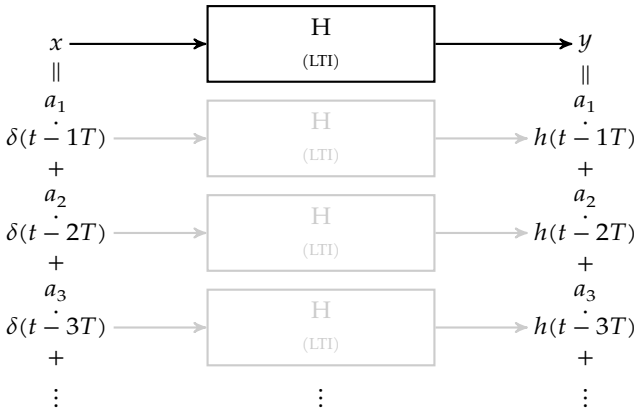
The starting point is the *linearity* of LTI systems, as depicted in the diagram below. Exciting the system with a linear combination of arbitrary but specific input signals x_1 and x_2 (with $x_1 \xrightarrow{H} y_1$ and $x_2 \xrightarrow{H} y_2$ results in an output that is exactly the same linear combination of the corresponding output signals y_1 and y_2. Graphically:



This opens the option of writing an arbitrary signal as a weighted sum of two or more basic signals, of which we hope it is more easy to calculate their corresponding output, then it is for the original signal.

The obvious question is: is it possible to find a set of basic input signals for which calculating their corresponding output signals is easy. Many attempts have been done, such as sine waves, complex exponentials or sinors. We will come back to these later, as they are the basis for the Fourier and the Laplace transform.

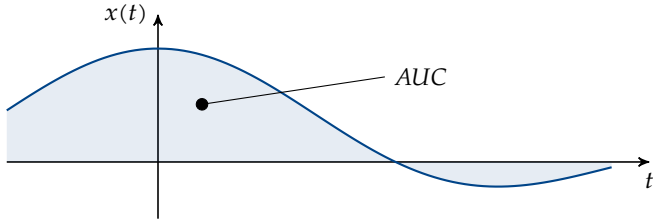
In view of this chapter, we focus on a specific one, and that is the impulse decomposition. The idea is that we decompose a signal as a sum of weighted and delayed Dirac impulses. Given the *time-invariant nature* of the system, the responses to the individual delayed Dirac impulses are delayed impulse responses. This has been illustrated in the figure below.



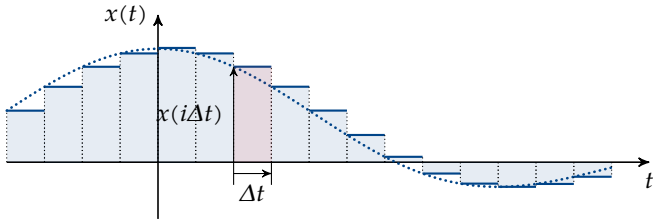
The key question is of course: is it possible to find such an impulse decomposition for an arbitrary signal? The answer is 'yes'. Oliver Heaviside proposed to model a signal as a sum of impulses. His insight was based on the observation that for a physical quantity, its intensity and its duration are what determines its impact. E.g., a force that works on an object only has impact if both its intensity and its duration are nonzero.

The conclusion was obvious for Heaviside: it is the *area under the curve* (*AUC*) that is important for a signal. He therefore tried to simplify a signal by replacing it with simpler models, maintaining the same *AUC*.

Consider as an example the signal below.

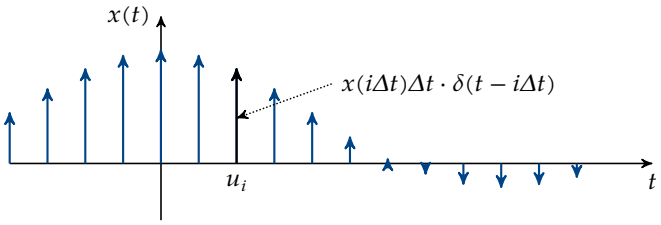


A Riemann sum was one of his attempts:



$$AUC \approx \sum_{i=-\infty}^{+\infty} x(i\Delta t) \cdot \Delta t \quad (\text{if } \Delta t \text{ small})$$

Replacing the Riemann rectangles with Dirac impulses with the same area, was another:



$$AUC \approx \int_{-\infty}^{\infty} \underbrace{\sum_{i=-\infty}^{+\infty} x(i\Delta t)\Delta t \cdot \delta(t - i\Delta t)}_{x_{approx}(t)} \, \mathrm{d}t \quad (\text{if } \Delta t \text{ small})$$

However, in both cases, we need to ensure that Δt is small enough in order not to miss any sudden evolution in the signal.

In the final approximation for the area A, we can readily see an approximation $x_{approx}(t)$ for the original curve $x(t)$ (marked with the underbrace).

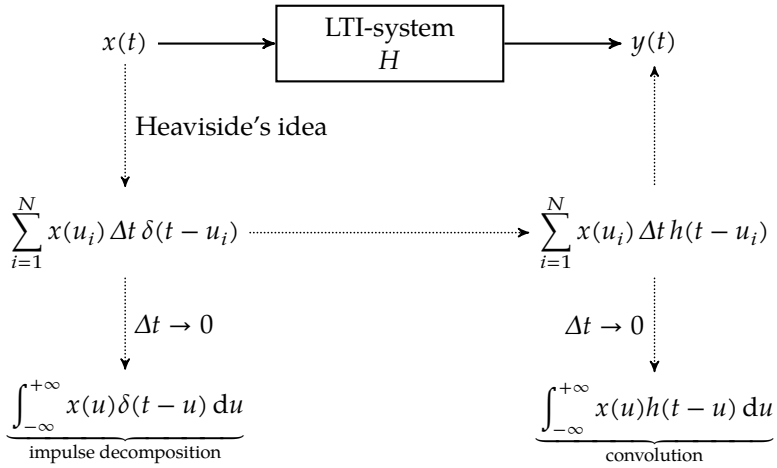
If we focus on this approximation with the idea of refining the partition of the domain t (i.e. $\Delta t \to 0$), we obtain:

$$\begin{aligned} x(t) &= \lim_{\Delta t \to 0} \sum_{i=-\infty}^{+\infty} x(i\Delta t)\Delta t \cdot \delta(t - i\Delta t) \\ &\Big\downarrow \quad \begin{array}{l} 1.\ \Delta t \to \mathrm{d}u \\ 2.\ i\Delta t \to u \\ 3.\ \sum \to \int \end{array} \\ &= \int_{-\infty}^{+\infty} x(u) \cdot \delta(t - u) \, \mathrm{d}u \end{aligned}$$

As stated earlier on page 20, this is what is known in mathematics as the *sifting property* of the Dirac impulse: it sifts the value of the function next to it for the argument where it peaks, out of the integral (in this case for $u = t$). In engineering, we call this the *impulse decomposition* of x: the signal x can be written as an infinitely dense sum of Dirac impulses.

Now let's return to the basic idea: when decomposing an input signal like this, does it allow to calculate the output more easily? Yes, it does. This is illustrated below. The input x is decom-

posed into a linear combination of impulses. The LTI property of the system allows to recompose the full output as the same linear combination of the impulse responses of the system.



And now, an insight that you might not understand right now, but that you may grasp after you have studied the Fourier and Laplace transforms: *the impulse decomposition is the forward transform from the time domain to the impulse (response) domain; the convolution is the inverse transform from the impulse (response) domain back to the time domain.*

4.3 The convolution operation and its properties

4.3.1 Definition

Let's start with a formal definition:

Convolution

The convolution of two functions f and g, results in a new function h that can be calculated as:

$$h(t) = \int_{-\infty}^{+\infty} f(u)g(t-u)\,\mathrm{d}u$$

We will use the five-pointed filled star as a shorthand for the convolution integral:

$$h(t) = f(t) \star g(t)$$

or more mathematically more correct:

$$h(t) = (f \star g)(t)$$

or for short: $h = f \star g$.

The convolution integral is so fundamentally important, that you better learn this integral by heart. The graphical example that follows next, might help you to remember the definition.

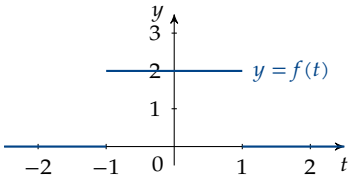
4.3.2 Graphical example

In this example we will use the rectangle function $\text{rect}_{a,b}(t)$ defined as:

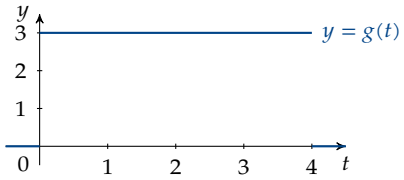
$$\text{rect}_{a,b}(t) = \text{u}(t-a) - \text{u}(t-b) \qquad \text{assuming} \quad a \leq b$$

Now, consider the following two functions:

$f(t) = 2\text{rect}_{-1,1}(t)$



$g(t) = 3\text{rect}_{0,4}(t)$

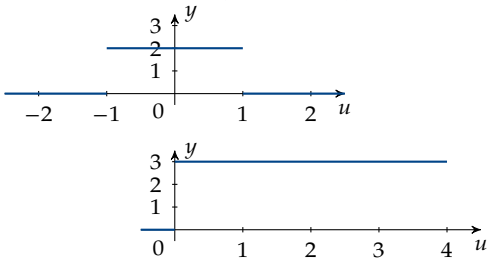


If we consider the convolution integral to be calculated:

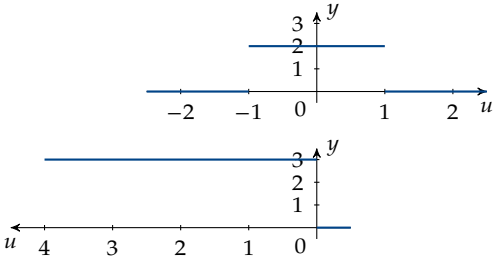
$$(f \star g)(t) = \int_{-\infty}^{+\infty} f(u)g(t-u)\,\mathrm{d}u$$

then we see that the first function is taken as is (with its argument t replaced by the variable u). The same happens for the second function but it has been flipped (u becomes $-u$) and then shifted over an amount t (because when $u = t$ then the argument $t - u$ becomes zero). This has been illustrated below.

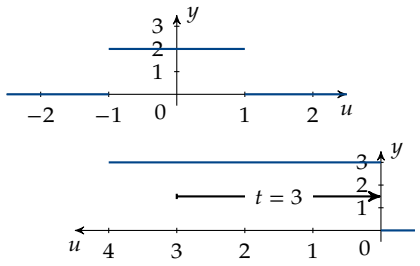
Step 1: replace arguments by u



Step 2: flip the second function



Step 3: shift the flipped function over a distance t (illustrated for t = 3)



Step 4: multiply f and g and integrate
Figure 4.1 illustrates this for several values of t. It may be clear that this is an involved procedure and that it is difficult to conduct this analysis in your mind when given two arbitrary functions. Therefore, don't be disappointed that you probably will not develop the skill to *guesstimate* the result of a convolution product. Our human brain has a limited capacity. We better learn to live with that.

4.3.3 Properties

In the same way that ordinary operations on a set (e.g., addition on the set of real numbers) have some properties, this convolution operation also has a number of properties. Knowing them can make life easier. In the discussion below, we use three functions f, g, and h. If we write a function f in a block in a block diagram, we mean "a system with impulse response f".

Convolution is commutative

This means that we can interchange the order in which functions appear in a convolution:

$$f \star g = g \star f$$

Illustrating this property with a block diagram

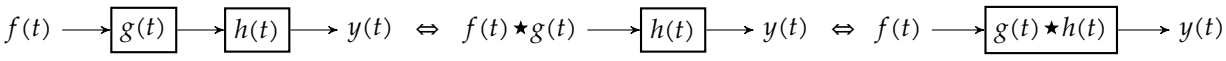
This means that a system with impulse response g and excited with a signal f will yield the same output as a system with impulse response f and excited with an input signal g.

Convolution is associative

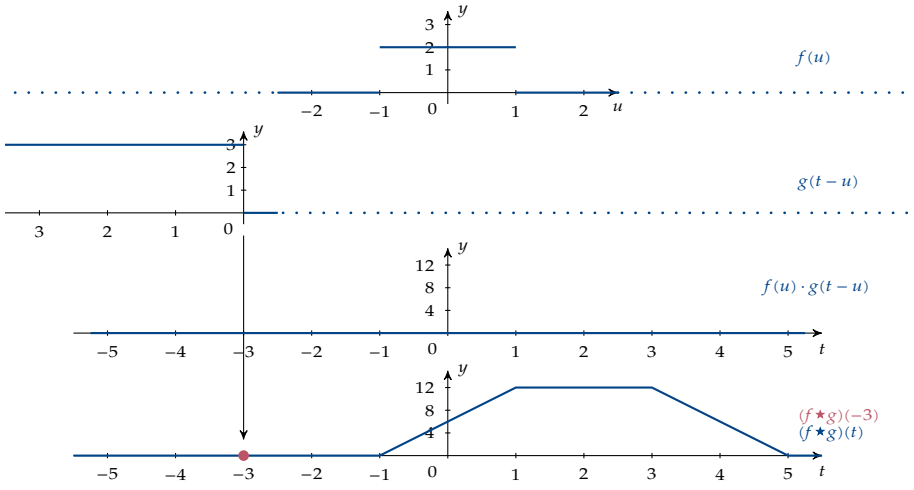
This means that the order in which you calculate a chain of convolutions is of no importance. Stated differently, you can put your brackets wherever you want:

$$f \star g \star h = f \star (g \star h) = (f \star g) \star h$$

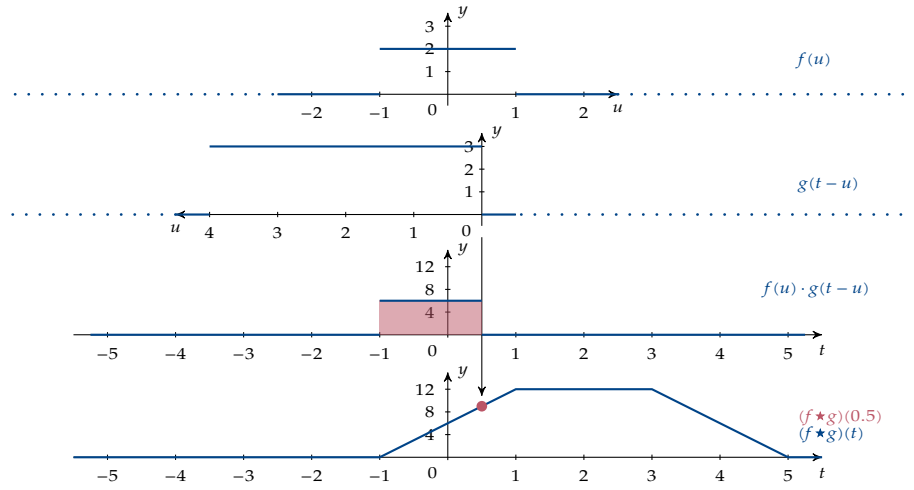
Graphically this can be represented as follows:



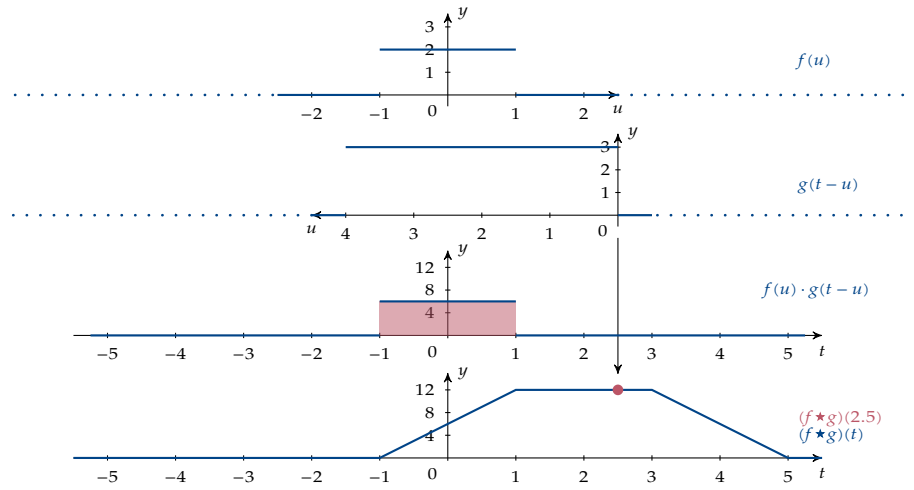
This means that you can simplify a cascade of two blocks as a single block with an impulse response that is the convolution of both impulse responses.



(a) Illustration for $t = -3$



(b) Illustration for $t = 0.5$



(c) Illustration for $t = 2.5$

Figure 4.1: Illustration of the calculation of the example convolution integral for several values of t

Convolution has a neutral element

The neutral element is the Dirac function. This property is due to the shifting property

$$f \star \delta = \delta \star f = f$$

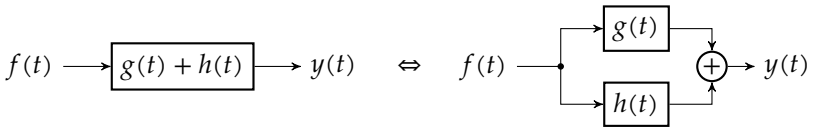
This means that a system with an impulse response equal to a Dirac impulse, is a pass-through system.

Convolution is distributive w.r.t. addition

Mathematically, we can summarize this as:

$$f \star (g + h) = f \star g + f \star h$$

Graphically:



This means that we can merge converging parallel paths into a block with the sum of the two impulse responses as its impulse response.

Simplification for causal waves

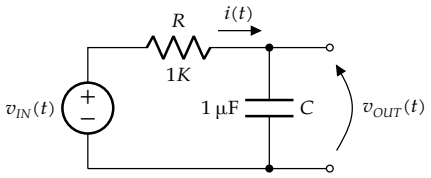
If f and g are causal, i.e. $f(t) = g(t) = 0$, for $t < 0$, we can limit the integration boundaries of the convolution integral:

$$(f \star g)(t) = \int_0^t f(u) g(t-u)\, \mathrm{d}u$$

This significantly reduces the complexity of calculating such an integral.

4.4 Example

Let's use convolution on a simple example to demonstrate how simple it can be to obtain the output. Consider the following simple network with input v_{IN} and output v_{OUT}:



We want to know how the system reacts when we excite it with a step input.

If you analyze this network using the methods explained in Chapter 3, you can obtain its impulse response:

$$h(t) = \mathrm{u}(t) \cdot \frac{1}{RC} \mathrm{e}^{-\frac{t}{RC}}$$

The step input is in fact the Heaviside function. We can now apply what we learned about convolution:

$$
\begin{aligned}
v_{OUT}(t) &= u(t) \star h(t) \\
&= \int_{-\infty}^{+\infty} u(\tau) \cdot h(t-\tau)\,\mathrm{d}\tau \\
&\quad \Big\downarrow \quad u(\tau) = \begin{cases} 0 & \text{if } \tau < 0 \\ 1 & \text{if } \tau \geq 0 \end{cases} \\
&\quad \Big\downarrow \quad h(t-\tau) = \begin{cases} 0 & \text{if } \tau > t \\ \frac{1}{RC}\,\mathrm{e}^{-\frac{t-\tau}{RC}} & \text{if } \tau \leq t \end{cases} \\
&= \int_0^t 1 \cdot \frac{1}{RC}\,\mathrm{e}^{-\frac{t-\tau}{RC}}\,\mathrm{d}\tau = \frac{1}{RC}\int_0^t \mathrm{e}^{-\frac{t}{RC}}\,\mathrm{e}^{\frac{\tau}{RC}}\,\mathrm{d}\tau \\
&= \frac{1}{RC}\,\mathrm{e}^{-\frac{t}{RC}}\int_0^t \mathrm{e}^{\frac{\tau}{RC}}\,\mathrm{d}\tau = \mathrm{e}^{-\frac{t}{RC}}\left[\mathrm{e}^{\frac{\tau}{RC}}\right]_0^t = 1 - \mathrm{e}^{-\frac{t}{RC}}
\end{aligned}
$$

Exercises

Exercise 4.4-1:
Consider a system that has the following impulse response:

$$h(t) = \mathrm{u}(t) \cdot \frac{1}{RC}\,\mathrm{e}^{-\frac{t}{RC}}$$

Determine the output signal y, assuming the following input signal:

$$x(t) = \mathrm{rect}_{0,1}(t)$$

Exercise 4.4-2:
Consider the same system as in the previous exercise. Determine the output signal y, assuming the following input signal:

$$x(t) = 3\,\delta(t)\,\sin(\omega t)$$

Exercise 4.4-3:
Consider the same system as in the previous exercise. Determine the output signal y, assuming the following input signal:

$$x(t) = -2\,\delta(t)$$

Exercise 4.4-4:
Prove the commutativity of the convolution operation.

Exercise 4.4-5:
Prove the associativity of the convolution operation.

Exercise 4.4-6:
Prove that the Dirac impulse is the neutral element for the convolution operation.

Exercise 4.4-7:
Prove the distributivity of the convolution operation w.r.t. addition.

Exercise 4.4-8:
Prove the simplicification of the convolution integral for causal signals.

4.5 Conclusion

In this chapter, we have learnt that LTI systems can be fully described using their impulse response. We have seen that the output of such a system can be calculated by convolving the input with the impulse response.

Chapter 5

The Continous-time Fourier transform

In this chapter, you will learn about:

- how Fourier theory is related to approximating functions,
- the different kinds of (continuous-time) Fourier transforms: Fourier series and the Fourier transform,
- how they are related,
- their properties, and
- how we can use Fourier theory for LTI-systems.

After having read/studied this chapter, you are expected to be able to

- use/apply the Fourier transform and its properties.

5.1 Introduction

The start of the 19^{th} century was a very interesting period for many reasons, one of them being the "discovery" of many interesting mathematical techniques to analyze signals and systems.

The well known Fourier, Laplace and Z-transform originated at that time. We won't try to be historically correct into how one scientist inspired the other, but fact is, that all these transforms are wonderfully connected. It's like they are all variations on the same theme. We also will not let ourselves be bothered by any convergence issues (though there are many!): our focus is on understanding the theory to a level on which we can safely use these transforms. If you ever get into convergence issues: find a good mathematician to help you out.

In this chapter, we will start from the Fourier series as "the mother of all signal transforms", and distill all other transforms from there. We'll do this according to the overall view of Figure 5.1 on the next page. We will only view the top half of the diagram, as these are the transforms 'in continuous time'. We will leave the discretization of time and the related transform for the course on digital signal processing (DSP) that will be on your menu later.

The diagram connects continuous signals and their transformed versions. Besides the mathematical beauty of this evolutionary diagram, the diagram and how transforms are related to each other is important to understand the effects introduced by any of these transforms. In the diagram, the inner ring represents signals in the time domain, while the outer ring represents signals in the frequency domain (see illustration in Figure 5.1).

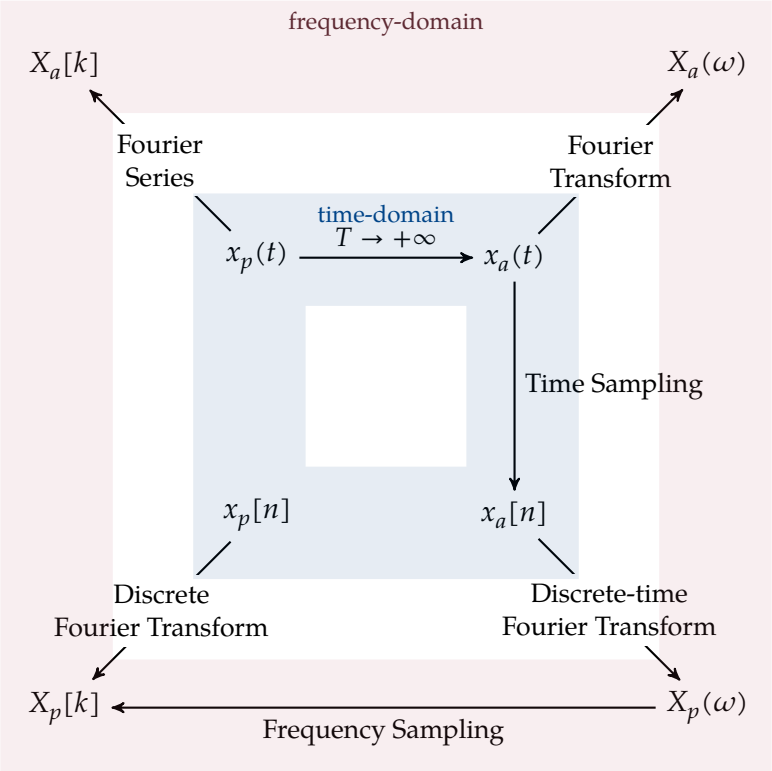


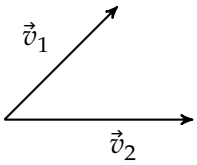
Figure 5.1: The Fourier family diagram; this course will cover the top half, i.e. Fourier series and Fourier transform; subscript p stands for periodic, subscript a stands for aperiodic

5.2 Vectors and approximating them

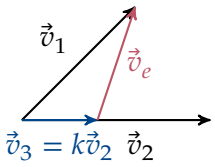
Fourier theory is closely related to approximating/decomposing vectors as a linear combination of base vectors. Therefore, we will first investigate this in a 2-dimensional vector space.

5.2.1 Vectors in 2D

Consider the vectors $\vec{v}_1$ and $\vec{v}_2$ below, that are both part of a two-dimensional vector space.



We'd like to find an approximation for $\vec{v}_1$ in the form of $\vec{v}_3 = k\vec{v}_2$. A possible candidate for such an approximation has been indicated in blue, below. The error committed $\vec{v}_e = \vec{v}_1 - k\vec{v}_2$ has been indicated in red.

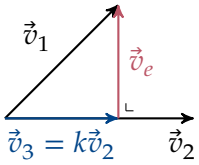


Intuitively, the best approximation for v_1 is the vector $k\vec{v}_2$ for which the error vector $\vec{v}_e$ is the smallest. This means that

$$\|\vec{v}_e\| = \|\vec{v}_1 - k\vec{v}_2\|$$

is a measure for the approximation error. We can use any valid concept of norm $\|\cdot\|$, but in this course, we will always use the Euclidean norm.

Intuitively, the best approximation is obtained if $\vec{v}_e \perp \vec{v}_2$:

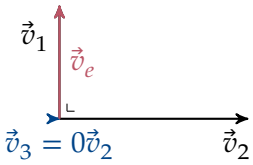


Therefore, we call this optimal approximation $\vec{v}_3$ the *orthogonal projection* of $\vec{v}_1$ on $\vec{v}_2$.

Note that if $\vec{v}_1$ would have been orthogonal to $\vec{v}_2$, a special situation arises, i.e.:

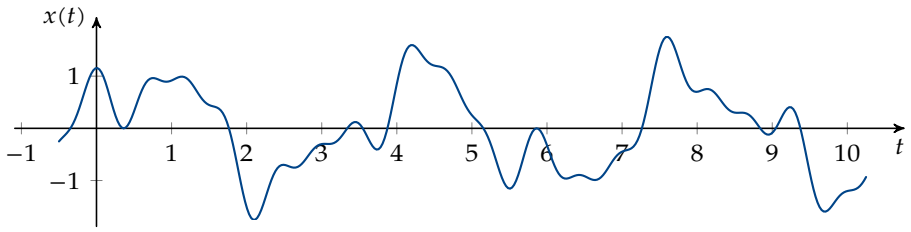
$$\vec{v}_1 \perp \vec{v}_2 \quad \Leftrightarrow \quad k = 0.$$

This has been illustrated below:

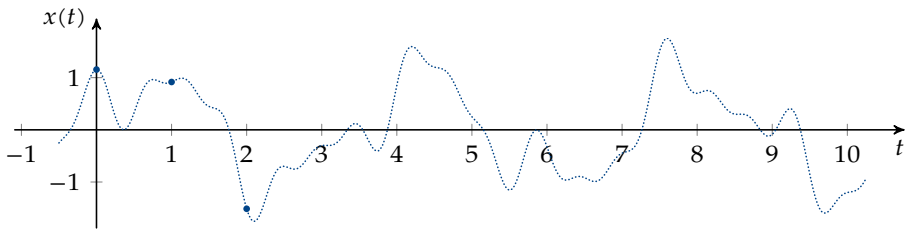


5.2.2 Signals as vectors

Consider the arbitrary signal below:



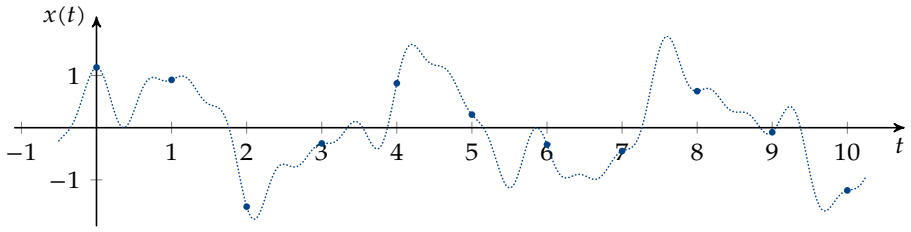
Specifying three samples of this function at known locations (indicated by the blue dots below) could be a simple way of describing the signal. Consider this to be a kind of finger print of the function.



This means, we can use a 3D-vector to represent this signal. The vector would in this case be

$$\vec{x} = [1.156, 0.917, -1.513]$$

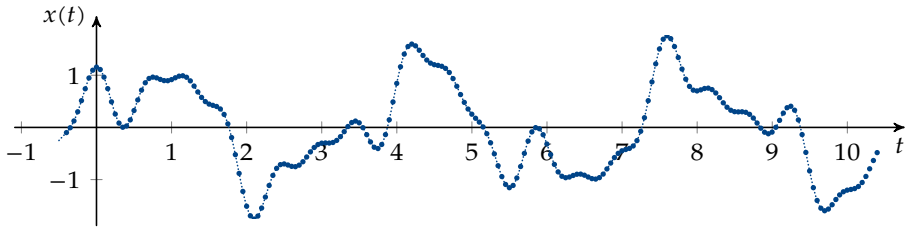
However, the level of detail that this vector provides is very limited. Many functions would have the same finger print. Indeed, we miss the entire range below 0 and above 2! Increasing the number of points could improve this. We might even consider an infinitely long vector, for all integer time positions:



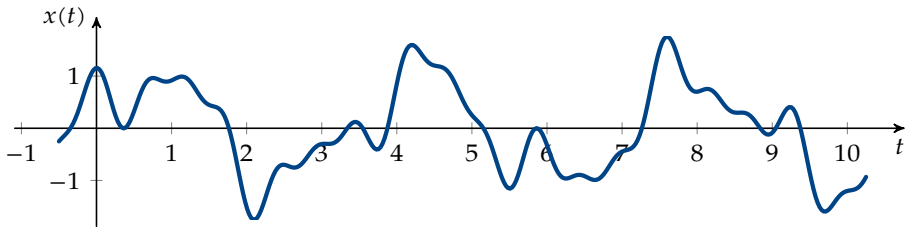
The vector would in this case be:

$$\vec{x} = [\ldots, 1.156, 0.917, -1.513, -0.301, 0.851, 0.251, -0.325, -0.447, 0.700, -0.087, -1.202, \ldots]$$

Yet we are still not satisfied, as the pitch of this point grid might be too coarse to grasp all interesting evolutions of the function that might occur in between any two points. In the course of DSP we will see what condition needs to be fulfilled in order to guarantee that we don't miss anything important. However, let's not go there for now. Let's be lazy: let us put an awful amount of points close together.



Or even better: let's put them infinitely close together, so that we obtain the original function again.



The conclusion is: we can consider a signal to be an infinitely long vector of points that are packed infinitely close together. If we gather all these signals into a set, we obtain the set of all possible real signals.

Still, we can go one step further. If we consider complex functions, we can even define the set of complex signals. These signals also can be seen as infinitely long and dense vectors with complex values.

As addition and scaling are well defined[1] on signals, we can truly call these sets *vector spaces*, the vector space of real signals and the vector space of complex signals.

5.2.3 Scalar product for signals

There is more. If we can define a proper scalar product (or inner product), we also may define a distance metric between the elements, such that we can compare signals to see how much alike or different they are.

We define the inner product of two complex functions as follows:

Scalar product of two complex functions

The scalar product $\langle f, g \rangle$ of two complex functions f and g on a domain Ω is defined as:

$$\langle f, g \rangle = \int_{\Omega} f(x)\overline{g(x)}\,\mathrm{d}x$$

Remember: the bar over g denotes taking the complex conjugate of g.
The domain can be the set of real numbers or any interval in that set on which the functions f and g are defined.

This definition can be simplified in case the two functions are real:

Scalar product of two real functions

The scalar product $\langle f, g \rangle$ of two real functions f and g on a domain Ω is defined as:

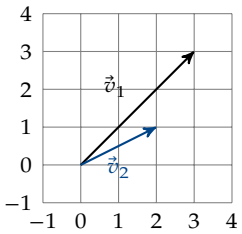
$$\langle f, g \rangle = \int_{\Omega} f(x)g(x)\,\mathrm{d}x$$

Note that the latter definition is just a special case of the former one.

The result of a scalar product is a complex number for complex signals, or a real number for real signals.

Note that this definition is very similar to the scalar product of ordinary vectors. Consider e.g. the 2D case below with vectors with coordinates $(3, 3)$ and $(2, 1)$:

[1] We don't go into detail on what it means to be well defined. We'll leave that for the mathematicians. Everyone's got his own turf.



The scalar product is calculated as:

$$\langle \vec{v}_1, \vec{v}_2 \rangle = 3 \cdot 2 + 3 \cdot 1 = 9$$

This is a sum of products of the corresponding values. The definition for signals is also a sum (the integral) of a product of the corresponding values (the product of the two functions for every value of the interval).

A vector space that is equipped with a proper scalar product is called a *metric space*. If it is also complete[2] then it is a so-called *Hilbert-space*. In our case, the Hilbert space of complex (or real) functions. We denote it as $L^2(\Omega, \mathbb{C})$ for complex functions on a domain Ω and $L^2(\Omega, \mathbb{R})$ for real functions an a domain Ω.

For functions f to be in this space, they must be *measurable*, i.e.

$$\langle f, f \rangle < \infty \qquad \Leftrightarrow \qquad \int_\Omega |f(x)|^2 \, \mathrm{d}x$$

The origin of the concept of measurability will become clear in the next section.

5.2.4 The length or size of signals

The scalar product is especially useful in defining the length of a vector, which in *vector space lingo* is called the *norm* of a vector. For signals, this can be seen as the size of the signal:

Norm of a signal

The norm $\|f\|$ of a complex or real signal f on a domain Ω is defined as:

$$\|f\| = \sqrt{\langle f, f \rangle}$$

It's not so hard to prove that the result of this norm is always real and positive and as such also corresponds well to what we expect of a length.

On the side, it is worth to note that:

$$\|f\| \Leftrightarrow f = 0 \text{ a.e.}$$

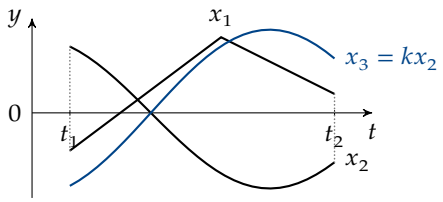
where 'a.e.' stands for 'almost everywhere'. With this, we mean that signals f that are only different from 0 in maximally a countable infinite number of points (i.e. *almost everwhere zero*), are considered to be of zero length and therefore are considered to be equivalent to the zero vector. In the practice of engineering, such a signals do not occur.

[2] Again, we will not go into detail into explaining completeness. Remember: everyone's got his own turf.

5.2.5 Approximating signals

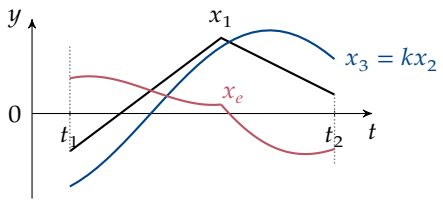
Given the fact that we can consider signals to be vectors, we can ask ourselves the question of the beginning of this section: given two signals x_1 and x_2, can we find an approximation for x_1 in the form of $x_3 = kx_2$?

Considering two functions x_1 and x_2 on an a real interval $[t_1, t_2]$, the following figures shows our intent:



Squinting our eyes a bit, our choice for $k = -1$ in the figure above seems not so bad. However, is there a formal way to determine the best k? Yes, there is.

To this end, we need to consider the error signal $x_e = x_1 - kx_2$, indicated in read below:



The smaller this function is, the better the approximation. Wait...what does it mean for a function to be small? Well, using our new concept of the norm of a function, we can exactly describe what we mean. We want

$$\|x_e\| = \sqrt{\int_{t_1}^{t_2} (x_e(t))^2 \, \mathrm{d}t}$$

to be as small as possible. Given the fact that norms are positive, requiring that $\|x_e\|^2$ is minimal is equally well.

Let's elaborate that idea. Our goal is to find a value for k causing $x_3 = kx_2$ to be such that the above squared norm becomes minimal. We know how to find an optimum of a function, don't we? The value of $\|x_e\|^2$ reaches extreme values when $\frac{\mathrm{d}\|x_e\|^2}{\mathrm{d}k} = 0$, i.e.

$$\frac{\mathrm{d}}{\mathrm{d}k}\left(\int_{t_1}^{t_2} (x_1(t) - kx_2(t))^2 \, \mathrm{d}t\right) = 0$$

$$\downarrow \text{ Leibniz's rule}$$

$$\begin{aligned}
&\Leftrightarrow \quad \int_{t_1}^{t_2} \frac{\mathrm{d}}{\mathrm{d}k}(x_1(t) - kx_2(t))^2 \, \mathrm{d}t = 0 \\
&\Leftrightarrow \quad \int_{t_1}^{t_2} 2(x_1(t) - kx_2(t))(-x_2(t)) \, \mathrm{d}t = 0 \qquad (5.1) \\
&\Leftrightarrow \quad -2\int_{t_1}^{t_2} x_1(t)x_2(t) \, \mathrm{d}t + 2k\int_{t_1}^{t_2} x_2^2(t) \, \mathrm{d}t = 0
\end{aligned}$$

Solving k from the latter equation, results in:

$$k = \frac{\int_{t_1}^{t_2} x_1(t)x_2(t)\,\mathrm{d}t}{\int_{t_1}^{t_2} x_2^2(t)\,\mathrm{d}t} \tag{5.2}$$

By mere analogy we can make the following observations:

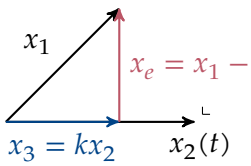
- we call $x_3 = kx_2$, with the above value of k, the orthogonal projection of x_1 onto x_2;
- if k turns out to be zero, i.e.
$$\int_{t_1}^{t_2} x_1(t)x_2(t)\,\mathrm{d}t = 0$$
then we say that x_1 and x_2 are orthogonal, in symbols: $x_1 \perp x_2$.
- Note that (5.1) in fact corresponds to:
$$\langle x_e(t), x_2(t)\rangle = 0$$

Note that we can rewrite (5.2) as:

$$k = \frac{\langle x_1, x_2\rangle}{\langle x_2, x_2\rangle}$$

which is an equation to remember! We call it the *projection formula*.

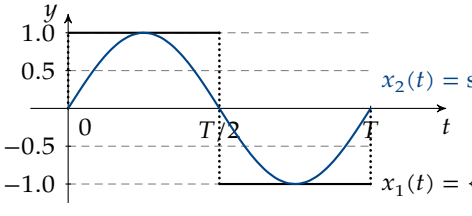
Don't try to find a 90° angle in this concept of orthogonality. You will not find any such angle. Remember at all times: this is just a very useful analogy. Therefore, it makes sense to keep the drawing below as a mental representation of how we approximate functions:



$$k = \frac{\langle x_1(t), x_2(t)\rangle}{\langle x_2(t), x_2(t)\rangle} \quad \text{and} \quad \langle x_e(t), x_2(t)\rangle = 0$$

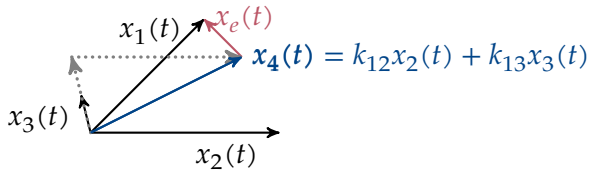
Exercises

Exercise 5.2.5-1:
Determine the best approximation of x_1 in terms of x_2:



5.2.6 Improving the approximation

Approximating a vector by scaling another one appropriately has no more secrets to us. What if we could use two vectors in the approximation (instead of one)?
Specifically: can we find values for k_{12} and k_{13} such that $x_4 = k_{12}x_2 + k_{13}x_3$ is the best approximation for x_1? This should improve the approximation, no?



The error made has been indicated in red. Again, it is our goal to make x_e as small as possible, i.e. minimizing

$$\|x_e(t)\|^2 = \int_{t_1}^{t_2} (x_1(t) - k_{12}x_2(t) - k_{13}x_3(t))^2 \, \mathrm{d}t$$

From optimization theory, we know that this function of k_{12} and k_{13} will reach extreme values on spots where the partial derivatives w.r.t. k_{12} and k_{13} are zero, symbolically $\frac{\partial \|x_e\|^2}{\partial k_{ij}} = 0$.

Let's elaborate this idea. We start with the partial derivative w.r.t. k_{12}:

$$\begin{aligned}
& \frac{\partial}{\partial k_{12}} \int_{t_1}^{t_2} (x_1(t) - k_{12}x_2(t) - k_{13}x_3(t))^2 \, \mathrm{d}t = 0 \\
\Leftrightarrow \quad & \int_{t_1}^{t_2} \frac{\partial}{\partial k_{12}} (x_1(t) - k_{12}x_2(t) - k_{13}x_3(t))^2 \, \mathrm{d}t = 0 \\
\Leftrightarrow \quad & \int_{t_1}^{t_2} 2\,(x_1(t) - k_{12}x_2(t) - k_{13}x_3(t))\,(-x_2(t)) \, \mathrm{d}t = 0 \\
\Leftrightarrow \quad & \int_{t_1}^{t_2} x_1(t)x_2(t) \, \mathrm{d}t - k_{12} \int_{t_1}^{t_2} x_2^2(t) \, \mathrm{d}t - k_{13} \int_{t_1}^{t_2} x_2(t)x_3(t) \, \mathrm{d}t = 0
\end{aligned}$$

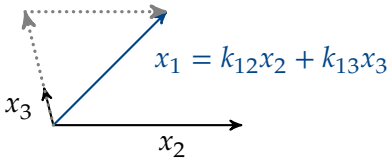
Similarly, using the partial derivative w.r.t. k_{13} we can find:

$$\int_{t_1}^{t_2} x_1(t)x_3(t) \, \mathrm{d}t - k_{12} \int_{t_1}^{t_2} x_2(t)x_3(t) \, \mathrm{d}t - k_{13} \int_{t_1}^{t_2} x_3^2(t) \, \mathrm{d}t = 0$$

This yields the following set of simultaneous equations that are linear in k_{12} and k_{13}:

$$\begin{cases}
\int_{t_1}^{t_2} x_1(t)x_2(t) \, \mathrm{d}t - k_{12} \int_{t_1}^{t_2} x_2^2(t) \, \mathrm{d}t - k_{13} \int_{t_1}^{t_2} x_2(t)x_3(t) \, \mathrm{d}t = 0 \\
\int_{t_1}^{t_2} x_1(t)x_3(t) \, \mathrm{d}t - k_{12} \int_{t_1}^{t_2} x_2(t)x_3(t) \, \mathrm{d}t - k_{13} \int_{t_1}^{t_2} x_3^2(t) \, \mathrm{d}t = 0
\end{cases}$$

Remembering that for a specific x_1 and x_2 the integrals above can be readily calculated and result in real numbers, we know that determining k_{12} and k_{13} is a piece of cake and will result in a 'perfect approximation':



Note that if x_2 and x_3 would have been orthogonal, i.e.

$$x_2 \perp x_3 \qquad \Leftrightarrow \qquad \int_{t_1}^{t_2} x_2(t)x_3(t)\,\mathrm{d}t = 0$$

life would be much simpler, as in that case the equation set can be simplified into:

$$\begin{cases} \int_{t_1}^{t_2} x_1(t)x_2(t)\,\mathrm{d}t - k_{12}\int_{t_1}^{t_2} x_2^2(t)\,\mathrm{d}t - \cancel{k_{13}\int_{t_1}^{t_2} x_2(t)x_3(t)\,\mathrm{d}t} = 0 \\ \int_{t_1}^{t_2} x_1(t)x_3(t)\,\mathrm{d}t - \cancel{k_{12}\int_{t_1}^{t_2} x_2(t)x_3(t)\,\mathrm{d}t} - k_{13}\int_{t_1}^{t_2} x_3^2(t)\,\mathrm{d}t = 0 \end{cases}$$

This removes x_3 from the first equation and x_2 from the second equation and as such allows to solve for k_{12} and k_{13} by solving every equation independently, resulting in:

$$\begin{cases} k_{12} = \dfrac{\langle x_1, x_2\rangle}{\langle x_2, x_2\rangle} \\ k_{13} = \dfrac{\langle x_1, x_3\rangle}{\langle x_3, x_3\rangle} \end{cases}$$

This means that x_3 does not influence the value of k_{12} and vice versa x_2 does not influence the value of k_{13}. Note that these equations again are instances of the projection formula that we have seen earlier.

Time for some intermediate conclusions:

- the more signals we can involve in the approximation, the better it gets;
- if the signals we use for the approximation are orthogonal, the mathematical derivation is greatly simplified;
- if the signals we use for the approximation are orthogonal, the k_i become independent of each other.

Some questions that we did not answer yet:

- Is an extra vector always helpful in improving the approximation? The answer is: no. If that new vector is not independent (i.e. if you can write it as a linear combination of the others) it will not help.
- How many independent vectors can you find? As many as the dimension of the vector space. In a 3D space this means three vectors; for the Hilbert space of complex functions the number of independent vectors is infinite.

A full set of independent vectors is called a *base* for the vector space. The base of a vector space is a set of vectors that is independent and spans the entire space, i.e. one can write any vector as a linear combination of the base vectors. To this end, one needs as many base vectors as the dimension of the vector space.

Though finding good bases is a good sport with a lot of fun, we are not going to devote time to that. If you ever need a method, look for *Gram-Schmidt-orthogonalization*. Any good mathematical book on function spaces will cover that. Instead we will investigate some good bases that have been composed for us.

For periodic functions with period T:

- the set of all sines and cosines with frequencies that are a positive multiple of $\omega_1 = 2\pi/T$:
$$\mathcal{A} = \{1, \cos k\omega_1 t, \sin k\omega_1 t \mid k \in \mathbb{N}_0\}$$
Note that for $k = 0$ the cosine results in 1 (the first element of the set) and the sine wave results in 0. We exclude the latter from $\mathcal{A}$, as the zero vector cannot be part of a set of base vectors.
- the set of all sinors with frequencies that are an integer multiple of $\omega_1 = 2\pi/T$:
$$\mathcal{B} = \left\{ \mathrm{e}^{\mathrm{j}k\omega_1 t} \mid k \in \mathbb{Z} \right\}$$

For arbitrary functions:

- the set of all possible unit Dirac impulses (that we already met in the previous chapter):
$$\mathcal{C} = \{\delta(t-\tau) \mid \tau \in \mathbb{R}\}$$
- the set of all sines and cosines:
$$\mathcal{D} = \{1, \cos \omega t, \sin \omega t \mid \omega \in \mathbb{R}\}$$
- the set of all sinors:
$$\mathcal{E} = \left\{ \mathrm{e}^{\mathrm{j}\omega t} \mid \omega \in \mathbb{R} \right\}$$

The sets $\mathcal{A}, \mathcal{B}, \mathcal{D}$ and $\mathcal{E}$ result in the Fourier series and the Fourier transform.

5.3 The Fourier series

Let's start with trying to approximate periodic signals, which is the playground for the *Fourier series*.

Several versions of the Fourier series exist. The most general, but least intuitive one is the exponential series. However, we will start with the trigonometric (or harmonic) series for real functions, as this avoids having to deal with complex numbers.

5.3.1 Trigonometric series

The base

One can prove that the following set of signals

$$\mathcal{A} = \{1, \cos k\omega_1 t, \sin k\omega_1 t \mid k \in \mathbb{N}\} \qquad \text{with } \omega_1 = \frac{2\pi}{T}$$

is an orthogonal set, i.e. for any two functions x_1 and x_2 in that set, we know that

$$\langle x_1, x_2 \rangle = 0 \qquad \Leftrightarrow \qquad \int_0^T x_1(t) x_2(t)\, \mathrm{d}t = 0$$

We also can prove that this set spans the entire Hilbert-space of real functions defined on the interval 0 to T (with the restriction that $x(0) = x(T)$). Therefore $\mathcal{A}$ is a base for the vector space.

We call ω_1 the fundamental frequency and $\omega_k = k\omega_1$ the *overtones* or *harmonics*. This also explains why we call this set the *harmonic set*. Obviously, the fundamental frequency is also the first harmonic.

The decomposition

Consider a real piecewise continuous signal $x(t)$ that is periodic with period T. This signal can be decomposed into a sum of sine and cosine signals, the so-called

Trigonometric Fourier series

$$x(t) \sim \frac{A_0}{2} + \sum_{k=1}^{+\infty} A_k \cos(\omega_k t) + \sum_{k=1}^{+\infty} B_k \sin(\omega_k t) \tag{5.3}$$

with

$$\omega_k = k\frac{2\pi}{T}$$

and

$$A_k = \frac{\langle x(t), \cos \omega_k t \rangle}{\langle \cos \omega_k t, \cos \omega_k t \rangle} = \frac{2}{T} \int_{-T/2}^{T/2} x(t) \cos(\omega_k t)\, \mathrm{d}t \tag{5.4}$$

$$B_k = \frac{\langle x(t), \sin \omega_k t \rangle}{\langle \sin \omega_k t, \sin \omega_k t \rangle} = \frac{2}{T} \int_{-T/2}^{T/2} x(t) \sin(\omega_k t)\, \mathrm{d}t \tag{5.5}$$

We denote equations (5.4) and (5.5), as the *analysis equations*, *decomposition equations* or just the *transform (equations)*. Equation (5.3) is the so-called *composition equation*, or *synthesis equation*, or just the *inverse transform (equation)*. The coefficients A_k and B_k are the so-called *trigonometric or harmonic Fourier coefficients*.

This naming convention can be used for all types of transforms.

Note that the equations for A_k and B_k are just instances of the general projection formula.

Many textbooks define ω_0 to be equal to $2\pi/T$ and write ω_k explicitly as $k\omega_0$. The notation used here, is a better one and has been chosen for its compactness and consistency.

This Fourier series can be understood as follows: the original signal $x(t)$ with period T can be written as sum of sines and cosines with frequencies that are a multiple of $2\pi/T$. This has been illustrated for an arbitrary function in Figure 5.2.

Remarks

1. The coefficients A_k and B_k have the same dimension as $x(t)$.
2. The integration boundaries may shift (as $x(t)$ and sines and cosines are periodic with period T), e.g.
$$B_k = \frac{2}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) \sin \omega_k t\, \mathrm{d}t = \frac{2}{T} \int_{a-\frac{T}{2}}^{a+\frac{T}{2}} x(t) \sin \omega_k t\, \mathrm{d}t.$$
3. When the function x is continuous at t_0 the series converges to the function value; when it contains a jump at t_0 the series converges to
$$\frac{x(t_0^+) + x(t_0^-)}{2}.$$
Therefore, we write a wiggle '$\sim$' instead of an equal sign.
4. At these jumps an effect called 'non-uniform convergence' occurs; as engineers we call this the 'Gibbs effect'.

The Gibbs effect

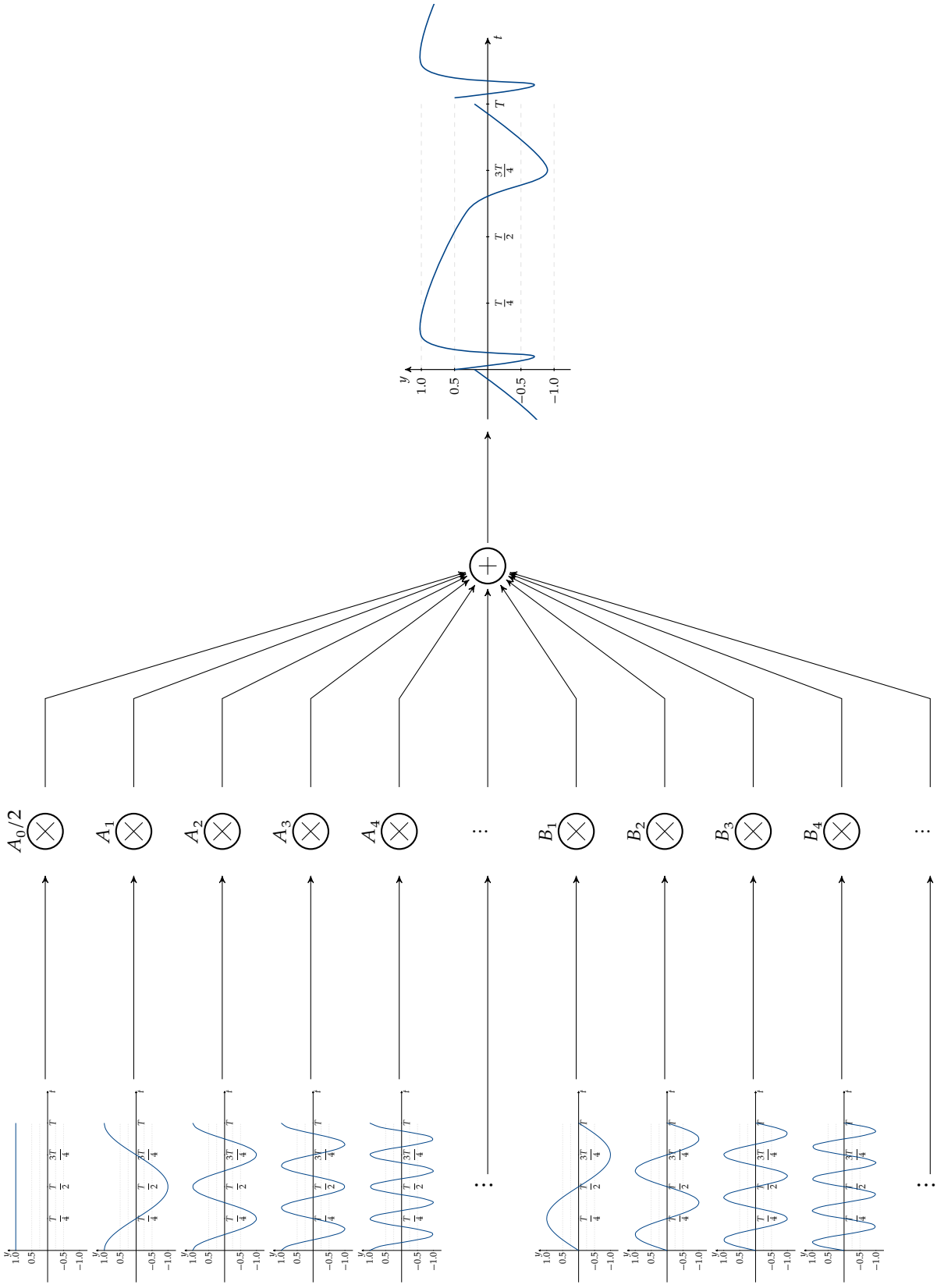
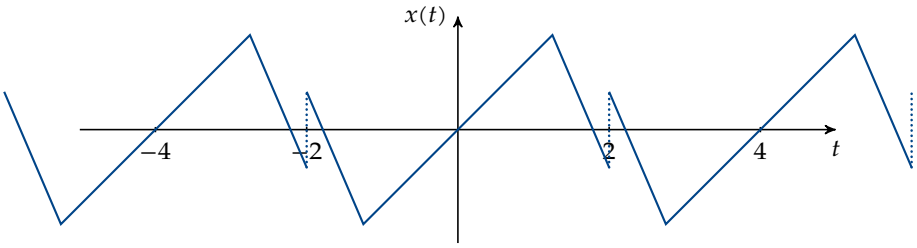
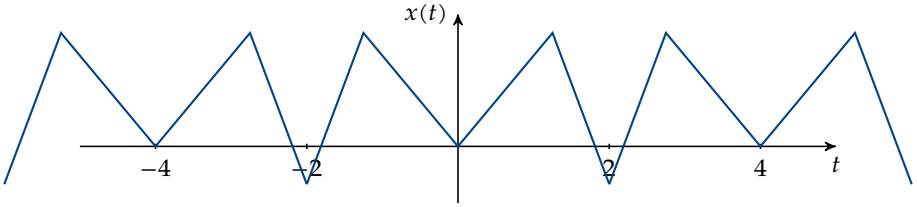


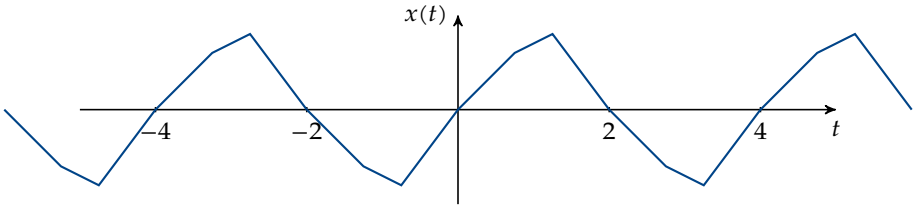
Figure 5.2: Illustration of how a periodic function can be written as a linear combination (with coefficients A_k and B_k) of sine and cosines



(a) Odd symmetry



(b) Even symmetry



(c) Half-wave symmetry

Figure 5.3: Illustration of some symmetry properties of periodic waveforms (with $T = 4$)

The summations in (5.3) are not just summations, but series running from 0 or 1 to plus infinity. Fourier was very much contested (a.o. by Lagrange), because those series do not fully converge for signals that exhibit discontinuities (jumps in the signal). This property became known afterwards as the Gibbs phenomenon. The phenomenon can be visualized by truncating the series after a chosen number of terms (see Figure 5.4 on page 66). The problem is, that the relative amplitude of the Gibbs phenomenon doesn't vanish for an increasing number of terms (it remains about 9% of the height of the jump). This isn't really a worry, as it can be shown that the energy in the Gibbs residu goes to zero for an increasing number of terms.

Symmetry allows simplification

In case the waveform exhibits some symmetries, the calculation of the coefficients A_k and B_k can be simplified.

Odd symmetry

If the signal is odd, i.e. $x(-t) = -x(t),\ \forall t$ (see Figure 5.3a), then the cosine coefficients all become zero:

$$A_k = \frac{2}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} x(t)\cos\omega_k t\,\mathrm{d}t = 0$$

Even symmetry

If the signal is even, i.e. $x(-t) = x(t),\ \forall t$, (see Figure 5.3b), then the sine coefficients all become

zero:

$$B_k = \frac{2}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} x(t)\sin\omega_k t\,\mathrm{d}t = 0$$

Half-wave symmetry

If the signal has the following property

$$x(t) = -x\left(t + \frac{T}{2}\right)$$

for all possible values of t, with T the period of x, then we call the signal *half-wave symmetrical*. This has been illustrated in Figure 5.3c. In that case all the even sine and cosine coefficients become zero:

$$A_{2k} = B_{2k} = 0$$

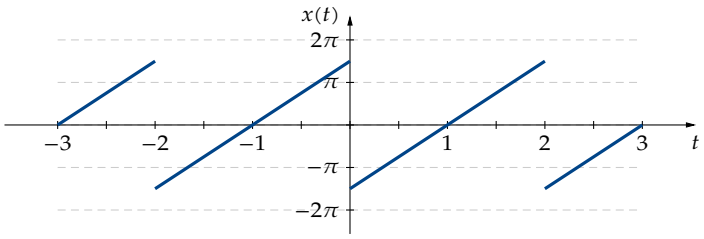
The example in Figure 5.3c is a special case: it is also odd and therefore for the example also $A_k = 0$.

Exercises

Some exercises on trigonometric Fourier series

Exercise 5.3.1-1:

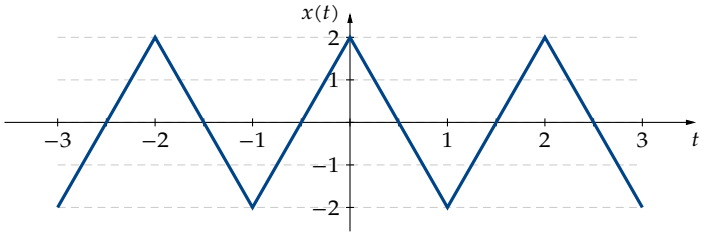
Calculate the trigonometric Fourier series of $x(t)$ as specified in the graph below:



Verify your result using Octave/Matlab by plotting a graph of the series you generated using the terms for $k = [-100, 100]$.

Exercise 5.3.1-2:

Calculate the trigonometric Fourier series of $x(t)$ as specified in the graph below:



Verify your result using Octave/Matlab by plotting a graph of the series you generated using the terms for $k = [-50, 50]$.

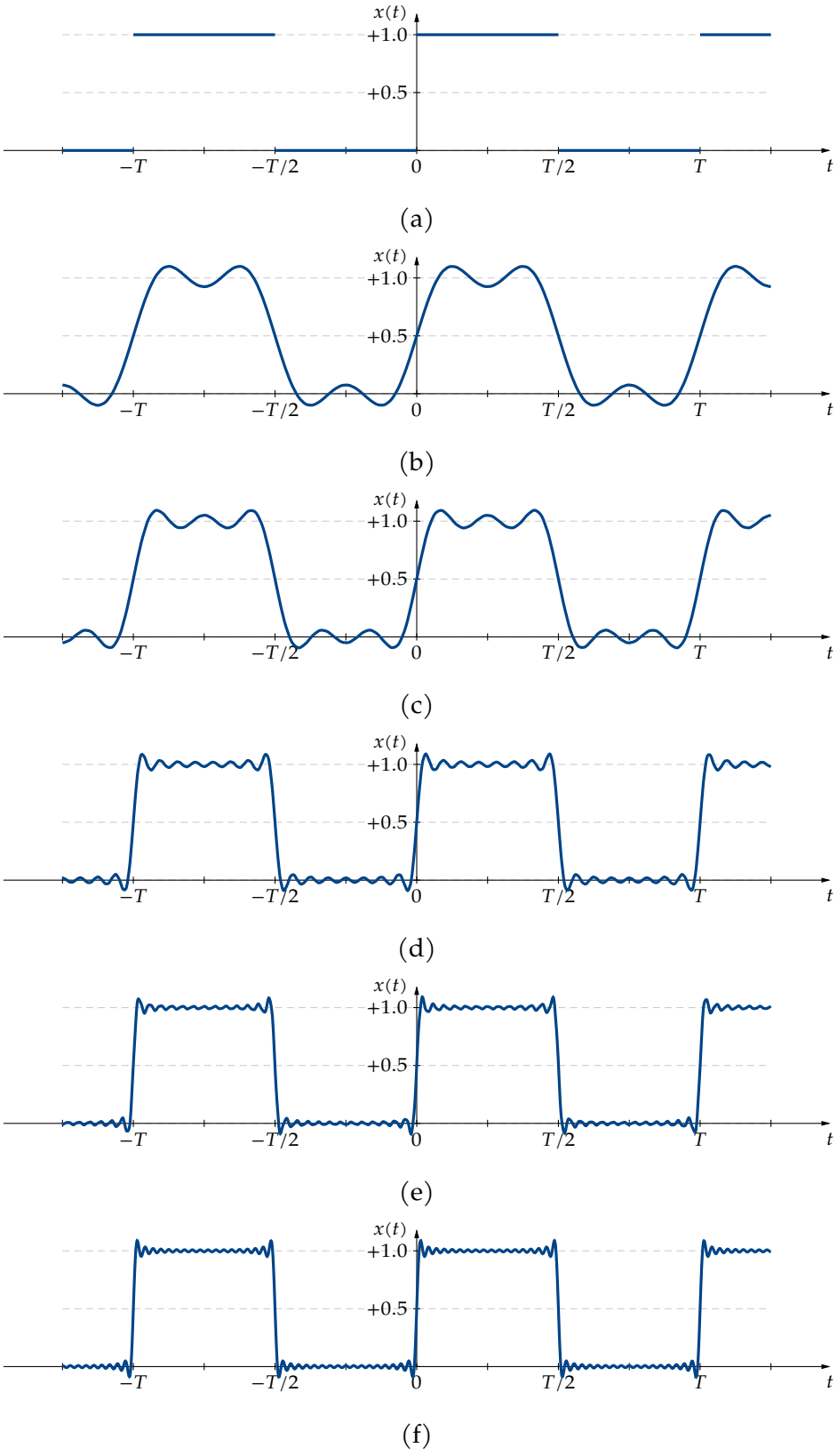


Figure 5.4: Illustration of the Gibbs phenomenon for a squarewave with period T and dutycyle of 50%: (a) original signal, (b) approximation $k \leq 3$, (c) $k \leq 5$, (d) $k \leq 15$, (e) $k \leq 25$, (f) $k \leq 35$

.

5.3.2 Alternative trigonometric series

We can rewrite the harmonic series a bit, by joining the sine and cosine series into one:

$$\begin{aligned} x(t) &\sim \frac{A_0}{2} + \sum_{k=1}^{+\infty} A_k \cos \omega_k t + \sum_{k=1}^{+\infty} B_k \sin \omega_k t \\ &\sim \frac{A_0}{2} + \sum_{k=1}^{+\infty} (A_k \cos \omega_k t + B_k \sin \omega_k t) \end{aligned}$$

If we focus on the generic term of this unified series, we can do some smart rewriting of this term:

$$T_k \equiv A_k \cos \omega_k t + B_k \sin \omega_k t$$

To this end, we use triangles, as explained below.

Towards an all cosine series

Given the fact that A_k and B_k are real numbers, they can be considered the Cartesian coordinates of point in the complex plain $(x, y) = (A_k, B_k)$. The polar coordinates of the same point $(r, \theta) = (M_k, \phi_k)$ are related to the Cartesian coordinates by

$$A_k = M_k \cos \phi_k \qquad\qquad B_k = M_k \sin \phi_k$$

This allows rewriting T_k as follows:

$$\begin{aligned} T_k &= A_k \cos \omega_k t + B_k \sin \omega_k t \\ &= M_k \cos \phi_k \cos \omega_k t + M_k \sin \phi_k \sin \omega_k t \\ &= M_k \left(\cos \phi_k \cos \omega_k t + \sin \phi_k \sin \omega_k t\right) \\ &= M_k \cos(\omega_k t - \phi_k) \end{aligned}$$

And this allows us to write an alternative Fourier series:

$$x(t) \sim \frac{A_0}{2} + \sum_{k=1}^{+\infty} M_k \cos(\omega_k t - \phi_k)$$

Towards an all sine series

In a similar way, we can consider A_k and B_k to be the cartesian coordinate s of a point in the complex plain $(x, y) = (B_k, A_k)$ such that the polar coordinates of the same point, $(r, \theta) = (M_k, \psi_k)$, are related by

$$B_k = M_k \cos \psi_k \qquad\qquad A_k = M_k \sin \psi_k$$

This allows rewriting T_k as follows:

$$\begin{aligned} T_k &= A_k \cos \omega_k t + B_k \sin \omega_k t \\ &= M_k \sin \psi_k \cos \omega_k t + M_k \cos \psi_k \sin \omega_k t \\ &= M_k \left(\sin \psi_k \cos \omega_k t + \cos \psi_k \sin \omega_k t\right) \\ &= M_k \sin(\omega_k t + \psi_k) \end{aligned}$$

And this allows us to write an alternative Fourier series:

$$x(t) \sim \frac{A_0}{2} + \sum_{k=1}^{+\infty} M_k \sin(\omega_k t + \psi_k)$$

5.3.3 Exponential series

The base

One can prove that the following set of signals

$$\mathcal{B} = \left\{ \mathrm{e}^{\mathrm{j}k\omega_1 t} \mid k \in \mathbb{Z} \right\}$$

is an orthogonal set, i.e. for any two functions x_1 and x_2 in the set, we know that

$$\langle x_1, x_2 \rangle = 0 \qquad \Leftrightarrow \qquad \int_0^T x_1(t) x_2(t)\, \mathrm{d}t = 0$$

We also can prove that this set spans the entire Hilbert-space of (complex) functions defined on the interval 0 to T. Therefore $\mathcal{B}$ is a base for the vector space.

We call this set the *exponential or sinor set*.

The decomposition

Consider a (complex) piecewise continuous signal $x(t)$ that is periodic with period T. This signal can be decomposed into a sum of exponential signals, the so-called exponential base functions. [3]

Exponential Fourier series

$$x(t) = \sum_{k=-\infty}^{+\infty} X_k \, \mathrm{e}^{\mathrm{j}\omega_k t} \tag{5.6}$$

with

$$\omega_k = k\frac{2\pi}{T}$$

and

$$X_k = \left\langle x, \mathrm{e}^{\mathrm{j}\omega t} \right\rangle = \frac{1}{T} \int_{-T/2}^{T/2} x(t)\, \mathrm{e}^{-\mathrm{j}\omega_k t}\, \mathrm{d}t \tag{5.7}$$

Again, we denote equation (5.7), as the *analysis equation*, *decomposition equation* or just the *transform (equation)*. Equation (5.6) is the so-called *composition equation*, or *synthesis equation*, or just the *inverse transform (equation)*. The coefficients X_k are the so-called *exponential Fourier coefficients* or *complex harmonic numbers*.

Note that the equation for X_k is just an instance of the general projection formula. Sometimes, we will write X_k as $X[k]$ suggesting that it is a discrete function (hence the square brackets) of the variable k.

Remarks

1. Given the fact that both $x(t)$ and $\mathrm{e}^{-j\omega_k t}$ are periodic with period T, also there product has the same period. This means that we can calculate (5.7) by integration over any period:

   $$X_k = \frac{1}{T} \int_a^{a+T} x(t)\, \mathrm{e}^{-\mathrm{j}\omega_k t}\, \mathrm{d}t$$

   for any arbitrary a.

[3] As probably all electrical engineers do, we use the character j to denote the imaginary number, i.e. the principal value that fulfills $\mathrm{j}^2 = -1$.

2. The Fourier series is often symbolized as an operator FS. In this case the coefficients of the series are written as a discrete function $X[k]$.

$$X[k] = \text{FS}(x(t)) \qquad\qquad x(t) = \text{FS}^{-1}(X[k])$$

3. Given this notation, the use of lowercase symbols for time-domain signals and corresponding uppercase symbols for their Fourier series coefficients is quite common. The pair therefore is often written as follows:

$$x(t) \xrightarrow{\text{FS}} X[k]$$

4. By applying Euler's[4]formula and taking into account the even/oddness properties of cosine and sine, one can readily see that for real periodic signals the real part of the coefficients X_k is even, the imaginary part is odd. We denote this by saying the coefficients are Hermitian w.r.t. k, i.e.,

$$\forall k \geq 0 : X_{-k} = \overline{X_k} \qquad \text{(only for real, periodic signals!)}$$

Examples

Example 1

To illustrate this, consider the *odd square wave* signal of Figure 5.5(a) with period T and a duty cycle of 50%.

The exponential Fourier coefficients can easily be found to be

$$X_k = \begin{cases} 0.5 & k = 0 \\ 0 & k \text{ even (but not zero)} \\ -\frac{\text{j}}{\pi k} & k \text{ odd} \end{cases}$$

Try to calculate these coefficients yourself! The harmonics diagram is shown in Figure 5.5(b).

Example 2

Another interesting example, which will be of use later on in the course on DSP, is the Fourier series decomposition of a *unit impulse train* with period T (see Figure 5.6(a)). An impulse train is often called a comb function (or more completely a *Dirac comb function*) and is often represented using the letter Sha (see Appendix B): $\text{Ш}_T(t)$.

The exponential Fourier coefficients can easily be found to be

$$X_k = \frac{1}{T}$$

The harmonics diagram is shown in Figure 5.6(b).

Given the Sha-notation and the arrow-notation to represent Fourier series pairs, we can abbreviate this conclusion as:

$$\text{Ш}_T(t) \xrightarrow{\text{FS}} \frac{1}{T}\,\text{Ш}_1[k]$$

If we consider the discrete values of X_k to be Dirac impulses on a continuous frequency axis (as we do when plotting harmonic diagrams), this becomes:

$$\text{Ш}_T(t) \xrightarrow{\text{FS}} \frac{1}{T}\,\text{Ш}_{\omega_1}(\omega)$$

[4]Euler's formula states that $e^{j\phi} = \cos\phi + j\sin\phi$.

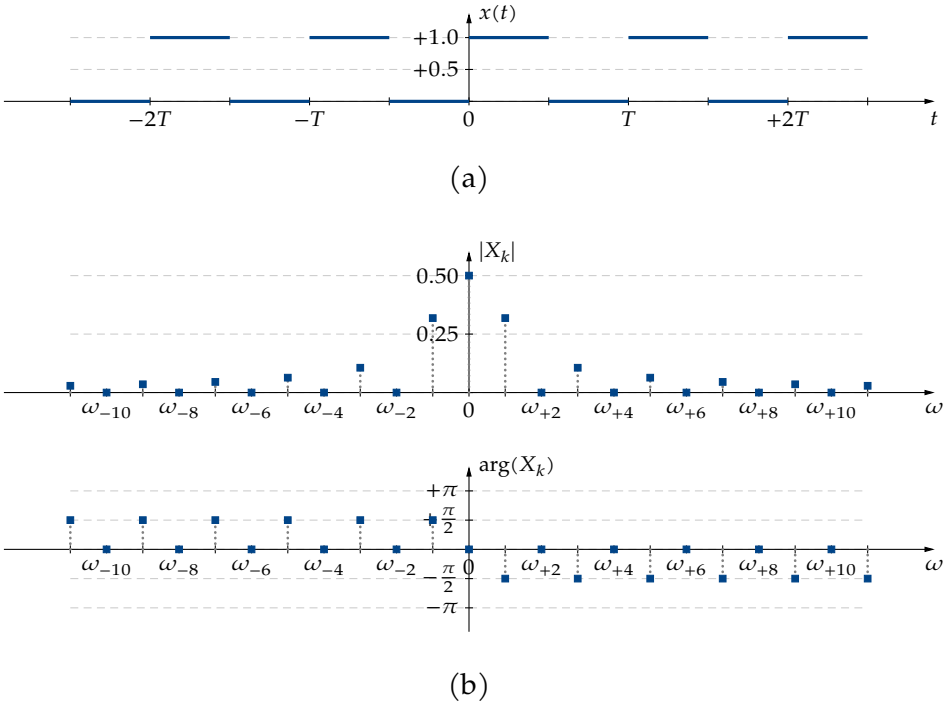


Figure 5.5: Odd square wave with period T and duty cycle of 50%: (a) time-domain representation, (b) harmonics diagram.

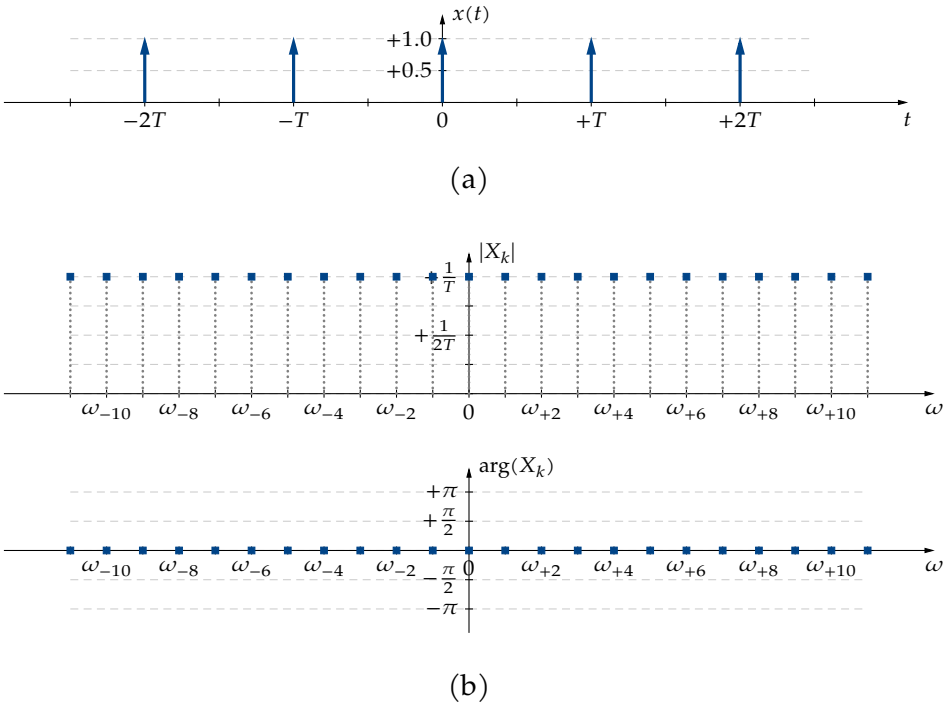


Figure 5.6: Unit pulse train with period T: (a) time-domain representation, (b) harmonics diagram.

The effect of time delay on the Fourier series

In example 1 (above), we have determined the exponential Fourier coefficients for square wave with duty cycle 50% that makes an upgoing jump at $t = 0$ (see Figure 5.5).

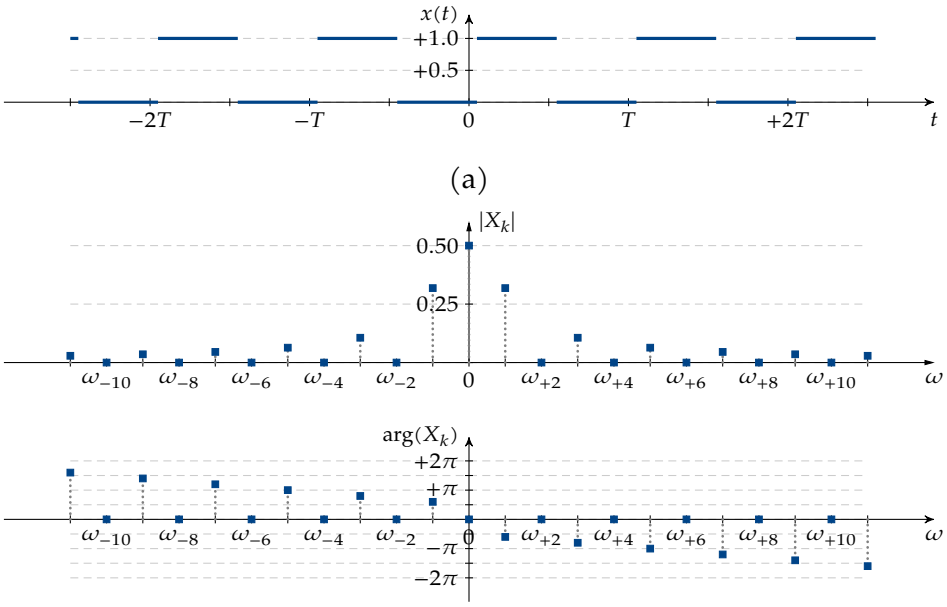
Let's see what happens to the those coefficients if we delay the square wave by an amount αT (with $0 < \alpha < 1$). Calculating the coefficients, yields the following:

$$X_k = \begin{cases} 0.5 & k = 0 \\ 0 & k \text{ even (but not zero)} \\ -\frac{\mathrm{j}}{\pi k} \cdot \mathrm{e}^{-jk\alpha 2\pi} & k \text{ odd} \end{cases}$$

which can be written as:

$$X_k = \mathrm{e}^{-jk\alpha 2\pi} \cdot C_{k,undelayed}$$

in which $C_{k,undelayed}$ corresponds to the coefficients of the series of the undelayed square wave. Therefore, we see that a delay corresponds in the Fourier series domain to multiplying the coefficients with $\mathrm{e}^{-jk\alpha 2\pi}$. This does not change the magnitude of the coefficients, it only adds a linear contribution (linear in k) to the phase of the coefficients. This has been illustrated in the graph below (for $\alpha = 5\%$):



(a)

(b)

Conclusion: we see the effect of time delay as an extra linear term in the phase.

Exercises

Some exercises on exponential Fourier series

Exercise 5.3.3-1:

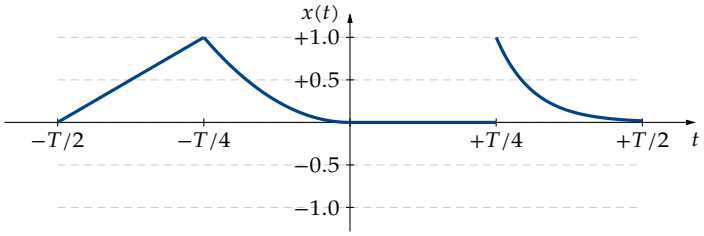
Calculate the exponential Fourier series coefficients of the signal specified in exercise 5.3.1-1.

Exercise 5.3.3-2:

Calculate the exponential Fourier series coefficients of the signal specified in exercise 5.3.1-2.

Exercise 5.3.3-3:
(*) Calculate the exponential Fourier series of the periodic waveform x(t) with period T:

$$x(t) = \begin{cases} \frac{4}{T}\left(t + \frac{T}{2}\right) & \text{if} \quad -\frac{T}{2} \leq t < -\frac{T}{4} \\ \frac{16}{T^2}t^2 & \text{if} \quad -\frac{T}{4} \leq t < 0 \\ 0 & \text{if} \quad 0 \leq t < \frac{T}{4} \\ \mathrm{e}^{-\frac{16}{T}\left(t-\frac{T}{4}\right)} & \text{if} \quad \frac{T}{4} \leq t < \frac{T}{2} \end{cases}$$



Physical interpretation of the exponential series for real periodic signals

It's nice to see that signals can be decomposed in complex exponentials (see section 5.3.3 on page 68). However, this is all rather complicated: can you imagine a complex exponential?

For *real periodic signals*, we can rewrite (5.6) in a simpler form by combining the terms containing equal $|k|$ and using the Hermitian property[5] of the complex Fourier coefficients X_k (see section 5.3.3 on page 68). You can see how it's done below:

$$\begin{aligned} x(t) &= \sum_{k=-\infty}^{+\infty} X_k \, \mathrm{e}^{\mathrm{j}\omega_k t} \\ &= X_0 + \sum_{k=1}^{+\infty} \left(X_k \, \mathrm{e}^{\mathrm{j}\omega_k t} + X_{-k} \, \mathrm{e}^{-\mathrm{j}\omega_k t} \right) \\ &\downarrow \; X_k = |X_k| \, \mathrm{e}^{\mathrm{j}\phi_k} \;\text{ and }\; X_{-k} = |X_k| \, \mathrm{e}^{-\mathrm{j}\phi_k} \\ &= X_0 + \sum_{k=1}^{+\infty} |X_k| \cdot \left(\mathrm{e}^{\mathrm{j}(\omega_k t + \phi_k)} + \mathrm{e}^{-\mathrm{j}(k\omega_1 t + \phi_k)} \right) \\ &\downarrow \; \cos\alpha = \frac{\left(\mathrm{e}^{\mathrm{j}\alpha} + \mathrm{e}^{-\mathrm{j}\alpha} \right)}{2} \\ &= X_0 + \sum_{k=1}^{+\infty} 2|X_k| \cos(\omega_k t + \phi_k) \end{aligned}$$

This leads us to an interpretable cosine decomposition for real periodic signals. The cosines that appear in this equation are called the *harmonics* of a signal. The signal thus is composed of a DC component equaling X_0, and a first, second, third, ...order harmonic with magnitude $2\,|X_k|$ and a phase lead of $\phi_k = \arg(X_k)$.

[5] Remember: Hermitian denotes even real parts and odd imaginary parts.

5.3.4 Conversion between exponential and trigonometric coefficients

One can go back and forth between the exponential and trigonometric representation using the following equations:

Exponential to trigonometric

$$\begin{aligned} A_k &= X_k + X_{-k} && k \in \mathbb{Z}^+ \\ B_k &= \mathrm{j}(X_k - X_{-k}) && k \in \mathbb{Z}_0^+ \end{aligned}$$

Trigonometric to exponential

$$\begin{aligned} X_0 &= \frac{A_0}{2} \\ X_k &= \frac{A_k - \mathrm{j}B_k}{2} && k \in \mathbb{Z}_0^+ \\ X_{-k} &= \frac{A_k + \mathrm{j}B_k}{2} && k \in \mathbb{Z}_0^+ \end{aligned}$$

Exercises

A. Some exercises on converting trigonometric into exponential harmonic numbers

Exercise 5.3.4-1:
Use the conversion formulae to convert the trigonometric harmonic numbers of exercise 5.3.1-1 into exponential harmonic numbers. If you solved 5.3.3-1 before, you can crosscheck your results.

Exercise 5.3.4-2:
Use the conversion formulae to convert the trigonometric harmonic numbers of exercise 5.3.1-2 into exponential harmonic numbers. If you solved 5.3.3-2 before, you can crosscheck your results.

B. Some exercises on converting exponential into trigonometric harmonic numbers

Exercise 5.3.4-3:
Use the conversion formulae to convert the exponential harmonic numbers of exercise 5.3.3-1 into trigonometric harmonic numbers. If you solved 5.3.1-1 before, you can crosscheck your results.

Exercise 5.3.4-4:
Use the conversion formulae to convert the exponential harmonic numbers of exercise 5.3.3-2 into trigonometric harmonic numbers. If you solved 5.3.1-2 before, you can crosscheck your results.

Exercise 5.3.4-5:
Try to prove the relationship between trigonometric and exponential Fourier coefficients, by starting from the exponential series, reworking that series into a trigonometric one (using Euler's formula).

5.3.5 Parseval's theorem

A theorem that we will not prove (but you can try yourself, based on your mathematical background) is Parseval's theorem. We distinguish two forms: one for the trigonometric series and one for the exponential series:

Parseval's theorem for the trigonometric Fourier series

Assuming $x(t) \sim \frac{A_0}{2} + \sum_{k=1}^{+\infty} A_k \cos \omega_k t + \sum_{k=1}^{+\infty} B_k \sin \omega_k t$, we can write the following identity:

$$\int_0^T |x(t)|^2 \, \mathrm{d}t = \frac{T}{2}\left[\frac{A_0^2}{2} + \sum_{n=1}^{+\infty}\left(A_n^2 + B_n^2\right)\right]$$

Parseval's theorem for the exponential Fourier series

Assuming $x(t) \sim \sum_{k=-\infty}^{+\infty} X_k \, \mathrm{e}^{\mathrm{j}\omega_k t}$, we can write the following identity:

$$\int_0^T |x(t)|^2 \, \mathrm{d}t = T \cdot \sum_{k=-\infty}^{+\infty} |X_k|^2$$

As the left part of the equation is the energy in one period of the signal, Parseval allows us to calculate that same energy using the Fourier coefficients.

5.3.6 Spectra

We often make drawings of the Fourier coefficients. We typically draw one of two things:

- a drawing of the magnitude and phase of X_k as a function of k, which we call the *magnitude* and *phase spectrum*
- a drawing of $|X_k|^2$ as a function of k, which we call the *power spectrum* (or more correctly the *power mass spectrum*)

The fact that we call the latter a power spectrum is because of Parseval's theorem, which states that:

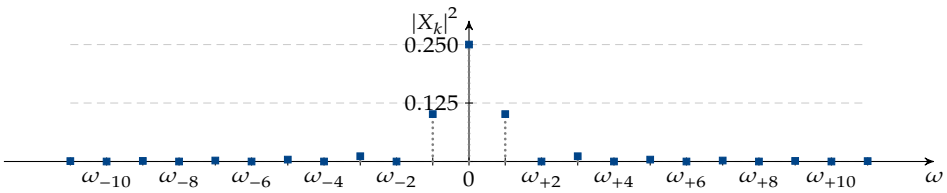
$$\int_0^T |x(t)|^2 \, \mathrm{d}t = T \cdot \sum_{k=-\infty}^{+\infty} |X_k|^2$$
$$\Leftrightarrow \quad \frac{1}{T}\int_0^T |x(t)|^2 \, \mathrm{d}t = \sum_{k=-\infty}^{+\infty} |X_k|^2$$

The left hand now contains the average power in a single period of the signal. The right hand part shows us that the squares of the magnitudes of the Fourier coefficients show us how that

power is divided over the different sinors that are part of the series.

Therefore making a drawing of these squares of magnitudes makes sense. Sometimes we use k on the X-axis, sometimes we use $k\frac{2\pi}{T}$ on the X-axis.

You can find an example of a magnitude and phase spectrum in Figure 5.5). The corresponding power (mass) spectrum has been drawn below:



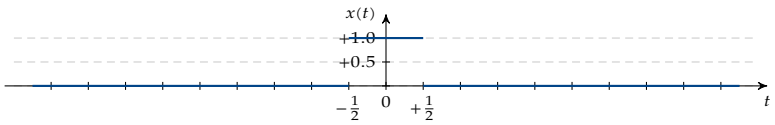
5.4 The Fourier transform

5.4.1 Derivation

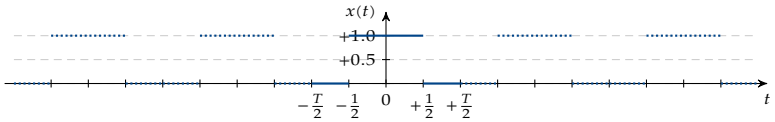
Now, let's take a look at aperiodic continuous signals. To start, we are going to assume that the aperiodic signal we want to analyze has a limited footprint. This ensures that we can package it as a periodic signal, by taking the footprint (and even a little bit more) as the primary period and making a periodic extension of that signal. Then we will calculate its Fourier series and see what happens to the Fourier coefficients if we increase the size of the period.

Graphical

A picture says more than a thousand words, so let's use a simple block pulse as our poster child example.



Now let's make it periodic:



The Fourier coefficients of this periodic signal are:

$$X_k = \frac{\sin\left(\frac{k\pi}{T}\right)}{k\pi} = \frac{1}{T}\,\mathrm{sinc}\left(\frac{k\pi}{T}\right) \tag{5.8}$$

Check what happens to these Fourier coefficients as we increase T in Figure 5.7. We can observe a few effects:

- the overall shape of the Fourier coefficients as a function of k stays the same, however
- the spectral components move closer together for increasing T, and
- the magnitude of the function diminishes for increasing T.

For $T \to +\infty$ the end result will be that all Fourier coefficients are zero, i.e. all information is gone. Looking back at (5.9), it is the first factor $1/T$ that is the culprit. To avoid the effect, we will work with modified Fourier coefficients, defined as:

$$X'_k = T \cdot X_k$$

In our example, this results in:

$$X'_k = \text{sinc}\left(\frac{k\pi}{T}\right) \tag{5.9}$$

The effect of this countermeasure can be observed in Figure 5.8.

Coming back to the second observation we made earlier, it is not hard to imagine that if $T \to \infty$ the curve of the Fourier coefficients will become a continuous line instead of separate points.

Mathematical

Let's derive the equations that describe this continuous line. We start by reconsidering the Fourier series equations (5.6) and (5.7):

$$\begin{cases} x(t) = \sum_{k=-\infty}^{+\infty} X_k \, \mathrm{e}^{\mathrm{j}\omega_k t} \\ X_k = \frac{1}{T} \int_{-T/2}^{T/2} x(t) \, \mathrm{e}^{-\mathrm{j}\omega_k t} \, \mathrm{d}t \end{cases}$$

with as usual $\omega_1 = \frac{2\pi}{T}$ and $\omega_k = k\omega_1$.

As mentioned earlier, the factor $1/T$ in (5.7) will make the coefficients X_k vanish as T goes to plus infinity. We will therefore use the modified Fourier coefficients as we defined them earlier:

$$X'_k = T \cdot X_k$$

This leads to the following set of equations:

$$\begin{cases} x(t) = \frac{1}{T} \sum_{k=-\infty}^{+\infty} X'_k \, \mathrm{e}^{\mathrm{j}\omega_k t} \\ X'_k = \int_{-T/2}^{T/2} x(t) \, \mathrm{e}^{-\mathrm{j}\omega_k t} \, \mathrm{d}t \end{cases}$$

Now, let's define $\Delta\omega = \frac{2\pi}{T}$ (the frequency distance in between two harmonics), and replace the factor $\frac{1}{T}$ by $\frac{\Delta\omega}{2\pi}$. This leads to

$$\begin{cases} x(t) = \frac{1}{2\pi} \sum_{k=-\infty}^{+\infty} X'_k \, \mathrm{e}^{\mathrm{j}\omega_k t} \Delta\omega \\ X'_k = \int_{-T/2}^{T/2} x(t) \, \mathrm{e}^{-\mathrm{j}\omega_k t} \, \mathrm{d}t \end{cases}$$

in which we moved the factor $\Delta\omega$ to the back.

Now, let's consider what happens if we keep increasing T:

$$\begin{aligned} T &\to +\infty \\ \Delta\omega &\to \mathrm{d}\omega \\ \omega_k &\to \omega \end{aligned}$$

What does this mean? The Fourier series harmonics X_k come closer and closer together until they become a continuous spectrum $X(\omega)$. By consequence, the Fourier *series* becomes a Riemann sum, i.e. in the limit an integral. This leads us to the definition of the Fourier transform.

5.4.2 Definition

Fourier transform

$$\begin{cases} x(t) = \dfrac{1}{2\pi} \displaystyle\int_{-\infty}^{+\infty} X(\omega)\, \mathrm{e}^{\mathrm{j}\omega t}\, \mathrm{d}\omega \\ X(\omega) = \displaystyle\int_{-\infty}^{+\infty} x(t)\, \mathrm{e}^{-\mathrm{j}\omega t}\, \mathrm{d}t \end{cases} \tag{5.12}$$

The latter equation is the so-called *Fourier transform*, the former is the so-called *inverse Fourier transform*.

Remarks

1. The units of the Fourier transform are signal units multiplied with seconds (as you integrated the signal over time), or signal units per Hertz, e.g.

$$[x(t)] = \mathrm{V} \quad \Rightarrow \quad [X(\omega)] = \mathrm{V\,s} = \mathrm{V/Hz} \text{ or } \left[\frac{X(\omega)}{2\pi}\right] = \mathrm{V/(rad/s)}$$

2. The Fourier transform is often symbolized as an operator $\mathcal{F}$.

$$X(\omega) = \mathcal{F}(x(t)) \qquad\qquad x(t) = \mathcal{F}^{-1}(X(\omega))$$

3. The use of lowercase symbols for time-domain signals and uppercase symbols for their Fourier transform is quite common. The pair is also often written as follows:

$$x(t) \overset{\mathcal{F}}{\to} X(\omega)$$

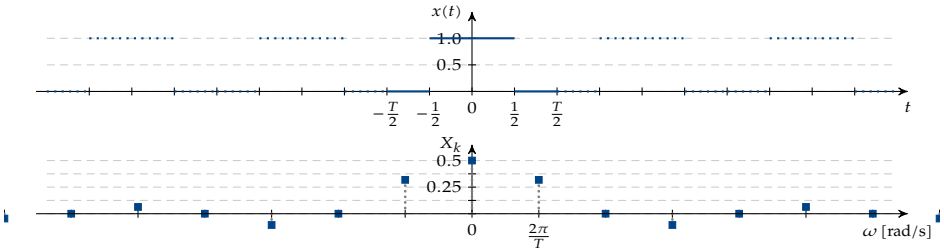
4. If the signal $x(t)$ is real, the Fourier transform is Hermitian, i.e.

$$\forall \omega \in \mathbb{R} : X(-\omega) = \overline{X(\omega)}$$

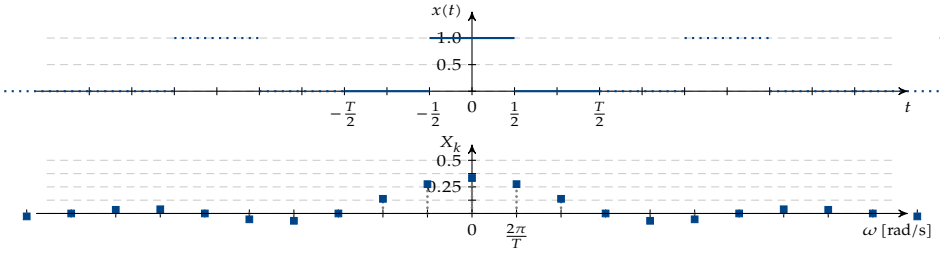
   You can easily prove this yourself by applying Euler's formula and taking into account the even/oddness properties of cosine and sine. A different way to describe the Hermiticity, is that the magnitude spectrum is even and the phase spectrum is odd:

$$|H(\omega)| = |H(-\omega)| \qquad\qquad \arg H(\omega) = -\arg H(-\omega)$$

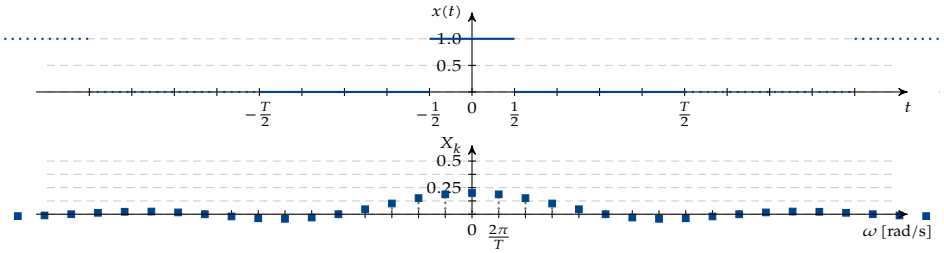
5. How does the Fourier transform compare to the Fourier series? The shuffling with the factor $1/T$ and increasing T to infinity kind of clutters the relationship. However, the relationship between them is really very simple.



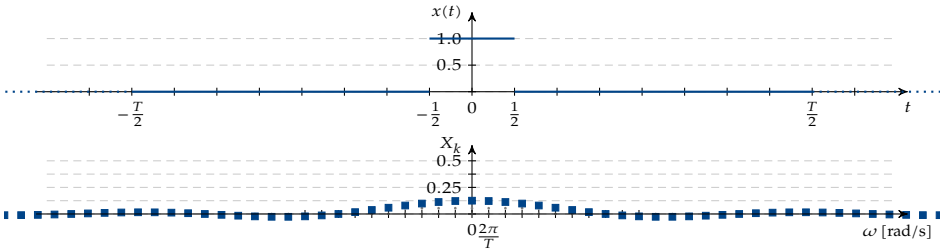
(a) $T = 2.0$



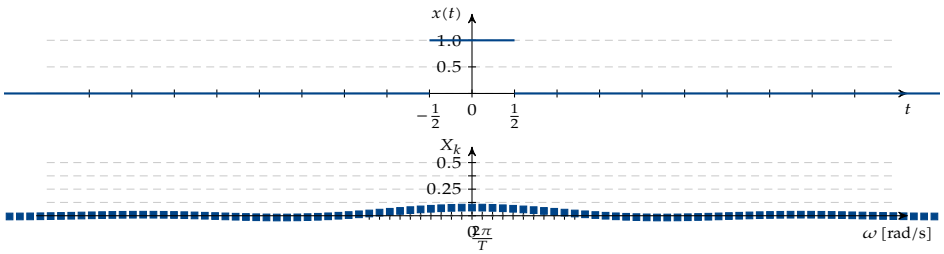
(b) $T = 3.0$



(c) $T = 5.0$

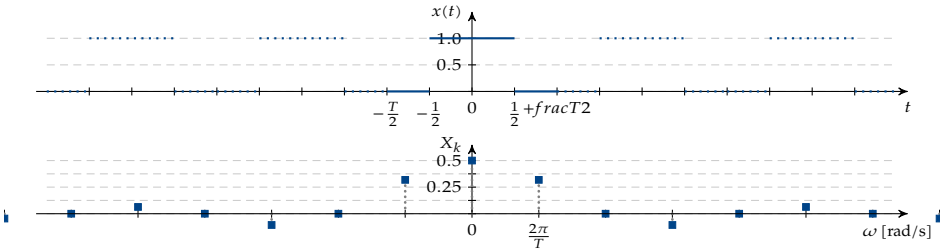


(d) $T = 8.0$

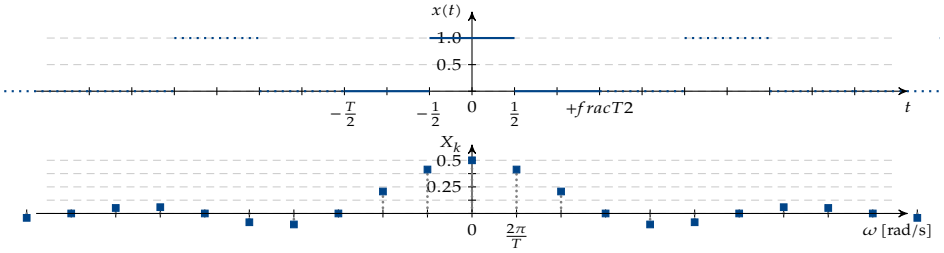


(e) $T = 13.0$

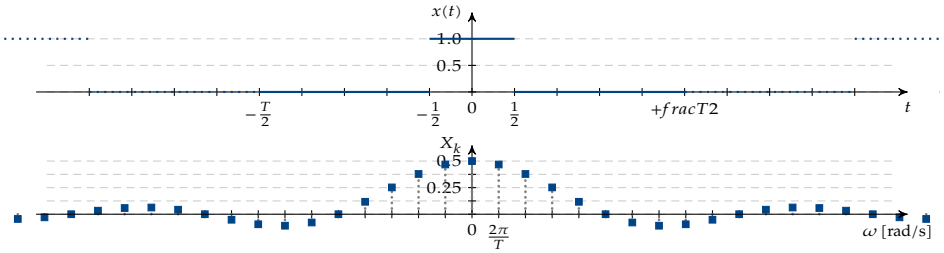
Figure 5.7: Evolution of the Fourier coefficients if T increases.



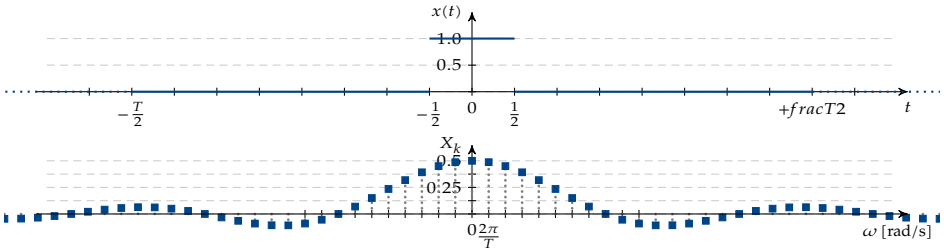
(a) $T = 2.0$



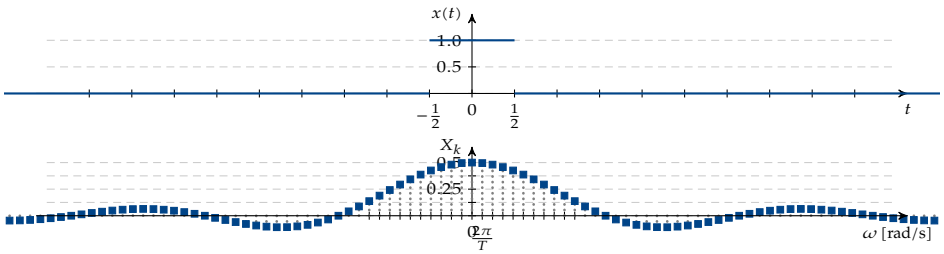
(b) $T = 3.0$



(c) $T = 5.0$



(d) $T = 8.0$



(e) $T = 13.0$

Figure 5.8: Evolution of the modified Fourier coefficients if T increases.

If, given the Fourier coefficients, we define:

$$X(\omega) = 2\pi \sum_{k=-\infty}^{+\infty} X_k \, \delta(\omega - \omega_k) \tag{5.13}$$

then it is easy to show that the inverse Fourier transform of $X(\omega)$ equals $x(t)$. Try!

This is also the reason why many people use Dirac impulses when graphing the harmonic numbers of a Fourier series. Alas, many of them forget about the factor 2π. Don't you!

The Fourier transform is often graphed in its polar form, which is called the *frequency spectrum* of the signal. An aperiodic continuous-time signal has a continuous frequency spectrum.

5.4.3 Examples

Some examples (that will be of use later on) can be found below.

Example 1

Let's take a new look at our *unit pulse train* (see Figure 5.9(a)). We can derive its Fourier transform by using the Fourier decomposition we derived earlier (in section 5.3.3) and applying (5.13), leading to:

$$X(\omega) = \frac{2\pi}{T} \sum_{k=-\infty}^{+\infty} \delta(\omega - \omega_k)$$

with

$$\omega_k = k\frac{2\pi}{T}$$

The frequency spectrum of this signal can be found in Figure 5.9(b).

This result is so fundamental that it is worth memorizing. The Sha notation gives us a tool to keep the focus on the important details:

$$\text{Ш}_T(t) \xrightarrow{\mathcal{F}} \omega_1 \, \text{Ш}_{\omega_1}(\omega)$$

or even better in a form that will be ready for use later on. At that time we will replace T by T_s (as it will represent the sample period) and ω_1 by ω_s (as it will represent the (angular) sample frequency).

$$T_s \, \text{Ш}_{T_s}(t) \xrightarrow{\mathcal{F}} 2\pi \, \text{Ш}_{\omega_s}(\omega)$$

Example 2

How does the Fourier transform of a *window function* (see Figure 5.10 on the next page) look

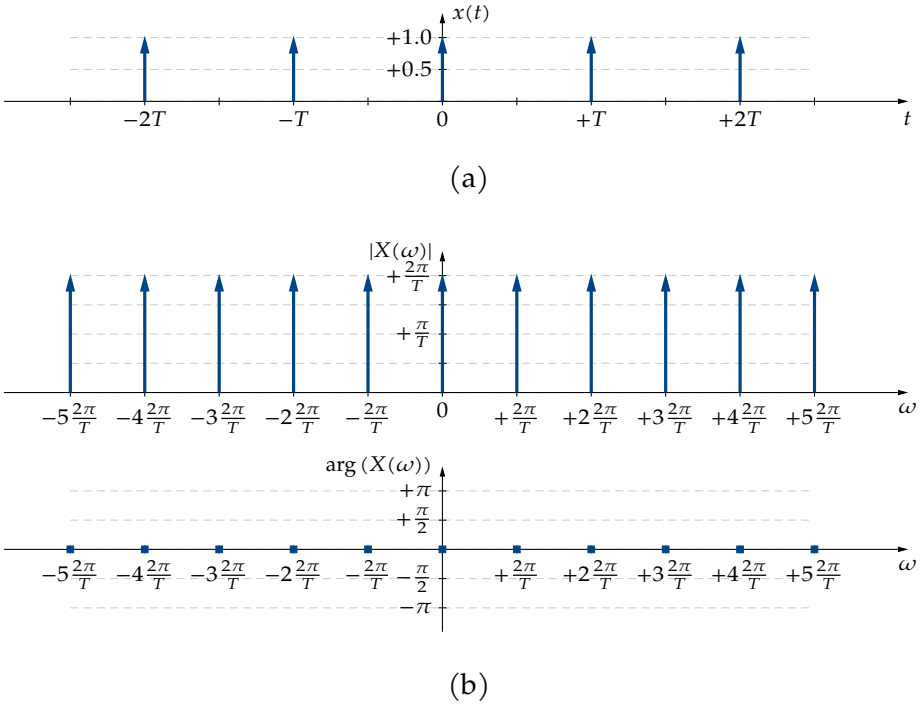


Figure 5.9: Unit pulse train with period T: (a) time-domain representation, (b) frequency spectrum.

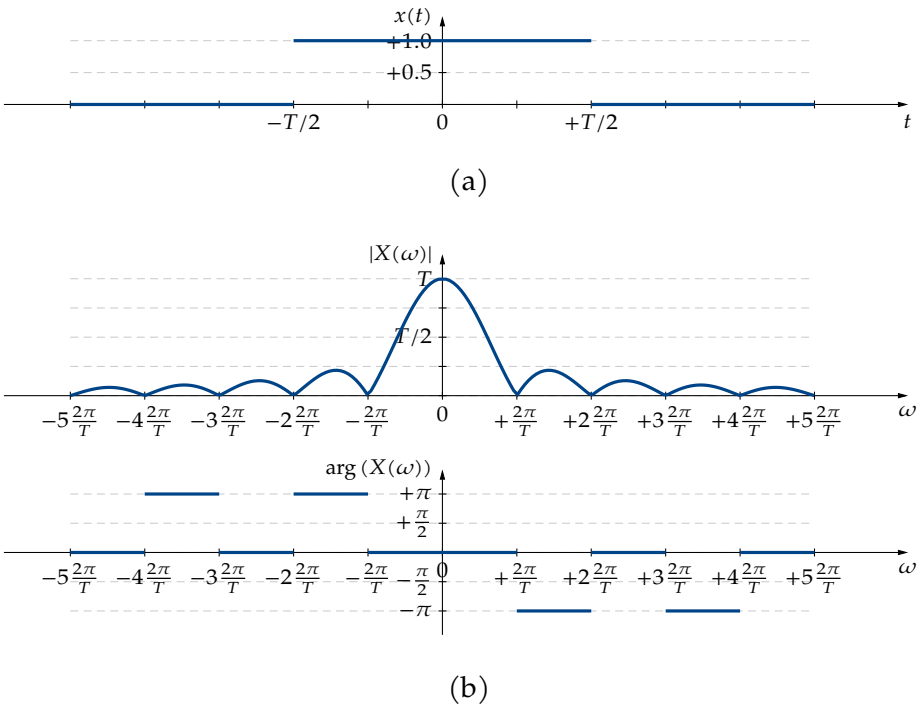


Figure 5.10: Symmetrical window function with length T: (a) time-domain representation, (b) frequency spectrum.

like? The derivation deserves a little more attention:

$$\begin{aligned} X(\omega) &= \int_{-\infty}^{+\infty} x(t)\, \mathrm{e}^{-\mathrm{j}\omega t}\, \mathrm{d}t \\ &= \int_{-T/2}^{T/2} \mathrm{e}^{-\mathrm{j}\omega t}\, \mathrm{d}t \\ &= \frac{1}{-\mathrm{j}\omega} \left[\mathrm{e}^{-\mathrm{j}\omega t} \right]_{t=-T/2}^{t=T/2} \\ &= \frac{2}{\omega} \left(\frac{\mathrm{e}^{\mathrm{j}\frac{\omega T}{2}} - \mathrm{e}^{-\mathrm{j}\frac{\omega T}{2}}}{2\mathrm{j}} \right) \\ &= \frac{2}{\omega} \sin\left(\frac{\omega T}{2} \right) \\ &= T \frac{\sin\left(\frac{\omega T}{2} \right)}{\frac{\omega T}{2}} = T\, \mathrm{sinc}\left(\frac{\omega T}{2} \right) \end{aligned}$$

Drawing the frequency spectrum yields Figure 5.10(b).

Example 3

Let's calculate the Fourier transform of a simple *sine wave* with period T (see Figure 5.11(a)). Let's define $\omega_1 = \frac{2\pi}{T}$.

$$\begin{aligned} X(\omega) &= \int_{-\infty}^{+\infty} \sin(\omega_1 t)\, \mathrm{e}^{-\mathrm{j}\omega t}\, \mathrm{d}t \\ &= \int_{-\infty}^{+\infty} \frac{\mathrm{e}^{\mathrm{j}\omega_1 t} - \mathrm{e}^{-\mathrm{j}\omega_1 t}}{2\mathrm{j}}\, \mathrm{e}^{-\mathrm{j}\omega t}\, \mathrm{d}t \\ &\Big\downarrow\ \mathrm{e}^{\mathrm{j}\omega_1 t} \xrightarrow{\mathcal{F}} 2\pi\, \delta(\omega - \omega_1) \\ &= \frac{\pi}{\mathrm{j}} \left(\delta(\omega - \omega_1) - \delta(\omega + \omega_1) \right) \end{aligned}$$

You can find the sine's frequency spectrum in Figure 5.11(b).

Example 4

The Fourier transform of a simple *cosine wave* with period T (see Figure 5.12(a)) can be obtained in similar way as for the sine wave, leading to

$$X(\omega) = \pi \left(\delta(\omega - \omega_1) + \delta(\omega + \omega_1) \right)$$

You can find the cosine's frequency spectrum in Figure 5.12(b).

Exercises

Some exercises on the Fourier transform

Exercise 5.4.3-1:

Calculate the Fourier transform of the signal $x(t)$ using the definition of (5.12), with

$$x(t) = \mathrm{e}^{-\frac{t^2}{2\sigma^2}}$$

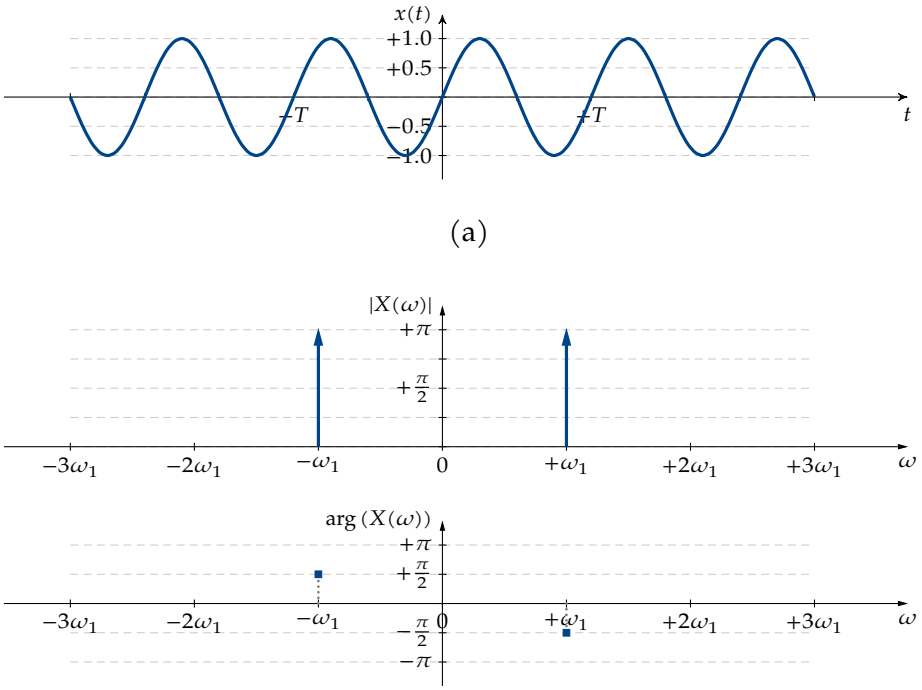


Figure 5.11: Sine with period T ($\omega_1 = \frac{2\pi}{T}$): (a) time-domain representation, (b) frequency spectrum.

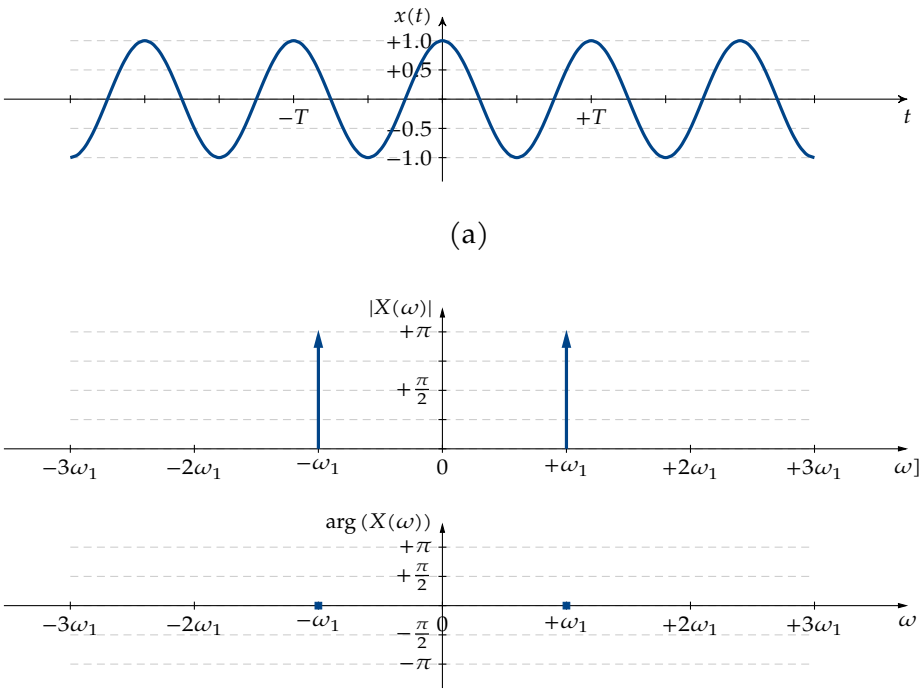


Figure 5.12: Cosine with period T ($\omega_1 = \frac{2\pi}{T}$): (a) time-domain representation, (b) frequency spectrum.

Exercise 5.4.3-2:
Calculate the Fourier transform of the signal $x(t)$ using the definition of (5.12), with

$$x(t) = \mathrm{u}(t)\ \mathrm{e}^{-\alpha t}$$

with $\alpha \in \mathbb{R}_0^+$.

Exercise 5.4.3-3:
Calculate the Fourier transform of the signal $x(t)$ using the definition of (5.12), with

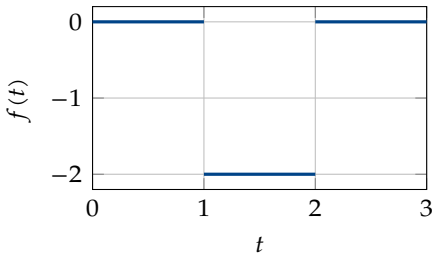
$$x(t) = \mathrm{u}(t)\, t\, \mathrm{e}^{-\alpha t}$$

with $\alpha \in \mathbb{R}_0^+$.

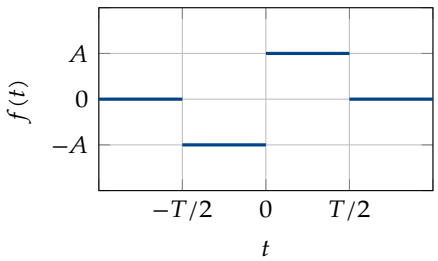
Exercise 5.4.3-4:
Calculate the Fourier transform of the signal $x(t)$ as specified in exercise 5.3.3-1.

Exercise 5.4.3-5:
Calculate the Fourier transform of the signal $x(t)$ as specified in exercise 5.3.3-2.

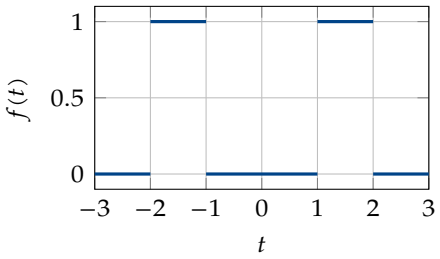
Exercise 5.4.3-6:
Determine the Fourier transform of the pulse $f(t)$:



Exercise 5.4.3-7:
Find the Fourier transform of the waveform $f(t)$:



Exercise 5.4.3-8:
Determine the Fourier transform of the waveform $f(t)$:



Exercise 5.4.3-9:
Determine the Fourier transform of a signal with $f(t) = At/B$ between $t = 0$ and $t = B$ and $f(t) = 0$ elsewhere. First, make a drawing of the signal.

Exercise 5.4.3-10:
Determine the Fourier transform of $f(t) = 10\sin(50t)$.

Exercise 5.4.3-11:
Find the Fourier transform of $f(t) = A\ \mathrm{e}^{-at}\,\mathrm{u}(t)$ with $a \in \mathbb{R}_0^+$ and $A \in \mathbb{R}$. First, make a drawing of the signal.

Exercise 5.4.3-12:
Find the Fourier transform of the function $f(t) = -\,\mathrm{u}(-t) + \mathrm{u}(t)$.
This function is often denoted as $\mathrm{sgn}(t)$, the so-called *signum* function. First, make a drawing of the function and find a logical explanation as to why we call this the *signum* function.

Exercise 5.4.3-13:
Determine the Fourier transform of $f(t) = \mathrm{u}(t)\cos(\omega_0 t)$ with $\omega_0 \in \mathbb{R}_0$.

Exercise 5.4.3-14:
Find the Fourier transform of $f(t) = A\ \mathrm{u}(t)$ with $A \in \mathbb{R}_0$.

5.4.4 Properties

The Fourier transform has many interesting properties. The derivation of these properties is often very simple (don't hesitate to challenge yourself and try some of them!). However, knowing the properties, their significance and being able to use them is much more important. Therefore, let's focus on the properties themselves.

Let's start by making the following assumptions:

$$x(t) \xrightarrow{\mathcal{F}} X(\omega)$$
$$y(t) \xrightarrow{\mathcal{F}} Y(\omega)$$
$$a, b \in \mathbb{R}$$
$$k \in \mathbb{R}_0$$

Knowing this, we can move to the properties themselves:

Linearity

$$ax(t) + by(t) \xrightarrow{\mathcal{F}} aX(\omega) + bY(\omega)$$

This property allows us to easily calculate the Fourier transform of signals that are linear combinations of other known signals. However, it also tells us that any linear combination of two signals just combines their frequency spectra in the same way.

Time-frequency symmetry

$$X(t) \xrightarrow{\mathcal{F}} 2\pi x(-\omega)$$

This property is a direct consequence of the similarity between the Fourier transform and the inverse Fourier transform. It allows us to calculate a new Fourier pair, starting from an existing one.

Time/Frequency scaling

$$x(kt) \xrightarrow{\mathcal{F}} \frac{1}{|k|} X\left(\frac{\omega}{k}\right)$$
$$\frac{1}{|k|} x\left(\frac{t}{k}\right) \xrightarrow{\mathcal{F}} X(k\omega)$$

This seems a complex property. However, it bears some clear messages:

- for aperiodic time-limited signals: short (sharp) time domain signals will exhibit a wide-band frequency spectrum, while smearing out (smoothing) the same time-domain signal over a longer period of time, will cause more narrow-band banded frequency spectra. The extreme example of this, is the Dirac impulse, having a maximally spread frequency content, and the opposite, a constant signal in time having a Dirac impulse as its frequency spectrum.
- for periodic signals: increasing the period, means decreasing the frequency and vice versa.

Time/Frequency shifting

$$x(t - t_0) \xrightarrow{\mathcal{F}} X(\omega)\, \mathrm{e}^{-\mathrm{j}\omega t_0}$$
$$x(t)\, \mathrm{e}^{+\mathrm{j}\omega_0 t} \xrightarrow{\mathcal{F}} X(\omega - \omega_0)$$

Let's start with the time shifting.
Delaying a signal, means adding an extra ωt_0 phase lag in the spectrum. This seems a rather unimportant property. However, the opposite tells the real story: if we're able to make a system that changes the phase of the incoming signal only by adding a linear component to the phase (a "linear phase filter"), then (except for the amplitude shaping) all frequency components will be delayed by an equal time amount, i.e. the overall shape of the signal will be maintained. In such a system, a step input will appear at the output with the same roundings at the start as

well as at the end of the transition. This is an important time-domain property. Linear phase filtering can't be accomplished in the analog domain. This is one of the areas where DSP shows its true power: it can accomplish things that can't be done using analog electronics alone.[6]

The frequency shifting is the base of frequency up- and downconversion. We will come back to this later.

Time/Frequency differentiation

$$\frac{\mathrm{d}}{\mathrm{d}t}x(t) \xrightarrow{\mathcal{F}} \mathrm{j}\omega X(\omega)$$
$$-\mathrm{j}tx(t) \xrightarrow{\mathcal{F}} \frac{\mathrm{d}}{\mathrm{d}\omega}X(\omega)$$

Time differentiation is an important operation in (analog) signal processing, as the device equations of many electronic devices (e.g., capacitor, inductor) contain derivatives. The factor j introduces an extra phase lead of $\pi/2$. Considering $\frac{\mathrm{d}}{\mathrm{d}t}\sin(\omega_0 t) = \omega_0 \cos(\omega_0 t)$, this property is not hard to understand. The frequency differentiation property is no more than the dual property of the time-domain differentiation.

Parseval's theorem

$$\int_{-\infty}^{+\infty} x(t)\overline{y(t)}\,\mathrm{d}t = \frac{1}{2\pi}\int_{-\infty}^{+\infty} X(\omega)\overline{Y(\omega)}\,\mathrm{d}\omega$$

Parseval's theorem in its general form looks somewhat abstract, but let's substitute $y(t)$ by $x(t)$, Parseval tells us that:

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

This means that we can calculate the energy in a signal, by integrating the squared magnitude of the frequency spectrum. So, phase does not alter the energy contents of a signal! That's quite a conclusion.
It also means that there's no energy in the frequency cross products. This due to the fact that the Fourier transform is an *orthogonal decomposition*.

Convolution/Multiplication

$$x(t) \star y(t) \xrightarrow{\mathcal{F}} X(\omega)Y(\omega)$$
$$x(t)y(t) \xrightarrow{\mathcal{F}} \frac{1}{2\pi}X(\omega) \star Y(\omega)$$

There's little to comment on this property. Though it seems complicated (convolution, what's that?), it shows how filtering in the frequency domain translates itself into the time domain. Read on, and you will learn that this is one of the properties that allows us to understand the concept of frequency transfer function, in relation to convolution with the impulse response of a system.

[6] Actually, this statement is somewhat incorrect as all electronics (including all digital signal processors) can be considered to be based on principles of analog electronics. The correct statement is that linear-phase filtering cannot be accomplished without using discrete-time signals (and discrete mathematics).

Exercises
Some exercises on the properties of the Fourier transform

Exercise 5.4.4-1:
Apply the time-frequency symmetry property to the following transform pair:

$$\delta(t) \xrightarrow{\mathcal{F}} 1$$

Exercise 5.4.4-2:
Knowing that

$$\sin(\omega_0 t) \xrightarrow{\mathcal{F}} \mathrm{j}\pi\left(\delta(\omega+\omega_0)-\delta(\omega-\omega_0)\right),$$

with $\omega_0 \in \mathbb{R}_0^+$, apply the time differentiation property to calculate the Fourier transform of:

$$\cos(\omega_0 t)$$

Exercise 5.4.4-3:
Knowing that

$$\cos(\omega_0 t) \xrightarrow{\mathcal{F}} \pi\left(\delta(\omega-\omega_0)+\delta(\omega+\omega_0)\right),$$

with $\omega_0 \in \mathbb{R}_0^+$, apply the time scaling property to calculate the Fourier transform of:

$$\cos\left(\frac{\omega_0}{10}t\right)$$

Exercise 5.4.4-4:
Knowing that

$$\mathrm{e}^{-\alpha t}\,\mathrm{u}(t) \xrightarrow{\mathcal{F}} \frac{1}{\mathrm{j}\omega+\alpha},$$

apply the time shifting property to calculate the Fourier transform of the delayed pulse:

$$\mathrm{e}^{-\alpha(t-t_0)}\,\mathrm{u}(t-t_0)$$

Exercise 5.4.4-5:
Knowing that

$$\mathrm{sgn}(t) \xrightarrow{\mathcal{F}} \frac{2}{\mathrm{j}\omega}$$

apply the time differentiation property to calculate the Fourier transform of:

$$\delta(t)$$

5.5 Description of LTI systems in the Fourier domain

The Fourier transform can be applied to any signal for which the Fourier integral converges. It certainly is not restricted to signals of LTI systems. However, the Fourier transform is especially useful for LTI systems. We will learn about that in this section.

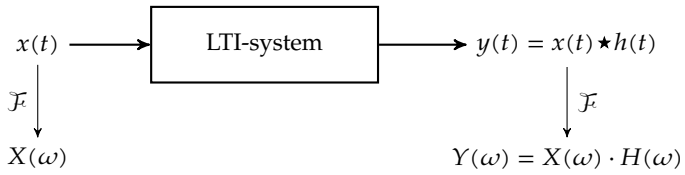
5.5.1 The frequency transfer function

Deriving the concept

From studying LTI systems and the convolution, we've learned that we can calculate the output of an LTI system by calculating the convolution of the input with its impulse response:

$$y(t) = x(t) \star h(t) = \int_{-\infty}^{+\infty} x(\tau) h(t-\tau)\, \mathrm{d}t$$

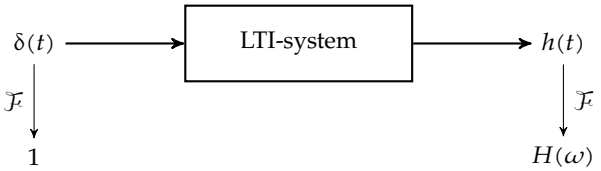
Given the convolution property of the Fourier transform (see page 87), this means:



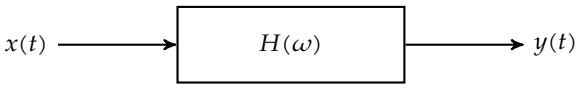
This allows for the creation of a very useful new concept, the *frequency transfer function*:

$$H(\omega) = \frac{Y(\omega)}{X(\omega)}$$

This frequency transfer function is not only the ratio (the *gain* from input to output), it is the Fourier transform of the impulse resonse, i.e. if you excite the system with a Dirac impulse, the response in the frequency domain will be the frequency transfer function.



Because the frequency transfer function $H(\omega)$ is a full description of the LTI-system (as was the impulse response), we often just write an H inside the box of an LTI-system to indicate that it is LTI with a frequency transfer function H:



Definition

In this way we come to the definition of the *frequency transfer function*:

Frequency transfer function

For an LTI-system the *frequency transfer function* $H(\omega)$ is the Fourier transform of the impulse response $h(t)$:

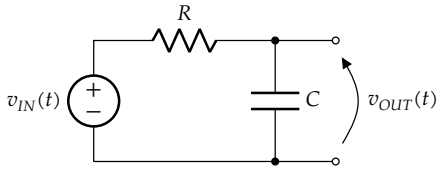
$$h(t) \xrightarrow{\mathcal{F}} H(\omega)$$

It describes the (linear) relationship between the input x and the output y of an LTI system in the frequency domain:

$$H(\omega) = \frac{Y(\omega)}{X(\omega)} \qquad \Leftrightarrow \qquad Y(\omega) = H(\omega) \cdot X(\omega)$$

Example

Consider the filter below:



We calculated the impulse response of this filter earlier (using v_{IN} as input and v_{OUT} as output) and found it to be:

$$h(t) = \mathrm{u}(t) \cdot \frac{1}{RC} \mathrm{e}^{-\frac{t}{RC}}$$

The frequency transfer function can be calculated as follows:

$$\begin{aligned} H(\omega) = \mathcal{F}(h(t)) &= \int_{-\infty}^{+\infty} \mathrm{u}(t) \frac{1}{RC} \mathrm{e}^{-\frac{t}{RC}} \mathrm{e}^{-\mathrm{j}\omega t} \,\mathrm{d}t \\ &= \frac{1}{RC} \int_{0}^{+\infty} \mathrm{e}^{-\frac{1+\mathrm{j}\omega RC}{RC} t} \,\mathrm{d}t \\ &= \frac{1}{RC} \left(-\frac{RC}{1+\mathrm{j}\omega RC} \left[\mathrm{e}^{-\frac{1+\mathrm{j}\omega RC}{RC} t} \right]_0^{+\infty} \right) \end{aligned}$$

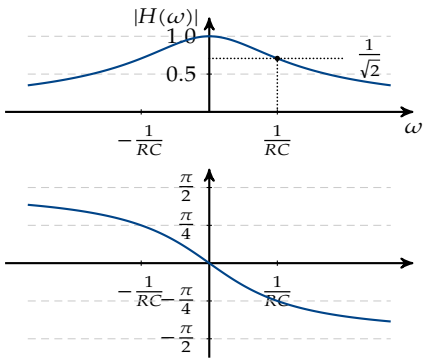
Note that when filling out the upper limit in the expression above, the imaginary part of the argument of the exponential is of no importance as the real part squeezes the magnitude of the exponential down to zero. This leads to:

$$\begin{aligned} H(\omega) &= \frac{1}{RC} \left(-\frac{RC}{1+\mathrm{j}\omega RC} (0-1) \right) \\ &= \frac{1}{1+\mathrm{j}\omega RC} = \frac{1-\mathrm{j}\omega RC}{1+(\omega RC)^2} \end{aligned}$$

Plotting the magnitude and the phase of this frequency transfer function, results in the following graphs. We clearly can see the low-pass filtering effect in the magnitude plot. The point where the gain drops by a factor $1/\sqrt{2}$ (i.e. -3 dB) has been indicated on the graph. We also can clearly see that the phase changes with frequency. But what does this phase change mean? To understand this, you need to read the next sections on eigenfunctions of LTI systems.

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

$$\arg|H(\omega)| = -\operatorname{atan}(\omega RC)$$



5.5.2 Eigenfunctions of LTI-systems

The eigenfunctions of a system are functions that don't change their nature when being treated by a system. They are in fact the eigenvectors of the corresponding (infinite dimensional) vector space. For LTI-systems, two sets of eigenfunctions exist:

- sinors
- sinusoids

As such, exciting the system with one of these eigenfunctions will result in a response that's of the same nature, i.e. a sinor will yield a sinor , a sine wave will yield a sine wave. The only thing that the system does with these eigenfunction signals is provide some (complex) gain. Of this gain, the modulus translates itself in an amplitude change of the sinor or sine wave, and the argument (or phase) translates itself in some phase change (which corresponds to delaying the eigenfunction).

Sinors

The most basic eigenfunctions are sinors. Let's start by calculating the Fourier transform of a sinor. We know that:

$$\delta(t) \xrightarrow{\mathcal{F}} 1$$

Therefore, because of time-frequency symmetry:

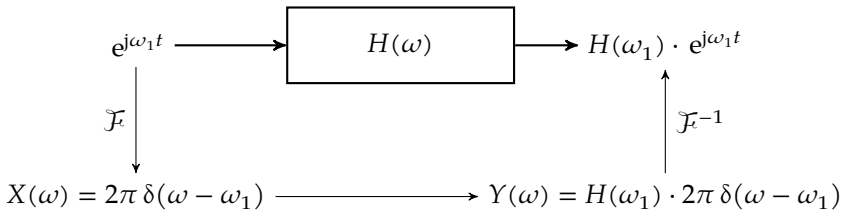
$$1 \xrightarrow{\mathcal{F}} 2\pi\,\delta(\omega)$$

Frequency shifting therefore results in:

$$e^{j\omega_1 t} \xrightarrow{\mathcal{F}} 2\pi\,\delta(\omega - \omega_1)$$

This is the Fourier pair we were looking for.

Using this information, we can easily complete the chain from input to output to find how the sinor is affected by the LTI system. In fact, we transform the problem to the frequency domain, solve it there and then transform the result back to the time domain.



Note that $H(\omega_1)$ was written, where one would expect $H(\omega)$. However, given the fact that it is the multiplier before a Dirac impulse, only the value when the Dirac impulse is not zero will matter. In general $H(\omega_1)$ will be a complex number. The relationship between input and output will therefore be:

$$e^{j\omega_1 t} \quad \xrightarrow{H} \quad |H(\omega_1)| \cdot e^{j\omega_1 t + \phi_1}$$

with $\phi_1 = \arg H(\omega_1)$. This means that the sinor will be amplified by a factor of $|H(\omega_1)|$ and delayed by an amount $\arg H(\omega_1)$.

Sine waves

Strangely, the sinor is not the only eigenfunction. So are sinusoids. Let's prove this for a sine wave (you can try the cosine wave yourself). A simple observation is that a sine wave can be written as the sum of two sinors:

$$\sin(\omega_1 t) = \frac{e^{j\omega_1 t} - e^{-j\omega_1 t}}{2j}$$

As the system is linear, the weighted sum of those two sine waves will appear at the output as the same weighted sum of their corresponding outputs. Therefore:

$$\sin(\omega_1 t) = \frac{e^{j\omega_1 t} - e^{-j\omega_1 t}}{2j} \quad \xrightarrow{H} \quad \frac{H(\omega_1) \cdot e^{j\omega_1 t} - H(-\omega_1) \cdot e^{-j\omega_1 t}}{2j}$$

If we assume the impulse response $h(t)$ is real, then $H(\omega)$ will be Hermitian, i.e.

$$H(-\omega_1) = \overline{H(\omega_1)}$$

or if we write $H(\omega_1) = A_1 e^{j\phi_1}$ then $H(-\omega_1) = A_1 e^{-j\phi_1}$.
Therefore:

$$\begin{aligned} \sin(\omega_1 t) = \frac{e^{j\omega_1 t} - e^{-j\omega_1 t}}{2j} \quad \xrightarrow{H} \quad & \frac{A_1 \cdot e^{j(\omega_1 t + \phi_1)} - A_1 \cdot e^{-j(\omega_1 t + \phi_1)}}{2j} \\ = \quad & A_1 \frac{e^{j(\omega_1 t + \phi_1)} - e^{-j(\omega_1 t + \phi_1)}}{2j} \\ = \quad & A_1 \sin(\omega_1 t + \phi_1) \end{aligned}$$

So, the output is again a sine wave, with some gain and some phase delay.

5.5.3 Determining a frequency transfer function

There are two ways by which we can discover the transfer function of an LTI system:

- by measuring it, and
- by calculating it.

For the former, we need measuring equipment and a physical LTI system that we can test. For the latter, we need a mathematical model of the system.

By measurement

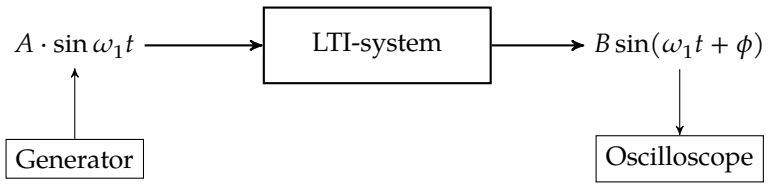
Measuring the frequency transfer function is not straightforward. The following does not work:

- apply a Dirac impulse as input and measure the impulse response (to transform it to the frequency domain): this does not work as we cannot generate analog Dirac impulses
- apply a sinor as input for every possible frequency and measure the amplitude gain and phase delay of the sinor at the output: this does not work as we cannot generate complex input signals

In fact, if we combine two sinors, we get a sine wave, and that approach does work:

- apply a sine wave as input for every possible frequency and measure the amplitude gain and phase delay of the sine wave at the output.[7]

This idea has been depicted in the following diagram:



Using this measurement setup we can perform measurements of A, B and ϕ for every desired frequency ω_1 and compose $H(\omega)$ using the following equations:

$$|H(\omega_1)| = \frac{B}{A} \qquad \arg\left(H(\omega_1)\right) = \phi$$

By calculation

The standard method that works for any LTI system is to execute the following procedure:

1. Compose the differential equations describing the system's behavior.
2. Solve the differential equations for $h(t)$.[8]
3. Calculate $h(t) \xrightarrow{\mathcal{F}} H(\omega)$.

However, for many LTI systems there is a quicker way. Instead of writing down the differential equations and spending costly time to solve these, we can also write the equations directly in the frequency domain, where solving them corresponds to solving algebraic equations instead of differential equations.

Especially for linear electronic cicruits, this works. Instead of writing down Kirchhoff's laws and the branch equations in the time domain, we will do this in the frequency domain. To this end, we transform the model of our electronic circuit to the frequency domain. This corresponds to replacing every network element by an equivalent network element in the frequency domain. Table 5.1 summarizes the transformation.

[7] A second possibility exists, using white noise. However, in this way you can only determine the magnitude spectrum and not the phase spectrum. We will not go into detail for this method.

[8] Remember that we have seen multiple ways to do this in Chapter 3.

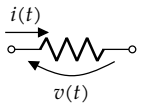
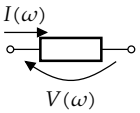
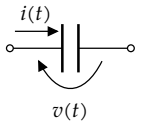
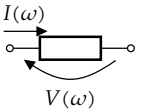
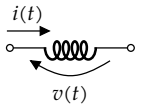
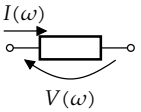
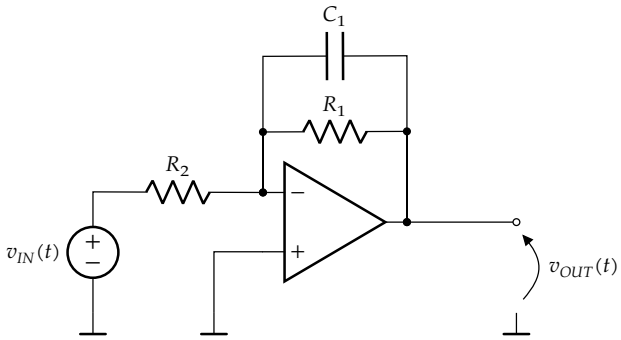
Time domain		Frequency domain
Resistor		
	$\mathcal{F}$	
$v(t) = R \cdot i(t)$	$\mathcal{F}$	$V(\omega) = \underbrace{R}_{\equiv Z} \cdot I(\omega)$
Capacitor		
	$\mathcal{F}$	
$i(t) = C \cdot \frac{\mathrm{d}v(t)}{\mathrm{d}t}$	$\mathcal{F}$	$I(\omega) = \underbrace{C \cdot \mathrm{j}\omega}_{\equiv 1/Z} \cdot V(\omega)$
Inductor		
	$\mathcal{F}$	
$v(t) = L \cdot \frac{\mathrm{d}i(t)}{\mathrm{d}t}$	$\mathcal{F}$	$V(\omega) = \underbrace{L \cdot \mathrm{j}\omega}_{\equiv Z} \cdot I(\omega)$

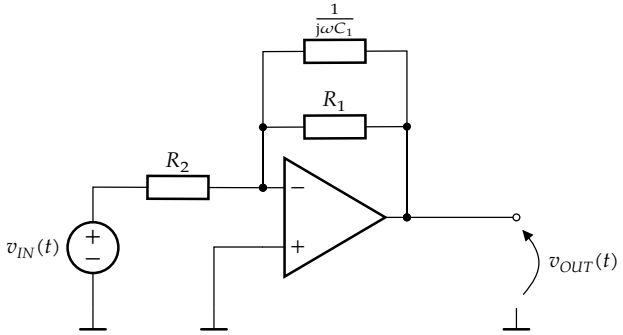
Table 5.1: Transformation of linear network elements to the frequency domain

You probably have seen this technique earlier in your courses on analog electronics.

As an example, consider the following circuit:



We can transform this network to the Fourier domain:

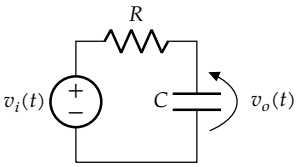


This allows calculating the following frequency-domain transfer function, without writing down or solving a single differential equation:

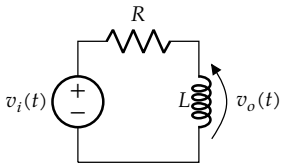
$$H(\omega) = \frac{V_{OUT}(\omega)}{V_{IN}(\omega)} = -\frac{R_1}{R_2}\frac{1}{1 + j\omega R_1 C_1}$$

Exercises

Exercise 5.5.3-1:
Consider the circuits below with input $v_i(t)$ and output $v_o(t)$. Determine their transfer functions using the given impulse response and the Fourier transform.

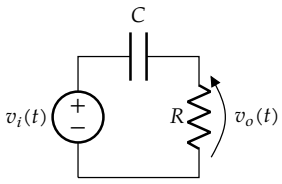


(a) $h(t) = \frac{1}{RC}\,\mathrm{e}^{-t/RC}\,\mathrm{u}(t)$

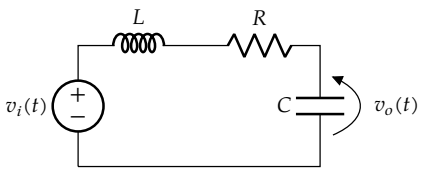


(b) $h(t) = \delta(t) - \frac{R}{L}\,\mathrm{e}^{-tR/L}\,\mathrm{u}(t)$

Exercise 5.5.3-2:
Consider the circuits below with input $v_i(t)$ and output $v_o(t)$. Determine their transfer functions using the complex impedance of the components.

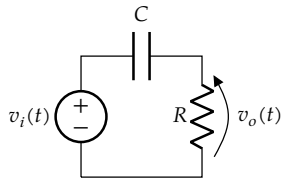


(a)

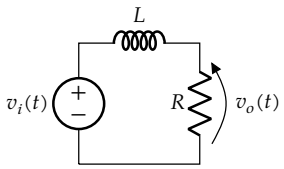


(b)

Exercise 5.5.3-3:
Consider the circuits below with input $v_i(t)$ and output $v_o(t)$. Determine their impulse responses.

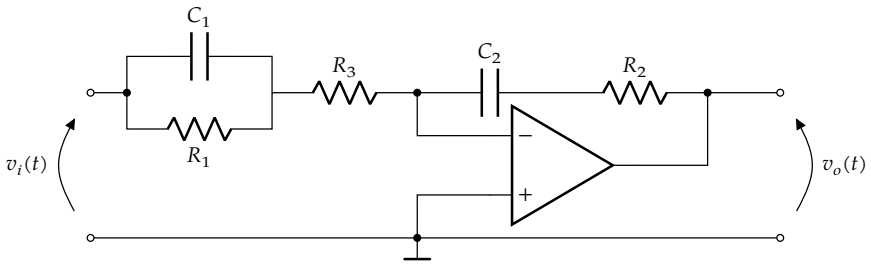


(a)

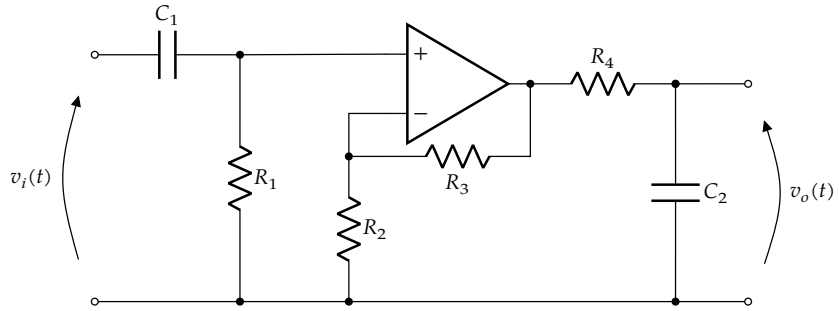


(b)

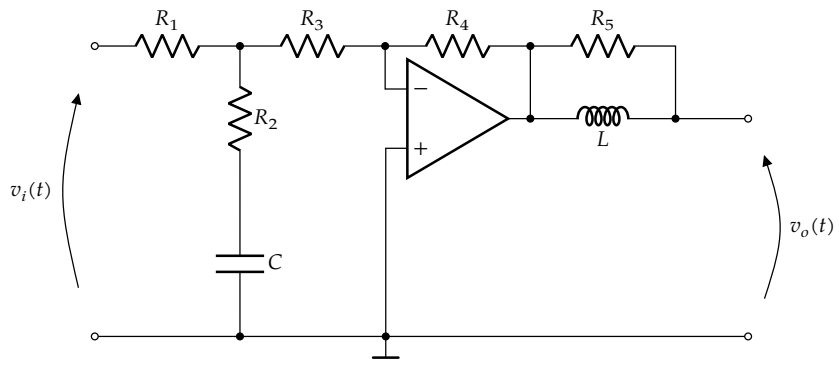
Exercise 5.5.3-4:
Determine the transfer function of the circuit below with input $v_i(t)$ and output $v_o(t)$.



Exercise 5.5.3-5:
Determine the transfer function of the circuit below with input $v_i(t)$ and output $v_o(t)$.

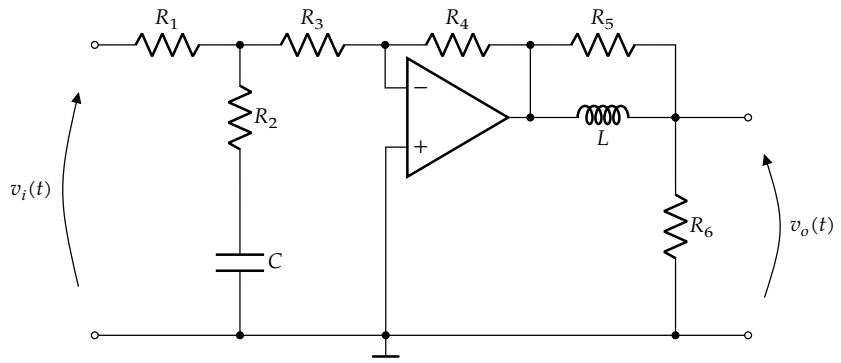


Exercise 5.5.3-6:
Determine the transfer function of the circuit below with input $v_i(t)$ and output $v_o(t)$. Assume $R_3 \gg R_2$.

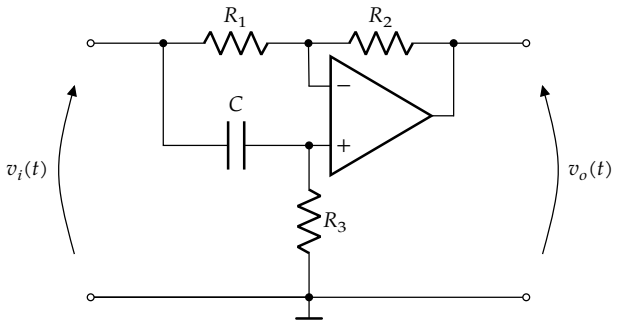


Exercise 5.5.3-7:

Determine the transfer function of the circuit below with input $v_i(t)$ and output $v_o(t)$. Assume $R_3 \gg R_2$.



Exercise 5.5.3-8:
Determine the transfer function of the circuit below with input $v_i(t)$ and output $v_o(t)$.



5.6 Conclusion

Now, take your time to review the Fourier family diagram of Figure 5.1 on page 52 again. By now it should be clear to you that the top two transforms of the Fourier family are very related.

Finally, a mathematical remark: the Fourier transform and its derivatives are *orthogonal signal decompositions*. In a finite-dimensional vector space a vector can be decomposed into basis vectors spanning the vector space. In the very same way a signal can be decomposed in the infinite-dimensional function space. The exponentials (or sines or cosines) are basis functions for this function space. They are also eigenfunctions of LTI systems.

Chapter 6

The Laplace transform

In this chapter, you will learn about:

- the Laplace transform, and how it is relelated to the Fourier Transform,
- why we need this transform,
- its properties,
- its stability, and finally
- the extended Fourier family diagram.

After having read this chapter, some questions will still be left unanswered:

- how do I describe systems using these transforms?
- how do I design such systems?

After having read/studied this chapter, you are expected to be able to

- calculate any (forward or inverse) Laplace transform given a signal or a spectrum,
- apply the transform properties,
- understand and compose pole-zero diagrams,
- assess the stability of systems based on a pole-zero diagram,
- reproduce (and find your way) in the extended Fourier family diagram.

6.1 From Fourier transform to Laplace transform

6.1.1 Forward...

The Fourier transform was defined in section **??** on page ?? as:

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} X_a(\omega)\, \mathrm{e}^{\mathrm{j}\omega t}\, \mathrm{d}\omega$$
$$X_a(\omega) = \int_{-\infty}^{+\infty} x(t)\, \mathrm{e}^{-\mathrm{j}\omega t}\, \mathrm{d}t$$

The subscript a has been added here to X_a to emphasize that it concerns is an *a*periodic spectrum. The Fourier transform of a signal is a decomposition of the signal in terms of a set of

orthonormal base functions. The analysis equation (5.12) can be seen as the projection equation which projects the signal vector onto the set of base vectors, or stated differently, as the equation that computes the correlation between the signal vector and the base vectors.

The entire theory, of course, only holds if the integrals in these equations converge. For a whole lot of cases (e.g., time-limited signals) these integrals do converge. However, in a whole lot of other cases these integrals do *not* converge: e.g., for sinusoidal signals (or any other periodic function) or exponential signals.

As an example, consider the constant function 1 and let's try to calculate its Fourier transform:

$$1 \xrightarrow{\mathcal{F}} F(\omega) = \int_{-\infty}^{+\infty} 1\, \mathrm{e}^{-j\omega t}\, \mathrm{d}t = \int_{-\infty}^{+\infty} \mathrm{e}^{-j\omega t}\, \mathrm{d}t$$

We can distinguish two cases:

1. $\omega = 0$

   $$F(\omega) = \int_{-\infty}^{+\infty} 1\, \mathrm{d}t = +\infty$$

   Conclusion: the integral is not bounded, and therefore does not converge
2. $\omega \neq 0$

   $$F(\omega) = \frac{1}{-j\omega} \left[\mathrm{e}^{-j\omega t} \right]_{-\infty}^{+\infty}$$

   Conclusion: when the boundaries that we need to fill out (in this primitive function) evolve to $\pm\infty$, the values of the primitive function keeps changing; the integral is bounded, yet it does not converge to a stable value.

This concludes our example.

An obvious idea is to give convergence a hand by multiplying with a convergence factor to $x(t)$. A very good choice, to this end, is adding an exponential convergence factor $\mathrm{e}^{-\sigma t}$ (with $\sigma \in \mathbb{R}$)[1], defining a new $x_c(t)$ as

$$x_c(t) = x(t)\, \mathrm{e}^{-\sigma t}$$

In this way, the Fourier transform relating $x_c(t) \xrightarrow{\mathcal{F}} X_{c,a}(\omega)$ can be written as:

$$\begin{aligned} x(t)\, \mathrm{e}^{-\sigma t} &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} X_{c,a}(\omega)\, \mathrm{e}^{\mathrm{j}\omega t}\, \mathrm{d}\omega \\ X_{c,a}(\omega) &= \int_{-\infty}^{+\infty} x(t)\, \mathrm{e}^{-\sigma t}\, \mathrm{e}^{-\mathrm{j}\omega t}\, \mathrm{d}t \end{aligned} \tag{6.3}$$

Now,

1. considering a fixed σ,
2. substituting $s = \sigma + \mathrm{j}\omega$, and
3. realizing that $X_{c,a}(\omega)$ also is a function of σ, so we'd better denote is as $X(s)$,

by rearranging (6.3) a little bit, we obtain the well-known

[1] To be really effective as convergence factor, σ in practice must be positive.

Laplace transform

$$x(t) = \frac{1}{2\pi \mathrm{j}} \int_{\sigma-\mathrm{j}\infty}^{\sigma+\mathrm{j}\infty} X(s)\, \mathrm{e}^{st}\, \mathrm{d}s$$
$$X(s) = \int_{-\infty}^{+\infty} x(t)\, \mathrm{e}^{-st}\, \mathrm{d}t$$

In this way, the Laplace transform checks for every value of σ the correlation between the exponentially modulated signal $x(t)\,\mathrm{e}^{-\sigma t}$ and complex exponentials $\mathrm{e}^{-\mathrm{j}\omega t}$ in the same way as the Fourier transform does. However, it is no longer an orthogonal signal decomposition.

The first integral seems a bit odd, it is a line integral in the complex plane, integrating along the straight line that connects lower and upper limit for a value of a σ for which the second integral converged.

Remarks

- The Laplace transform is often symbolized as an operator $\mathcal{L}$:

$$X(s) = \mathcal{L}(x(t)) \qquad\qquad x(t) = \mathcal{L}^{-1}(X(s))$$

- The use of lowercase symbols for time-domain signals and uppercase symbols for their Laplace transform is quite common. The pair is also often written as follows:

$$x(t) \overset{\mathcal{L}}{\rightarrow} X(s)$$

6.1.2 ...and back

In case the integrals above converge for $\sigma = 0$, these equations reduce to the original Fourier transform equations. It is therefore (under these circumstances) very simple to derive the Fourier transform (and hence the frequency spectrum it represents) from the Laplace transform: just set $\sigma = 0$ in the Laplace transform.

Example

We can illustrate this by considering as example the transform pair:

$$x(t) = \begin{cases} 1 & \text{if } |t| \le 1 \\ 0 & \text{if } |t| > 1 \end{cases} \overset{\mathcal{L}}{\rightarrow} X(s) = \frac{\mathrm{e}^{s} - \mathrm{e}^{-s}}{s} \tag{6.4}$$

For $\sigma = 0$ the Laplace transform integrals still converge. Therefore the Fourier transform can be derived as:

$$\begin{aligned} X(\omega) &= X(s)|_{s=\mathrm{j}\omega} \\ &= \frac{\mathrm{e}^{\mathrm{j}\omega} - \mathrm{e}^{-\mathrm{j}\omega}}{\mathrm{j}\omega} \\ &= 2\frac{\sin(\omega)}{\omega} = 2\,\mathrm{sinc}(\omega) \end{aligned}$$

In view of the substitution of s by $\mathrm{j}\omega$, very often, the Fourier transform is denoted as $X(\mathrm{j}\omega)$ instead of $X_a(\omega)$.

6.2 One-sided Laplace transform

Very often, we only consider signals that start at $t = 0$. In that case we obtain the

One-sided Laplace transform

$$x(t) = \frac{1}{2\pi \mathrm{j}} \int_{\sigma-\mathrm{j}\infty}^{\sigma+\mathrm{j}\infty} X(s)\, \mathrm{e}^{st}\, \mathrm{d}s \tag{6.5}$$
$$X(s) = \int_{0^-}^{+\infty} x(t)\, \mathrm{e}^{-st}\, \mathrm{d}t$$

The lower bound is a left-side limit to zero, to allow a Dirac impulse at $t = 0$ to be fully part of the integral. In most cases the one-sided Laplace transform is simply labeled as *the* Laplace transform, while the two-sided version is explicitly described as the *two-sided* or *bilateral* Laplace transform.

To clearly state that we're considering causal signals, we should multiply the functions we use with $u(t)$. E.g. e^{at} should become $u(t)\,\mathrm{e}^{at}$, $\sin \omega t$ should become $u(t) \sin \omega t$, a.s.o. In many cases, we will explicitly write $u(t)$, but in some cases (because it is evident from the context) and in most of the literature, the factor $u(t)$ is omitted. Keep in mind that when using the one-sided Laplace transform that all signals under consideration are causal. In the remainder of this chapter, if we don't mention it explicitly, we will use the term 'Laplace transform' for the one-sided Laplace transform.

6.3 Region of convergence

Our definition of the Laplace transform has the form of an improper integral. The question rises whether this integral converges or not. For a generic causal signal $x(t)$ that is exponentially bounded and piecewise continuous, we can use the following lemma:

Lemma: convergence of the Laplace transform

For any signal $x(t)$ that is

1. causal,
2. piecewise continuous on the range $t \in [0, t_0]$, and
3. exponentially bounded otherwise, i.e.

$$\exists t_0, M, \alpha \in \mathbb{R}, \forall t > t_0 : |x(t)| < M\, \mathrm{e}^{\alpha t},$$

the corresponding Laplace transform $X(s) = \mathcal{L}(x(t))$ converges for $\mathrm{Re}(s) > \alpha$.

Proof

We take a start with the definition of the Laplace transform:

$$X(s) = \int_{0^-}^{+\infty} x(t)\, \mathrm{e}^{-st}\, \mathrm{d}t$$

We can split this integral in two parts:

$$X(s) = \underbrace{\int_{0^-}^{t_0} x(t)\,\mathrm{e}^{-st}\,\mathrm{d}t}_{X_1(s)} + \underbrace{\int_{t_0}^{+\infty} x(t)\,\mathrm{e}^{-st}\,\mathrm{d}t}_{X_2(s)}$$

The first part $(X_1(s))$ is piecewise continuous and therefore integrable.

The second part

$$X_2(s) = \int_{t_0}^{+\infty} x(t)\,\mathrm{e}^{-st}\,\mathrm{d}t$$

is elaborated below.

Taking the absolute value of both sides of the equation yields:

$$\begin{aligned}
|X_2(s)| &= \left|\int_{t_0}^{+\infty} x(t)\,\mathrm{e}^{-st}\,\mathrm{d}t\right| \\
&\leq \int_{t_0}^{+\infty} \left|x(t)\,\mathrm{e}^{-st}\right|\,\mathrm{d}t \\
&= \int_{t_0}^{+\infty} |x(t)| \cdot \left|\mathrm{e}^{-st}\right|\,\mathrm{d}t \\
&< \int_{t_0}^{+\infty} M\,\mathrm{e}^{\alpha t} \cdot \left|\mathrm{e}^{-st}\right|\,\mathrm{d}t \\
&\downarrow \quad \left|\mathrm{e}^{-st}\right| = \left|\mathrm{e}^{-\sigma t}\right| \underbrace{\left|\mathrm{e}^{+\mathrm{j}\omega t}\right|}_{=1} = \mathrm{e}^{-\sigma t} \\
&= \int_{t_0}^{+\infty} M\,\mathrm{e}^{\alpha t} \cdot \mathrm{e}^{-\sigma t}\,\mathrm{d}t \\
&= \frac{M}{\alpha - \sigma}\left[\,\mathrm{e}^{(\alpha-\sigma)t}\right]_{t_0}^{t\to+\infty} \\
&\downarrow \quad \sigma > \alpha \\
&= -\frac{M}{\alpha - \sigma}\,\mathrm{e}^{(\alpha-\sigma)t_0}
\end{aligned}$$

Therefore, under these circumstances $X_2(s)$ converges absolutely and therefore also converges as such. As also X_1 can also be integrated, the conclusion is that $X(s) = X_1(s) + X_2(s)$ converges. ■

Note that because of the inequality in the proof above, also the following must hold for exponentially bounded signals $x(t) \xrightarrow{\mathcal{L}} X(s)$:

$$\lim_{s\to\infty} X(s) = 0$$

This also means that when the Laplace transform $X(s)$ of an exponentially bound signal $x(t)$ is a rational function, i.e.

$$X(s) = \frac{N(s)}{D(s)}$$

that the following must hold:

$$\text{degree}\,(N(s)) \leq \text{degree}\,(D(s))$$

6.4 Properties

From the definition of the (one-sided) Laplace transform a number of properties can be derived.

The interpretation of these properties is similar to the interpretation of the Fourier transform properties. Therefore, we will be brief and just list the properties. Try to formulate the interpretation yourself and when needed, take a peek at the Fourier transform properties.

Let's start by making the following assumptions:

$$\begin{gathered} x(t) \xrightarrow{\mathcal{L}} X(s) \\ y(t) \xrightarrow{\mathcal{L}} Y(s) \\ a, b \in \mathbb{R} \\ k \in \mathbb{R}_0 \end{gathered}$$

and $x(t)$ and $y(t)$ are causal.

Linearity

$$ax(t) + by(t) \xrightarrow{\mathcal{L}} aX(s) + bY(s)$$

Time/Frequency scaling

$$\begin{aligned} x(kt) &\xrightarrow{\mathcal{L}} \frac{1}{|k|} X\left(\frac{s}{k}\right) \\ \frac{1}{|k|} x\left(\frac{t}{k}\right) &\xrightarrow{\mathcal{L}} X(ks) \end{aligned}$$

Time/Frequency shifting

$$\begin{aligned} x(t - t_0) &\xrightarrow{\mathcal{L}} X(s)\, \mathrm{e}^{-st_0} \\ x(t)\, \mathrm{e}^{+s_0 t} &\xrightarrow{\mathcal{L}} X(s - s_0) \end{aligned}$$

Time/Frequency differentation

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} x(t) &\xrightarrow{\mathcal{L}} sX(s) - x(0) \\ -tx(t) &\xrightarrow{\mathcal{L}} \frac{\mathrm{d}}{\mathrm{d}s} X(s) \end{aligned}$$

Time/Freqency division

$$\begin{aligned} \frac{x(t)}{t} &\xrightarrow{\mathcal{L}} \int_s^\infty X(u)\, \mathrm{d}u \\ \int_0^t x(u)\, \mathrm{d}u &\xrightarrow{\mathcal{L}} \frac{X(s)}{s} \end{aligned}$$

Laplace transform of a periodic function

$$\left.\begin{aligned} &f(t) \text{ is periodic with period } T \\ &\phi(t) = (u(t) - u(t - T)) f(t) \xrightarrow{\mathcal{L}} \Phi(s) \end{aligned}\right\} \Rightarrow f(t) \xrightarrow{\mathcal{L}} F(s) = \frac{\Phi(s)}{1 - \mathrm{e}^{-sT}}$$

Convolution/Multiplication

$$x(t) \star y(t) \xrightarrow{\mathcal{L}} X(s)Y(s)$$
$$x(t)y(t) \xrightarrow{\mathcal{L}} \frac{1}{2\pi \mathrm{j}} X(s) \star Y(s)$$

Initial/Final value theorem

$$x(0) = \lim_{s \to \infty} sX(s)$$
$$\lim_{t \to +\infty} x(t) = \lim_{s \to 0} sX(s)$$

This is under the condition that these limits exist.

Exercises

Exercise 6.4-1:
Prove the property of time-scaling.

Exercise 6.4-2:
Prove the property of time shifting.

Exercise 6.4-3:
Prove the property of time differentiation.

Exercise 6.4-4:
Prove the property of the Laplace transform of a periodic function.

Exercise 6.4-5:
Find the initial and final value of $f(t)$ when $F(s) = \dfrac{2s^2 - 3s + 4}{s^3 + 3s^2 + 2s}$.

Exercise 6.4-6:
Find the initial and final value of $f(t)$ when $V(s) = \dfrac{s + 16}{s^2 + 4s + 12}$.

Exercise 6.4-7:
Find the initial and final value of $f(t)$ when $V(s) = \dfrac{s + 10}{3s^3 + 2s^2 + s}$.

Exercise 6.4-8:
Find the initial and final value of $f(t)$ when $F(s) = \dfrac{-2(s + 7)}{s^2 - 2s + 10}$.

6.5 Practical ways to calculate the (inverse) Laplace transform

6.5.1 Forward

The most straightforward way to calculate the Laplace transform is to calculate the integral of its definition. However in most cases, a smart use of the basic properties can help.

Example 1
Compose the Laplace transform of $x(t) = \mathrm{e}^{ct}$ with $c \in \mathbb{C}$.

We start using the definition:

$$\begin{aligned} X(s) &= \int_{0^-}^{+\infty} \mathrm{e}^{ct}\,\mathrm{e}^{-st}\,\mathrm{d}t \\ &= \frac{1}{a-s}\left[\,\mathrm{e}^{(c-s)t}\right]_{0^-}^{t\to+\infty} \\ &\downarrow \quad \mathrm{Re}(s) > \mathrm{Re}(c) \\ &= \frac{1}{s-c} \end{aligned}$$

Example 2
Compose the Laplace transform of $x(t) = \delta(t)$.

Again, starting from the definition:

$$\begin{aligned} X(s) &= \int_{0^-}^{+\infty} \delta(t)\,\mathrm{e}^{-st}\,\mathrm{d}t \\ &= \mathrm{e}^{-s\cdot 0} = 1 \end{aligned}$$

Example 3
Compose the Laplace transform of $x(t) = \sin \omega t$.

$$\begin{aligned} X(s) &= \int_{0^-}^{+\infty} \sin\omega t \cdot \mathrm{e}^{-st}\,\mathrm{d}t \\ &\downarrow \quad \sin\omega t = \frac{\mathrm{e}^{\mathrm{j}\omega t} - \mathrm{e}^{-\mathrm{j}\omega t}}{2\mathrm{j}} \\ &= \frac{1}{2\mathrm{j}}\left(\int_{0^-}^{+\infty} \mathrm{e}^{\mathrm{j}\omega t}\,\mathrm{e}^{-st}\,\mathrm{d}t - \int_{0^-}^{+\infty} \mathrm{e}^{-\mathrm{j}\omega t}\,\mathrm{e}^{-st}\,\mathrm{d}t\right) \\ &\downarrow \quad \text{see example 1} \\ &= \frac{1}{2\mathrm{j}}\left(\frac{1}{s-\mathrm{j}\omega} - \frac{1}{s+\mathrm{j}\omega}\right) \\ &= \frac{1}{2\mathrm{j}}\frac{s+\mathrm{j}\omega - s + \mathrm{j}\omega}{s^2+\omega^2} \\ &= \frac{\omega}{s^2+\omega^2} \end{aligned}$$

You will appreciate that finding Laplace transforms requires some calculus and algebraic skills.

Of course, it makes sense to keep common transform pairs handy in a table, such that we can avoid doing the same calculations over and over again (see section 6.1). It turns out that the table gets us quite a way.

Exercises

Exercise 6.5.1-1:
Find the Laplace transform for the following time-domain signals:

a) $x(t) = t^2$
b) $x(t) = \cos 5t$
c) $x(t) = \sinh \frac{t}{3}$
d) $x(t) = u(t-1)$

Exercise 6.5.1-2:
Find the Laplace transform for the following time-domain signals:

a) $x(t) = t + 1$
b) $x(t) = 5\sin 2t + 3\cos 5t$
c) $x(t) = u(t) - \delta(t)$
d) $x(t) = 1 - 2\,\mathrm{e}^{-3t}$
e) $x(t) = \cos\left(10t + \frac{\pi}{3}\right)$

Exercise 6.5.1-3:
Find the Laplace transform for the following time-domain signals:

a) $x(t) = t^2 \sin t$
b) $x(t) = \mathrm{e}^{-4t}\cos 2t$
c) $x(t) = \frac{1 - \mathrm{e}^{-t}}{t}$
d) $x(t) = (t+1)\,\mathrm{e}^{-3t}$
e) $x(t) = t\,\mathrm{e}^{-3t}\cos 4t$

6.5.2 Inverse

Whereas the forward transform is not so difficult (especially when you get to master the basic properties), the inverse transform starts out to be almost impossible. Indeed, solving (6.5) is a daunting task.

To get around this, in general, we have several options:

1. Use the properties and a table of common transform pairs to recognize known Laplace transforms in the parts of an expression.
2. Use the convolution property. This allows to inversely transform a product of two Laplace transforms to a convolution of their individual transforms.

However, if $X(s)$ is a ratio of two polynomials in s (which in engineering covers over 90% of all cases), there is a simple procedure that we can use.

We have seen earlier (on page 103) that when $X(s)$ is a rational function with

$$X(s) = \frac{N(s)}{D(s)}$$

that

$$\text{degree}\,(N(s)) \leq \text{degree}\,(D(s))$$

in case $x(t)$ is exponentially bounded.

Procedure to calculate the inverse Laplace transform for rational $X(s)$:

Procedure:

1. Perform partial fraction expansion (PFE) on $X(s)$.

$$X(s) = X_1(s) + X_2(s) + X_3(s) + \ldots$$

2. Each of the terms $X_j(s)$ will have one of the following forms:

   2.1. $\dfrac{A}{(s+a)^i}$ with $A, a \in \mathbb{R}$ and $i \in \mathbb{N}_0$.

   2.2. $\dfrac{Bs+C}{(s^2+bs+c)^i}$ with $B, C, b, c \in \mathbb{R}$ and $i \in \mathbb{N}_0$, and $b^2 - 4c < 0$.

3. Because of linearity, the inverse Laplace transform is the sum of the inverse transforms of the individual terms.

Terms in the form of 2.1:
These are the transforms of t^{i-1} that have been translated in frequency:

$$A \cdot \mathrm{e}^{-at} \cdot \frac{t^{i-1}}{(i-1)!} \cdot u(t) \xrightarrow{\mathcal{L}} \frac{A}{(s+a)^i}$$

Terms in the form of 2.2:
In case $i = 1$, we can rewrite the term as:

$$\begin{aligned}\frac{Bs+C}{s^2+bs+c} &= \frac{Bs+C}{(s+b/2)^2 + \underbrace{(c-b^2/4)}_{d^2}} \\ &= B\frac{s+b/2}{(s+b/2)^2+d^2} + \underbrace{\left(\frac{C}{d} - \frac{Bb}{2d}\right)}_{D}\frac{d}{(s+b/2)^2+d^2}\end{aligned}$$

This allows to recognize the transform of a sine and a cosine wave:

$$B\,\mathrm{e}^{-\frac{b}{2}t}\cos dt + D\,\mathrm{e}^{-\frac{b}{2}t}\sin dt \xrightarrow{\mathcal{L}} B\frac{s+b/2}{(s+b/2)^2+d^2} + D\frac{d}{(s+b/2)^2+d^2}$$

In case $i = 2$, we can consider the term as a product of two identical terms. Therefore, we can use the same pattern, but need to calculate an additional convolution. For $i = 3, 4, \ldots$, the same principle holds, but the calculus becomes quite cumbersome.

Exercises

Exercise 6.5.2-1:
Calculate the inverse Laplace transform of:

a) $\dfrac{1}{s+7}$

b) $\dfrac{1}{s^6}$

c) $\dfrac{s}{s^2+16}$

d) $\dfrac{7}{s^2+144}$

e) $\dfrac{1}{(s+4)^7}$

f) $\dfrac{1}{s^2+6s+11}$

g) $\dfrac{3s\,e^{-6s}}{s^2+4}$

h) $\dfrac{e^{-10s}}{(s-2)^4}$

i) $\dfrac{s}{(s^2-25)^2}$

j) $\dfrac{2}{s(s^2+4)}$

k) $\dfrac{20}{s(s^2+2s+2)}$

l) $\dfrac{s+1}{4s^2+4s+5}$

Exercise 6.5.2-2:
Calculate the inverse Laplace transform of:

a) $\dfrac{1}{s^2(s^2+1)}$

b) $\dfrac{s}{(s^2+9)^2}$

c) $\dfrac{s^2}{(s^2+a^2)^2}$

d) $\dfrac{1}{s^3(s^2+a^2)}$

e) $\dfrac{1}{(s^2+a^2)^2}$

using the convolution lemma.

Exercise 6.5.2-3:
Calculate the inverse Laplace transform of:

a) $\dfrac{1}{s^2+s}$

b) $\dfrac{s^2+1}{s(s+1)(s+2)}$

c) $\dfrac{4s+22}{(s+3)^3(s^2+1)}$

d) $\dfrac{2^2+3s-8}{(s-1)(s+1)(s+4)}$

e) $\dfrac{s^3}{s^4-5s^2+4}$

f) $\dfrac{2a^4}{s^4-a^4}$

using partial fraction expansion (PFE).

6.6 Common transform pairs

Without going into the mathematical details of calculating Laplace transforms (you took mathematics courses to that end), fact is that knowing some simple transform pairs in combination with the properties in the previous section, helps in using the Laplace transform successfully. Table 6.1 on the following page displays the most common transform pairs.[2]

6.7 Description of LTI-systems in the Laplace domain

The Laplace transform can be applied to any signal for which the Laplace integral converges. It certainly is not restricted to signals of LTI systems. However, the Laplace transform is especially useful for LTI systems. Let's see why.

[2]The use of $u(t)$ makes these formulas also valid for the (generic) two-sided transform.

$x(t)$	$\xrightarrow{\mathcal{L}}$	$X(s)$
$\delta(t)$	$\xrightarrow{\mathcal{L}}$	1
$u(t)$	$\xrightarrow{\mathcal{L}}$	$\frac{1}{s}$
$t \cdot u(t)$	$\xrightarrow{\mathcal{L}}$	$\frac{1}{s^2}$
$\frac{t^n}{n!} \cdot u(t)$	$\xrightarrow{\mathcal{L}}$	$\frac{1}{s^{n+1}}$
$t^p \cdot u(t)$	$\xrightarrow{\mathcal{L}}$	$\frac{\Gamma(p+1)}{s^{p+1}}$
$\mathrm{e}^{-at} \cdot u(t)$	$\xrightarrow{\mathcal{L}}$	$\frac{1}{s+a}$
$\sin(at) \cdot u(t)$	$\xrightarrow{\mathcal{L}}$	$\frac{a}{s^2+a^2}$
$\cos(at) \cdot u(t)$	$\xrightarrow{\mathcal{L}}$	$\frac{s}{s^2+a^2}$

Table 6.1: Common Laplace transform pairs, given $n \in \mathbb{N}, a, p \in \mathbb{R}, p > -1$

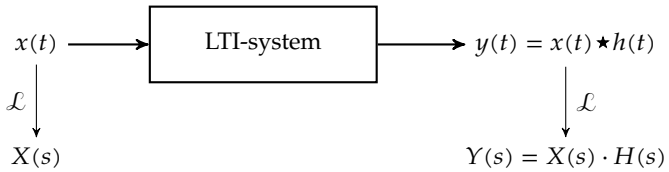
6.7.1 The transfer function

Deriving the concept

From studying LTI systems and the convolution, we've learned that we can calculate the output of an LTI system by calculating the convolution of the input with its impulse response:

$$y(t) = x(t) \star h(t) = \int_{-\infty}^{+\infty} x(\tau) h(t-\tau)\,\mathrm{d}t$$

Given the convolution property of the Laplace transform (see page 105), this means:



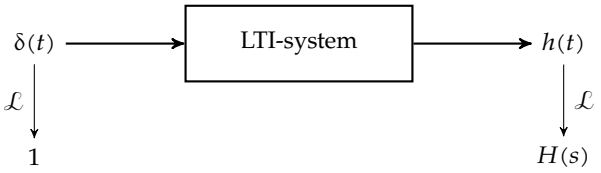
This allows for a very useful new concept, the *transfer function*:[3]

$$H(s) = \frac{Y(s)}{X(s)}$$

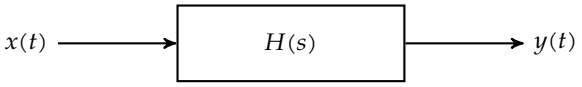
This transfer function is not only the constant gain ratio (from input to output), it is the Laplace

[3]To be very specific, one could label it as the *complex frequency transfer funnction*. However, we rarely do that and assume that the term *transfer function* belongs in the complex frequency domain (as opposed to the *frequency transfer function* in the Fourier Domain).

transform of the impulse response, i.e. if you excite the system with a Dirac impulse, the response in the Laplace domain will be the transfer function.



Because the transfer function $H(s)$ is a full description of the LTI-system (as was the impulse response), we often just write an H inside the box of an LTI-system to indicate that it is LTI with a transfer function $H(s)$:



Definition

In this way we come to the definition of the *transfer function*:

Transfer function

For an LTI-system the *transfer function* $H(s)$ is the Laplace transform of the impulse response $h(t)$:

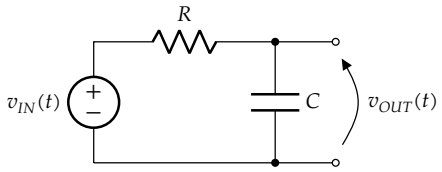
$$h(t) \overset{\mathcal{L}}{\longrightarrow} H(s)$$

It describes the (linear) relationship between the input x and the output y of an LTI system in the complex frequency domain:

$$H(s) = \frac{Y(s)}{X(s)} \qquad \Leftrightarrow \qquad Y(s) = H(s) \cdot X(s)$$

Example

Consider the filter below:



We calculated the impulse response of this filter earlier (using v_{IN} as input and v_{OUT} as output) and found it to be:

$$h(t) = \mathrm{u}(t) \cdot \frac{1}{RC} \mathrm{e}^{-\frac{t}{RC}}$$

The frequency transfer function can be calculated as follows:

$$\begin{aligned} H(s) = \mathcal{L}(h(t)) &= \int_0^{+\infty} \mathrm{u}(t)\, \frac{1}{RC}\, \mathrm{e}^{-\frac{t}{RC}}\, \mathrm{e}^{-st}\, \mathrm{d}t \\ &= \frac{1}{RC} \int_0^{+\infty} \mathrm{e}^{-\frac{1+sRC}{RC}t}\, \mathrm{d}t \\ &= \frac{1}{RC} \left(-\frac{RC}{1+sRC} \left[\mathrm{e}^{-\frac{1+sRC}{RC}t} \right]_0^{+\infty} \right) \\ &\downarrow s = \sigma + \mathrm{j}\omega \\ &= \frac{1}{RC} \left(-\frac{RC}{1+sRC} \left[\mathrm{e}^{-\frac{1+\sigma RC}{RC}t}\, \mathrm{e}^{-\mathrm{j}\omega t} \right]_0^{+\infty} \right) \end{aligned}$$

Considering this final line, we can see that the primitive function within the square brackets is a sinor with frequency ω and an exponential amplitude. If $\sigma > -1/RC$, this amplitude will go down for increasing t. Therefore:

$$\begin{aligned} H(s) &= \frac{1}{RC} \left(-\frac{RC}{1+sRC} (0-1) \right) \\ &= \frac{1}{1+sRC} \end{aligned}$$

6.7.2 Rational Laplace transforms - zeros and poles

In many cases the form of the Laplace transform will be a rational function in s, i.e. a ratio of two polynomials in s:

$$X(s) = \frac{N(s)}{D(s)}$$

This holds almost always for the transfer functions of LTI systems and also for many signals. To convince yourself, consider the table of common transform pairs (Table 6.1).

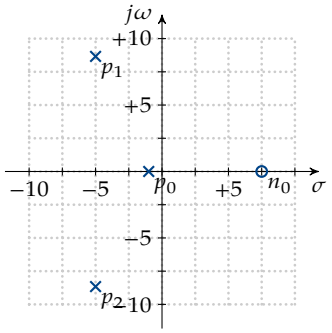
We define *zeros* and *poles* of $X(s)$ as follows:

- the roots of $N(s)$ are the *zeros* of $X(s)$, and
- the roots of $D(s)$ are the *poles* of $x(s)$.

Often, these poles and zeros are depicted in the complex plane. As an example, consider the following transfer function:

$$H(s) = \frac{s - 7.5}{(s+1)(s^2 + 10s + 100)}$$

This yields the following pole-zero plot:



with

- ●× indicating the poles
- ●○ indicating the zeros

Note that for Laplace transforms of real signals (or real impulse responses) poles and zeros always come in complex conjugate pairs.

6.7.3 Stability

A function $x(t)$ is called stable if its "end-value" is finite, i.e. iff

$$\lim_{t\to+\infty} x(t) \in \mathbb{R}$$

For signals whose Laplace transform can be described by a ratio of polynomials in s this translates into the requirement that *all poles need to be located in the left half-plane*.

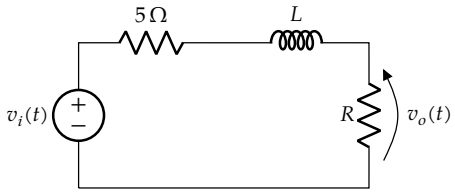
This requirement is easy to understand if you reconsider the forms 2.1 and 2.2 of the procedure to calculate the inverse of the Laplace transform (see page 108). The elements that can cause instability in these terms are the exponentials. As long as $a > 0$ for the form 2.1 and $b > 0$ for the form 2.2, the exponentials are decaying with time. For the latter the same holds, even for $i > 1$, though we won't elaborate this in detail.

However, the conclusion is simple: *poles in the left-half plane are stable poles; If the poles lie on the border* ($\sigma = 0$), *the system will exhibit oscillatory behavior.*

The examples in Figure 6.1 and Figure 6.2 illustrate this.

Exercises

Exercise 6.7.3-1:
The step response of the circuit below with input $v_i(t)$ is $v_o(t) = \frac{3}{4}(1 - \mathrm{e}^{-100t})\,\mathrm{u}(t)$.
Determine (a) the impulse response of this circuit, (b) the value of L and R and (c) whether it is stable.



Exercise 6.7.3-2:
The step response of a circuit is $v_o(t) = [5 - 5\,\mathrm{e}^{-2t}(1 + 2t)]\,\mathrm{u}(t)$.
Determine the transfer function of the circuit and determine if it is stable.

Exercise 6.7.3-3:
The step response of a circuit is $v_o(t) = (40 + 1.03\,\mathrm{e}^{8t} - 41\,\mathrm{e}^{-320t})\,\mathrm{u}(t)$.
Determine the transfer function of the circuit and determine if it is stable.

Exercise 6.7.3-4:
The step response of a circuit is $v_o(t) = \frac{5}{3}(\mathrm{e}^{-5t} - \mathrm{e}^{-20t})\,\mathrm{u}(t)$.
Determine the transfer function of the circuit and determine if it is stable.

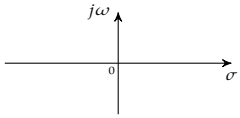
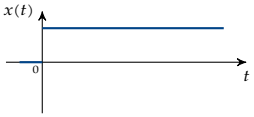
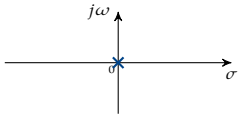
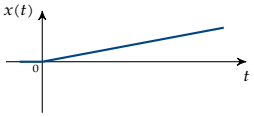
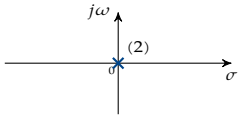
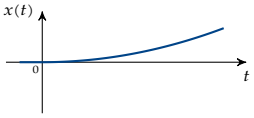
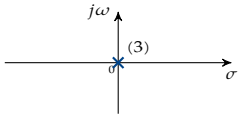
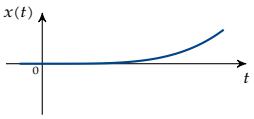
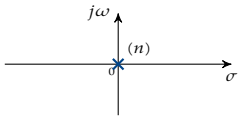
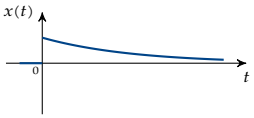
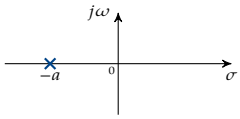
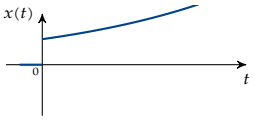
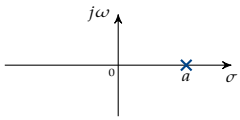
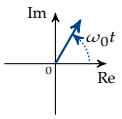
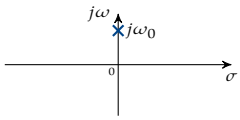
$x(t)$	Time domain	$X(s)$	Pole/zero	Note
$\delta(t)$		1		non-physical
$u(t)$		$\frac{1}{s}$		stable
$t\,u(t)$		$\frac{1}{s^2}$		unstable
$\frac{t^2}{2}\,u(t)$		$\frac{1}{s^3}$		unstable
$\frac{t^n}{n!}\,u(t)$		$\frac{1}{s^{n+1}}$		unstable
$e^{-at}\,u(t)$ $(a>0)$		$\frac{1}{s+a}$		stable
$e^{at}\,u(t)$ $(a>0)$		$\frac{1}{s-a}$		unstable
$e^{j\omega_0 t}\,u(t)$		$\frac{1}{s-j\omega_0}$		oscillating

Figure 6.1: Illustration of some time-domain signals with their corresponding Laplace transforms and an indication of the location of the poles and zeros

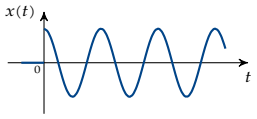
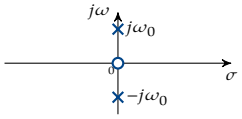
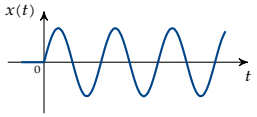
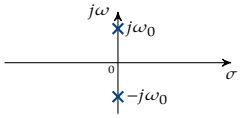
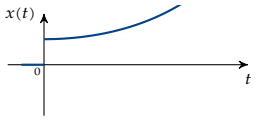
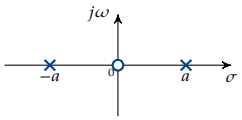
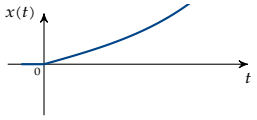
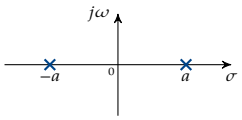
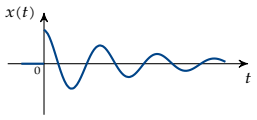
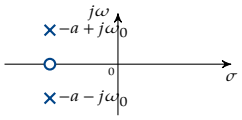
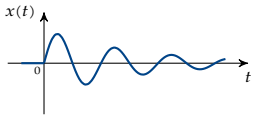
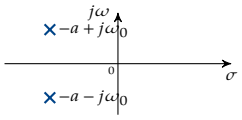
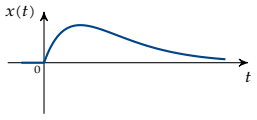
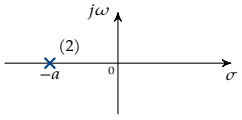
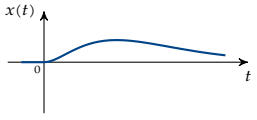
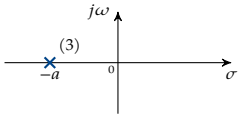
$x(t)$	Time domain	$X(s)$	Pole/zero	Note
$\cos(\omega_0 t)\, u(t)$		$\frac{s}{s^2+\omega_0^2}$		oscillating
$\sin(\omega_0 t)\, u(t)$		$\frac{\omega_0}{s^2+\omega_0^2}$		oscillating
$\cosh(at)\, u(t)$		$\frac{s}{s^2-a^2}$		unstable
$\sinh(at)\, u(t)$		$\frac{a}{s^2-a^2}$		unstable
$\cos(\omega_0 t)\ \mathrm{e}^{-at}\, u(t)$ $(a>0)$		$\frac{s+a}{(s+a)^2+\omega_0^2}$		stable
$\sin(\omega_0 t)\ \mathrm{e}^{-at}\, u(t)$ $(a>0)$		$\frac{\omega_0}{(s+a)^2+\omega_0^2}$		stable
$t\ \mathrm{e}^{-at}\, u(t)$ $(a>0)$		$\frac{1}{(s+a)^2}$		stable
$t^2\ \mathrm{e}^{-at}\, u(t)$ $(a>0)$		$\frac{2}{(s+a)^3}$		stable

Figure 6.2: Illustration of some time-domain signals with their corresponding Laplace transforms and an indication of the location of the poles and zeros

6.8 Why do we need the Laplace transform?

It must be said that even though we originally stated having added the convergence factor $e^{-\sigma t}$ to help convergence, the factor may well be added even if the original Fourier transform equations *did* converge.

Convergence or not, the net result is that now we have a transform that maps the original signal to a two-dimensional space. Why would we want this complication if we find ourselves served equally well by the Fourier transform?

Example
Consider the graphical plots of the Laplace transform of the example of (6.4) in Figure 6.3 on the next page.

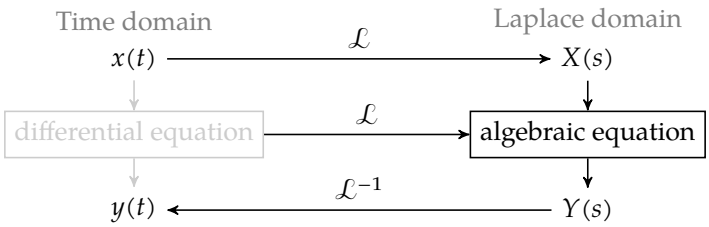
There are two things to note on these plots:

1. we can see the $|\text{sinc}(\omega)|$ for $\sigma = 0$, corresponding to the Fourier transform of a rectangular window function,
2. apart from this, the plots are not really more insightful than the ones of the "original" Fourier transform.

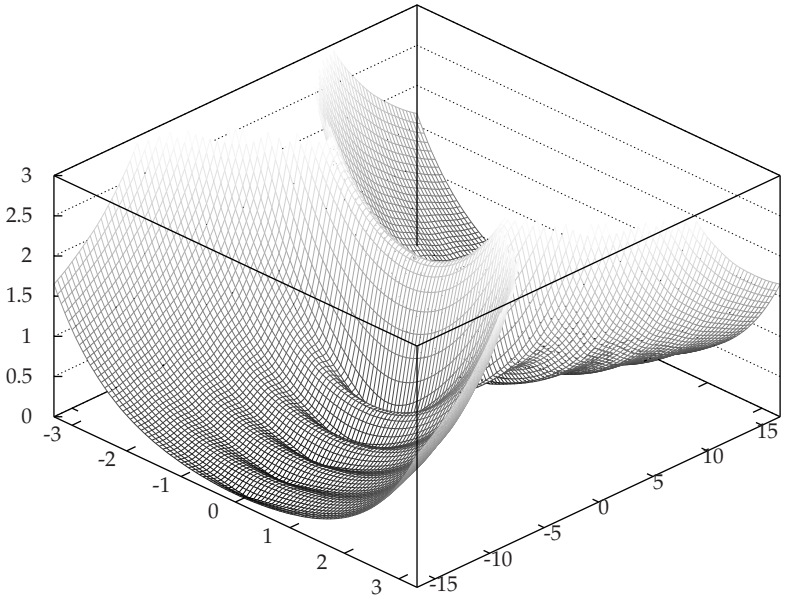
Therefore, once again: why would we put up with this extra complication if there's no benefit for it?

Well, many physical systems can be described using systems of (linear) differential equations. The solutions to these equations can be written in terms of sinusoidal and exponential functions. A particular fact is that the Laplace transforms of these types of functions become very simple expressions. This leads to an easy solution of the original system in the Laplace domain.

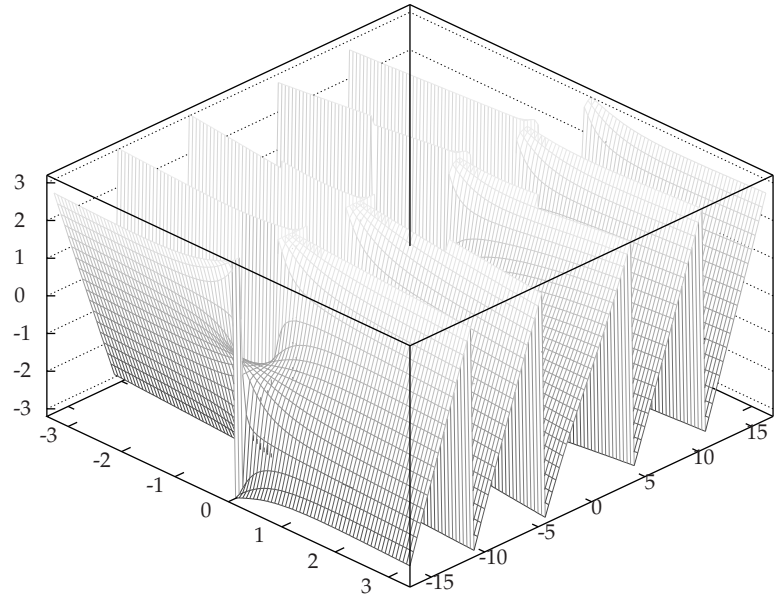
The basic idea of using the Laplace transform for systems that are described in terms of differential equations is represented in the diagram below:



In order to determine the output signal $y(t)$, we will not solve the differential equation, but transform the input signal and the system's description to the Laplace domain, and solve for the output signal there. The output signal can then be inversely transformed to obtain it in the time domain. In case the one-sided Laplace transform is considered, the initial conditions can also be taken into account easily as we will see in the applications below.



(a) Magnitude plot



(b) Phase plot

Figure 6.3: Graphical representation of the Laplace transform of example (6.4)

6.9 Applications

6.9.1 Solving differential equations

One of the major applications for the Laplace transform is to solve differential equations.

Differential equation

A differential equation is a relationship of the following type:

$$F(y^{(n)}, \dots, y^{(3)}, y'', y', y, t) = x(t) \tag{6.6}$$

with $n \in \mathbb{N}$.
In this equation $y^{(n)}$ represents the n-th derivative of $y(t)$ w.r.t. t. Any signal $y(t)$ that satisfies (6.6) is a so-called *solution* of the differential equation.
The value of n is referred to as the *order* of the differential equation.

Example

Consider the differential equation:

$$y' - y = 0$$

Obviously $y = \mathrm{e}^t$ is a solution to this equation. However, this is not the only possible solution. Consider for example another possible solution: $y(t) = 2\,\mathrm{e}^t$. In fact, any signal of the form $y(t) = c\,\mathrm{e}^t$ with $c \in \mathbb{C}$ is a solution.

Finding solutions for a differential equation is referred to as *solving* it. A major question is how to solve such a differential equation? There is a plethora of solution methods, even to such an extent that almost every type has its own solution method. However, it turns out that for the category of linear differential equations (i.e. linear in the factors $y^{(i)}$), the Laplace transform offers a very smooth way to solve the entire category.

Solving differential equations using the Laplace transform

Consider the differential equation:

$$y''(t) + 2y'(t) + y(t) = \mathrm{e}^{-t}$$

subject to $y(0) = 0$ and $y'(0) = 1$.

We will attempt to solve this equation by transforming both sides of the equation, using the Laplace transform. Keep in mind that by using the (one-sided) Laplace transform, we implicitly assume $y(t)$ to be causal!

$$\mathcal{L}\left(y''(t) + 2y'(t) + y(t)\right) = \mathcal{L}\left(\mathrm{e}^{-t}\right)$$

linearity of the Laplace transform $\downarrow$

$$\mathcal{L}(y''(t)) + 2\mathcal{L}(y'(t)) + \mathcal{L}(y(t)) = \mathcal{L}\left(\mathrm{e}^{-t}\right)$$

Let's set $\mathcal{L}(y(t)) = Y(s)$ and calculate some of the terms involved:

$$\begin{aligned}
\mathcal{L}(y'(t)) &= sY(s) - y(0) \\
\mathcal{L}(y''(t)) &= s(sY(s) - y(0)) - y'(0) = s^2Y(s) - sy(0) - y'(0) \\
\mathcal{L}(\mathrm{e}^{-t}) &= \frac{1}{s+1}
\end{aligned}$$

Therefore:

$$\begin{aligned}
& s^2Y(s) - sy(0) - y'(0) + 2(sY(s) - y(0)) + Y(s) = \frac{1}{s+1} \\
\Leftrightarrow \quad & s^2Y(s) - 1 + 2sY(s) + Y(s) = \frac{1}{s+1}
\end{aligned}$$

Solving for $Y(s)$ yields:

$$\begin{aligned}
& \left(s^2 + 2s + 1\right) Y(s) = \frac{1}{s+1} + 1 \\
\Leftrightarrow \quad & (s+1)^2 Y(s) = \frac{s+2}{s+1} \\
\Leftrightarrow \quad & Y(s) = \frac{s+2}{(s+1)^3}
\end{aligned}$$

Now, we can determine $y(t)$ by inverse Laplace transform. To this end, we use the procedure for ratios of polynomials.

$$\begin{aligned}
Y(s) &= \frac{s+2}{(s+1)^3} \\
&\downarrow \quad \text{partial fraction expansion (PFE)} \\
Y(s) &= \frac{1}{(s+1)^2} + \frac{1}{(s+1)^3} \\
y(t) &= \mathcal{L}^{-1}\left(\frac{1}{(s+1)^2} + \frac{1}{(s+1)^3}\right) \\
&= \mathcal{L}^{-1}\left(\frac{1}{(s+1)^2}\right) + \mathcal{L}^{-1}\left(\frac{1}{(s+1)^3}\right) \\
&= t\,\mathrm{e}^{-t}u(t) + \frac{1}{2}t^2\,\mathrm{e}^{-t}u(t)
\end{aligned}$$

Exercises

Exercise 6.9.1-1:
Solve the following differential equations, subject to the listed boundary conditions:

a)	$y'(t) + y(t) = 4$	with $y(0) = 2$	
b)	$y'(t) + y(t) = t\,\mathrm{e}^{-t}$	with $y(0) = 1$	
c)	$y'(t) + y(t) = \mathrm{e}^{-t}\sin t$	with $y(0) = 1$	
d)	$y''(t) + 4y'(t) + 3y(t) = \mathrm{e}^{-t}\sin t$	with $y(0) = 0$	and $y'(0) = 0$
e)	$y''(t) + 3y'(t) + 2y(t) = 4\,\mathrm{e}^{t}$	with $y(0) = 1$	and $y'(0) = -1$
f)	$y''(t) - 3y'(t) - 4y(t) = 2\,\mathrm{e}^{-t}$	with $y(0) = 1$	and $y'(0) = 1$
g)	$y''(t) + 4y'(t) + 5y(t) = 0$	with $y(0) = 1$	and $y'(0) = -2$
h)	$y''(t) - y'(t) - 2y(t) = 2\,\mathrm{e}^{t}$	with $y(0) = 3$	and $y'(0) = 7$

Exercise 6.9.1-2:
Solve the following differential equations, subject to the listed boundary conditions:

a)	$ty''(t) + (1-2t)y'(t) - 2y(t) = 0$	with $y(0) = 1$	and $y'(0) = 2$	
b)	$ty''(t) + y'(t) + ty = 0$	with $y(0) = 5$	and $y'(0) = 0$	
c)	$ty''(t) - 4ty'(t) - 4y(t) = 0$	with $y(0) = 0$	and $y'(0) = 1$	
d)	$ty''(t) + 2y'(t) + 16ty = 8\cos 4t$	with $y(0) = 0$	and $y'(0) = 4$	
e)	$ty''(t) + 2y'(t) + ty = 3t - 2\cos t$	with $y(0) = 3$	and $y'(0) = -1$	
f)	$ty''(t) - y'(t) = -2$	with $y(0) = 3$	and $y'(0) = 2$	and $y''(0) = 2$

6.9.2 Solving linear electronic networks

Instead of writing down the differential equations to solve them afterwards, we can take a quicker route that works for many LTI systems: we can also write the equations directly in the Laplace domain, where solving them corresponds to solving algebraic equations instead of differential equations. We can even take into account initial conditions without extra effort.

Especially for linear electronic circuits, this works. Instead of writing down Kirchhoff's laws and the branch equations in the time domain, we will do this in the Laplace domain. To this end, we transform the model of our electronic circuit to the Laplace domain.

We will discuss the next two situations separately:

- zero initial conditions
- non-zero initial conditions

Zero initial conditions

This corresponds to replacing every network element by an equivalent network element in the Laplace domain. Table 6.2 summarizes the transformation.

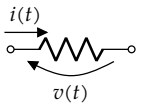
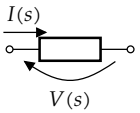
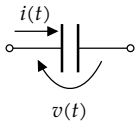
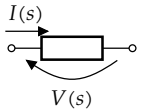
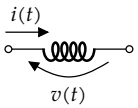
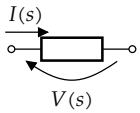
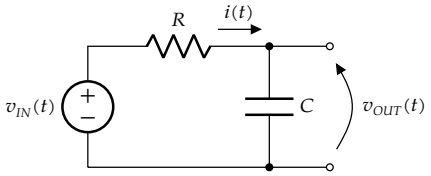
Time domain		Laplace domain
Resistor		
	$\mathcal{L}$ →	
$v(t) = R \cdot i(t)$	$\mathcal{L}$ →	$V(s) = \underbrace{R}_{\equiv Z} \cdot I(s)$
Capacitor		
	$\mathcal{L}$ →	
$i(t) = C \cdot \frac{dv(t)}{dt}$	$\mathcal{L}$ →	$I(s) = \underbrace{Cs}_{\equiv 1/Z} \cdot V(s)$
Inductor		
	$\mathcal{L}$ →	
$v(t) = L \cdot \frac{di(t)}{dt}$	$\mathcal{L}$ →	$V(s) = \underbrace{Ls}_{\equiv Z} \cdot I(s)$

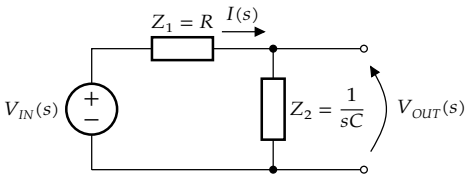
Table 6.2: Transformation of linear network elements to the Laplace domain

Example 1

Consider the RC filter below:



Applying the transformation to the Laplace domain, results in:



The transfer function (i.e. the ratio of the output over the input) can now be readily calculated:

$$\frac{V_{OUT}(s)}{V_{IN}(s)} = \frac{Z_2}{Z_1 + Z_2} = \frac{1/sC}{R + 1/sC} = \frac{1}{1 + sRC}$$

Example 2

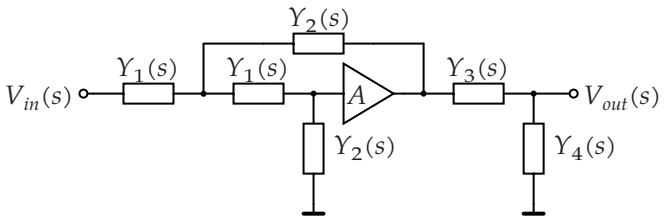


Figure 6.4: Schematics of the Sallen-Key topology augmented with a passive output filter

As an example, consider the well-known Sallen-Key circuit (augmented with a passive output filter) of Figure 6.4.

Assuming the circuit starts with zero initial state, the voltage transfer function of this circuit can be derived to be:

$$\frac{V_{out}(s)}{V_{in}(s)} = \frac{Y_3(s)}{Y_3(s)+Y_4(s)} \cdot \frac{AY_1^2(s)}{Y_1^2(s)+(3-A)Y_1(s)Y_2(s)+Y_2^2(s)} \tag{6.7}$$

If we assume

$$Y_1(s) = 1/R_1 \qquad Y_2(s) = sC_1$$
$$Y_3(s) = sC_2 \qquad Y_4(s) = 1/R_2$$

we can rework (6.7) to

$$\frac{V_{out}(s)}{V_{in}(s)} = \frac{sR_2C_2}{1+sR_2C_2} \cdot \frac{A}{s^2R_1^2C_1^2+s(3-A)R_1C_1+1}$$

From this equation, the location of poles and zeros can be determined:

$$p_0 = -\frac{1}{R_2C_2} \qquad n_0 = 0$$
$$p_{1,2} = \frac{-(3-A) \pm \sqrt{(A-1)(A-5)}}{2R_1C_1}$$

If we take

$$R_1 = 1\,\text{k}\Omega \qquad R_2 = 10\,\Omega \qquad A = 2$$
$$C_1 = 1\,\text{pF} \qquad C_2 = 1\,\text{nF}$$

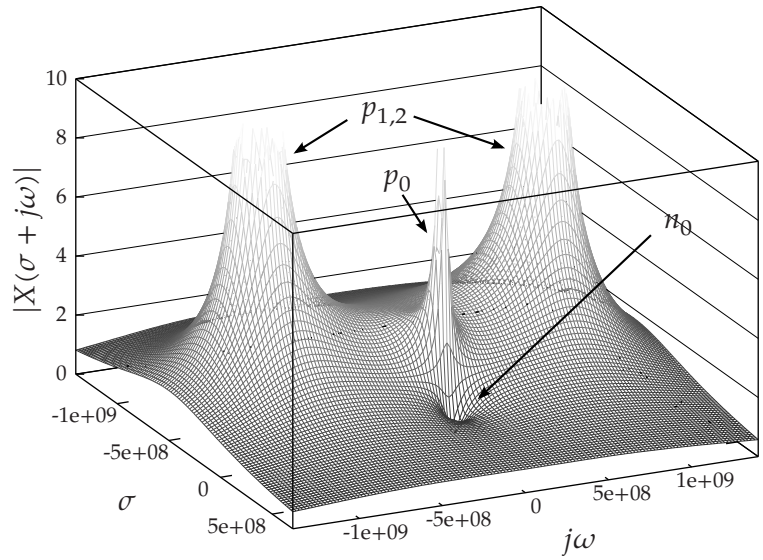
we can generate the graphical plots of Figure 6.5 on the facing page. In these plots, the location of the poles and zeros are clearly visible. The poles and zeros can be calculated to be:

$$p_0 = -100 \times 10^6\,\text{rad/s} \qquad n_0 = 0$$
$$p_{1,2} = (-500 \times 10^6 \pm \text{j}866 \times 10^6)\text{rad/s}$$

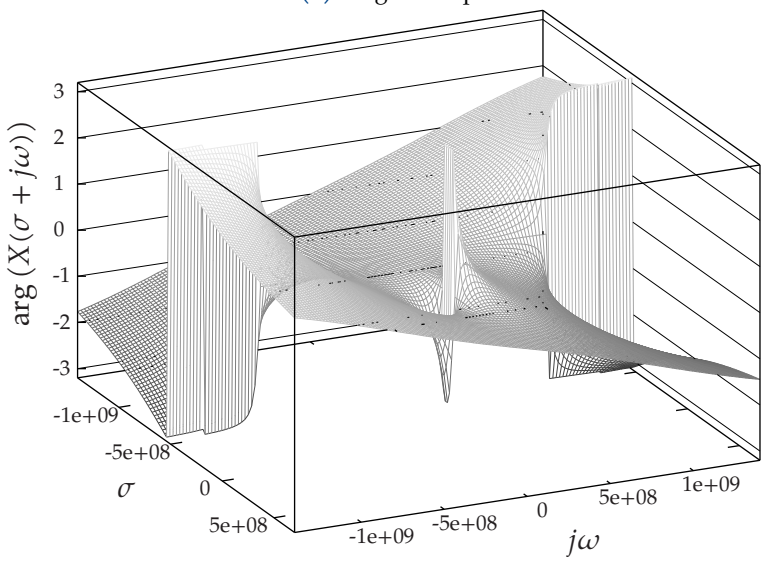
Very often, these plots are displayed in a simplified form, as a 2-D pole-zero diagram, as in Figure 6.6.

Nonzero initial conditions

This corresponds to replacing every network element by a set of equivalent network elements in the Laplace domain. For a resistor, this is the same mapping as in the case of zero-initial conditions (as the resistor is not capable of keeping a state), while for a capacitor and an inductor, the situation is more complicated.



(a) Magnitude plot



(b) Phase plot

Figure 6.5: Graphical representation of the Laplace transform of the augmented Sallen-Key filter of (6.7)

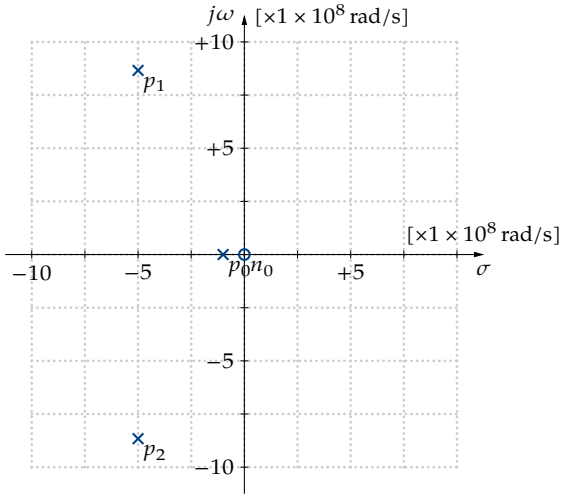


Figure 6.6: Pole-zero diagram of the augmented Sallen-Key filter of (6.7)

We start by writing down the branch equation of a capacitor, both in differential and integral form. Then we transform those equations to the Laplace domain:

$$\begin{aligned} i(t) &= C\frac{\mathrm{d}v(t)}{\mathrm{d}t} & v(t) &= \frac{1}{C}\int_0^t i(\tau)\,\mathrm{d}\tau + v(0) \\ &\downarrow \mathcal{L} & &\downarrow \mathcal{L} \\ I(s) &= C\left(sV(s) - v(0)\right) & V(s) &= \frac{1}{Cs}I(s) + \frac{v(0)}{s} \end{aligned}$$

Table 6.3 summarizes the transformation. Please, note the direction of the arrows and the voltage and current sources! The above equations on the left correspond to the parallel variant (B), the equations on the right to series variant (A).

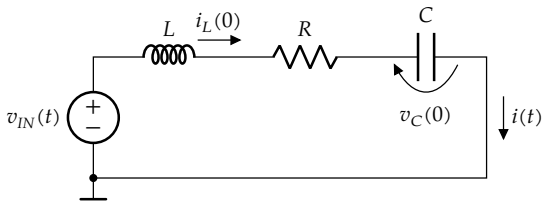
Let's do the same for the branch equation of the inductor:

$$\begin{aligned} v(t) &= L\frac{\mathrm{d}i(t)}{\mathrm{d}t} & i(t) &= \frac{1}{L}\int_0^t v(\tau)\,\mathrm{d}\tau + i(0) \\ &\downarrow \mathcal{L} & &\downarrow \mathcal{L} \\ V(s) &= L\left(sI(s) - i(0)\right) & I(s) &= \frac{1}{Ls}V(s) + \frac{i(0)}{s} \end{aligned}$$

Table 6.3 summarizes the transformation. Please, note the direction of the arrows and the voltage and current sources! The above equations on the left correspond to series variant (B), the equations on the right to parallel variant (A).

Example

Consider the network below. Given the input $v_{IN}(t)$ let's try to calculate the output $i(t)$ taking into account the nonzero initial conditions of the inductor and the capacitor.



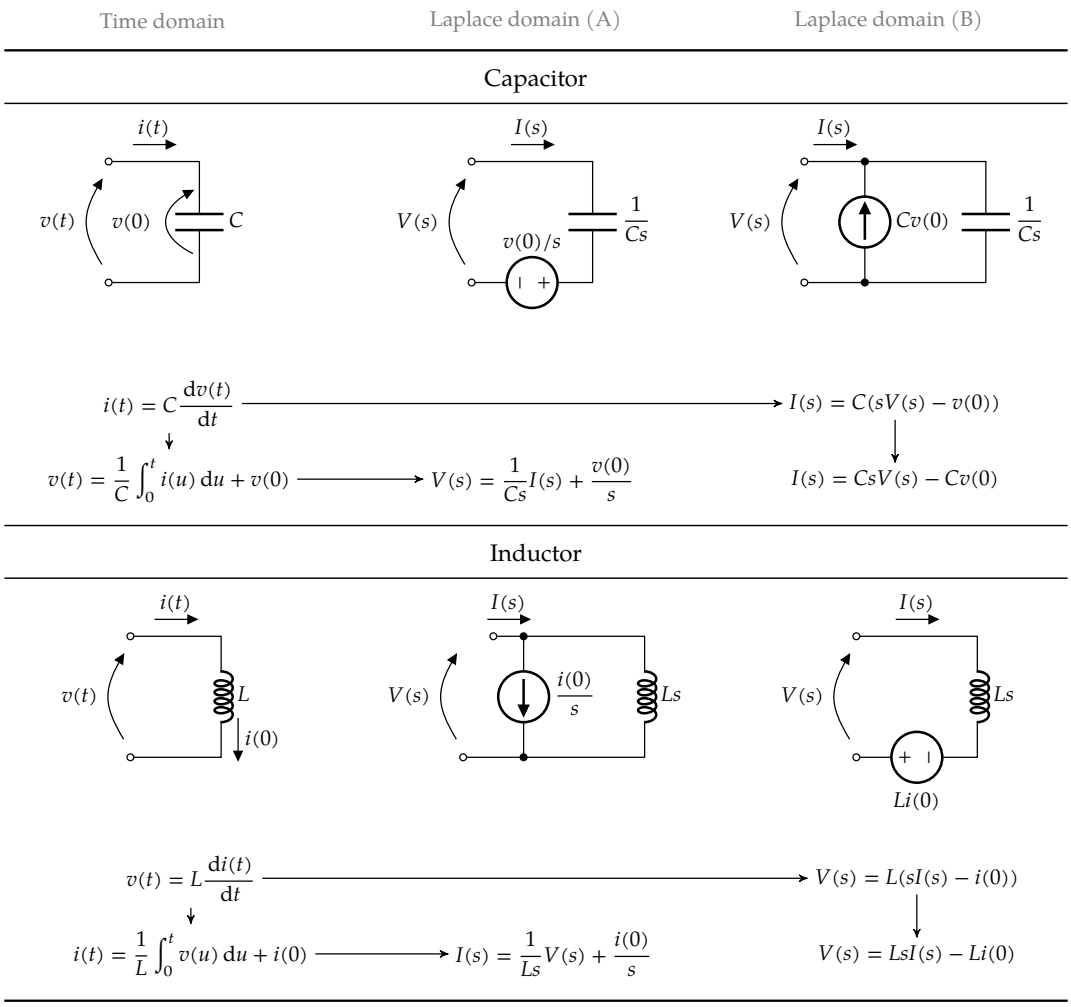
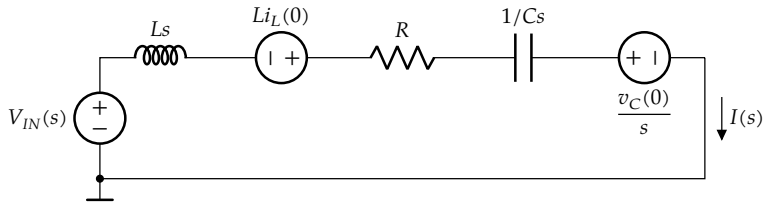
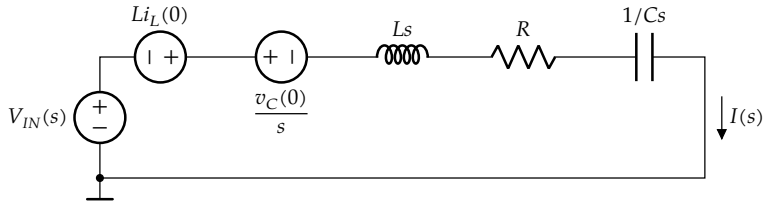


Table 6.3: Transformation of linear network elements to the Laplace domain taking into account nonzero initial conditions

Next, we apply the transformations of Table 6.3. We opt for the series equivalents as we only have a single loop. Using the parallel variants would increase the number of loops in the network, making its solution more involved.



We can rearrange this network to put all sources on the left-hand side an all passive elements on the right-hand side:



This allows a straightforward calculation of the output:

$$I(s) = \frac{V_{IN}(s) + Li_L(0) - v_c(0)/s}{Ls + R + 1/Cs}$$

If the initial conditions would have been zero, this equation reduces to:

$$I(s) = \frac{V_{IN}(s)}{Ls + R + 1/Cs}$$

and once again, the transfer function can be drived in that case:

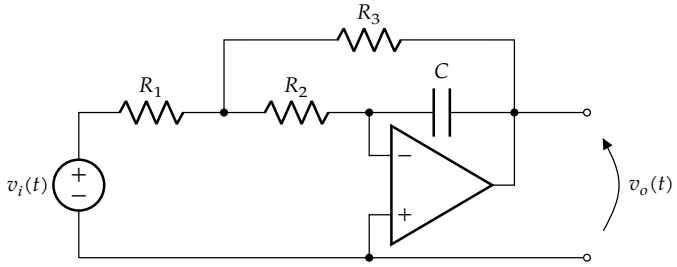
$$H(s) = \frac{I(s)}{V_{IN}(s)} = \frac{Cs}{1 + RCs + LCs^2}$$

Exercises

Note that the exercises below are only on determining transfer functions (in the Laplace domain). We will keep the initial value problems for Chapter 7.

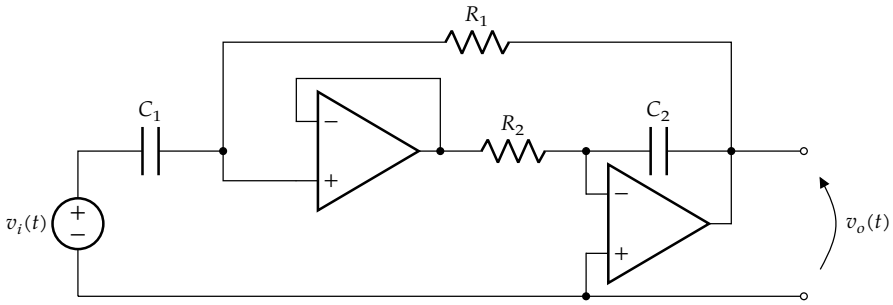
Exercise 6.9.2-1:

Determine the transfer function of the circuit below with $v_i(t)$ as input and $v_o(t)$ as output.



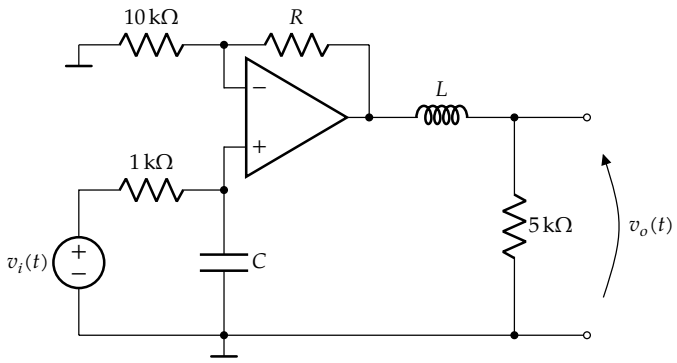
Exercise 6.9.2-2:

Determine the transfer function of the circuit below with $v_i(t)$ as input and $v_o(t)$ as output.



Exercise 6.9.2-3:

Consider the circuit below with $v_i(t)$ as input and $v_o(t)$ as output. Determine R, C and L such that the transfer function of this circuit is given by: $H(s) = \dfrac{15 \times 10^6}{(s + 2000)(s + 5000)}$



6.10 The extended Fourier family diagram

In this chapter, we saw how the Laplace transform can be derived from the Fourier transform. Because of these relationships, it is logical to add the Laplace transform to the Fourier family diagram.

You can find the extended Fourier family diagram in Figure 6.7.

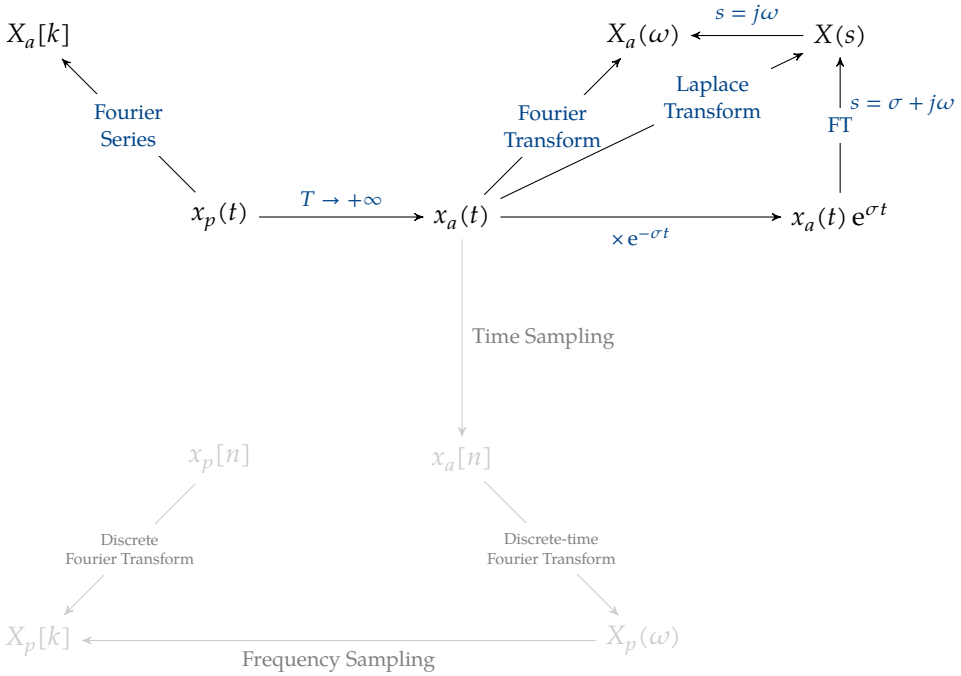


Figure 6.7: The Fourier family photo with the nephew Laplace on it; the bottom half has been grayed out as this will be the subject of one of your next courses on DSP.

Chapter 7

Analyzing LTI systems — time domain

In this chapter, you will:

- learn how to analyze LTI systems in the time domain (taking into account some initial conditions),
- study some simple prototype systems (first order, second order, with and without zeros).

After having read/studied this chapter, you are expected to be able to

- analyze LTI systems in the time-domain yourself,
- use your knowledge of the prototype systems to understand the behavior of more complicated systems.

7.1 Introduction

In this chapter, we will study LTI systems in the time domain. However, the Laplace transform will be one of the blades of our Swiss army knife that will allow us to avoid solving differential equations explicitly.

Once we master the basic technique, we will study some prototype all pole systems (first order, second order) and then see how zeros change their behavior. Having a good understanding of these prototype systems in conjunction with the location of their poles and zeros will allow us to get a good basic understanding of the behavior of more complex systems. We will study these basic systems in the setting of *zero initial conditions* i.e. there is no energy stored in the system at $t = 0$.

Afterwards, we will also see how easy it is to study these systems given a number of nonzero initial conditions. This will prove the added value of the Laplace transform over the Fourier transform.

We will finish the chapter with a number of examples that illustrate the techniques that we have learnt throughout this chapter.

7.2 Basic analysis technique

7.2.1 Principle

The basic technique to analyze LTI systems in the time-domain has been displayed below.

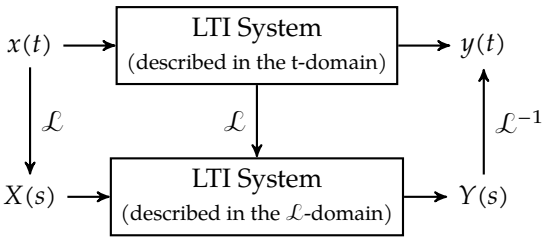


Figure 7.1: Basic principle to analyze a system in the time-domain: convert the problem to the Laplace domain and solve it there.

We transform the input signal and the system's description to the Laplace domain and solve for the output:

$$Y(s) = H(s) \cdot X(s)$$

Afterwards, we can obtain the output signal in the time domain by inverse Laplace transformation. In general, we call this a *transient analysis*, as we analyze how the system's output evolves over time given a certain input and possibly an initial internal state of the system. The term *transient* suggests that we go from one stable state to another through some intermediary states. Indeed, that is exactly what we will observe in many cases.

To complete the picture, if the constellation of what is known about the system is different, e.g., we know the input and the output, but not the system description of the LTI system, we can also derive that easily by transforming input and output to the Laplace domain and determining the transfer function as:

$$H(s) = \frac{Y(s)}{X(s)}$$

Likewise, when a desired output and the system is known, we can try to find what input we need to provide, using:

$$X(s) = \frac{Y(s)}{H(s)}$$

Of course, in both cases we can find the corresponding unknowns in the time-domain (the impulse response and the input signal) by inverse Laplace transform.

7.2.2 Transfer function

In general, an LTI system will have a rational transfer function, i.e. a ratio of two polynomials in s:

$$H(s) = \frac{b_m s^m + b_{m-1} s^{m-1} + \cdots + b_2 s^2 + b_1 s + b_0}{a_n s^n + a_{n-1} s^{n-1} + \cdots + a_3 s^3 + a_2 s^2 + a_1 s + a_0} = \frac{\sum_{j=0}^{m} b_j s^j}{\sum_{i=0}^{n} a_i s^i} \tag{7.1}$$

We call the denominator of this transfer function the *characteristic polynomial* of the LTI system as it describes the nature of the system, irrespective of the specific input and output of the system:

it characterizes the system. The degree of the characteristic polynomial is called the *order* of the system.

Every feasible system will exhibit the following property:

Feasible system condition lemma

A feasible LTI system with transfer function

$$H(s) = \frac{\sum_{j=0}^{m} b_j s^j}{\sum_{i=0}^{n} a_i s^i}$$

will obey the following property:

$$m \leq n$$

We can easily prove this lemma by contradiction:

Proof
Let's calculate the initial value of this system when confront it with a step input $x(t) = \mathrm{u}(t)$. We know that

$$y(0) = \lim_{s\to\infty} sY(s) = \lim_{s\to\infty} s \cdot H(s) \cdot X(s)$$

Knowing that $\mathrm{u}(t) \xrightarrow{\mathcal{L}} 1/s$, we can deduce:

$$y(0) = \lim_{s\to\infty} H(s)$$

Therefore, if $m > n$, then $y(0) \to \infty$, which cannot be the case for a feasible system. This proves our lemma by contradiction. ■

We can factorize (7.1) into:

$$H(s) = K \cdot \frac{\prod_{j=1}^{m}(s - z_j)}{\prod_{i=1}^{n}(s - p_i)}$$

with $K = b_m/a_n$, and z_j and p_i complex.

We can even go one step further as in:

$$H(s) = K \cdot \frac{\prod_{j=1}^{m_1}(s - z_j) \cdot \prod_{j=1}^{m_2}(s - (\sigma_{z,j} + \mathrm{j}\omega_{z,j}))(s - (\sigma_{z,j} - \mathrm{j}\omega_{z,j}))}{\prod_{i=1}^{n_1}(s - p_i) \cdot \prod_{i=1}^{n_2}(s - (\sigma_{p,i} + \mathrm{j}\omega_{p,i}))(s - (\sigma_{p,i} - \mathrm{j}\omega_{p,i}))}$$

with all $z_j, p_i, \sigma_{z,j}, \omega_{z,j}, \sigma_{p,i}, \omega_{p,i}$ real.

Indeed, all poles and zeros come in complex conjugate pairs. We have m zeros of which m_1 are real and $2m_2$ are complex conjugate pairs. We have n poles of which n_1 are real and $2n_2$ are complex conjugate pairs.

Of course: $m = m_1 + 2m_2$ and $n = n_1 + 2n_2$.

We will see later that the *intrinsic behavior* of a system is mainly governed by the characteristic polynomial (i.e. the location of the poles). Therefore, let's first study simple all-pole systems:

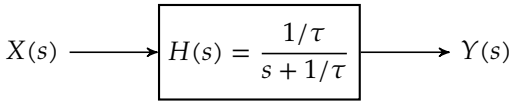
- first-order systems: systems containing a single real pole
- second-order systems: systems containing a complex-conjugate pole pair.

These will prove to be the building blocks of all LTI-systems. We'll add zeros later.

7.3 First-order systems

7.3.1 Description

Consider the following generic first-order system (with $\tau \in \mathbb{R}_0$):

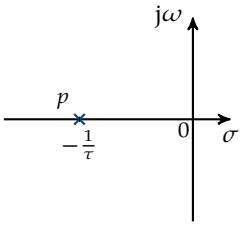


This particular form (in which the coefficient of s is one and $|H(0)| = 1$) is called the *normal form*. We call τ the *time constant* of the first-order system. It will not suprise you that its unit is second.

The pole of this system is determined by setting the characteristic polynomial to zero and solving for s. let's call the solution p:

$$s + \frac{1}{\tau} = 0$$
$$\Leftrightarrow \qquad p = -\frac{1}{\tau}$$

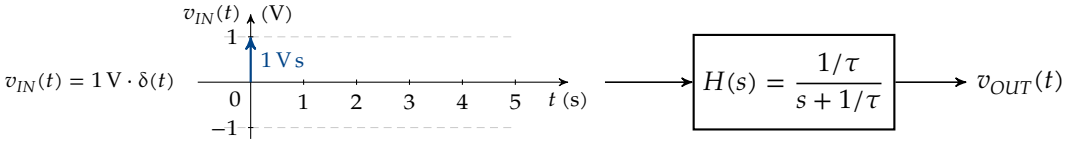
For stable systems, the pole $p = -\frac{1}{\tau}$ is located in the left-half plane (i.e. $\tau > 0$), as indicated below:



If $\tau < 0$ and hence $|p| > 0$, we have a pole in the right-half plane, i.e. then the system is not stable.

7.3.2 Impulse response

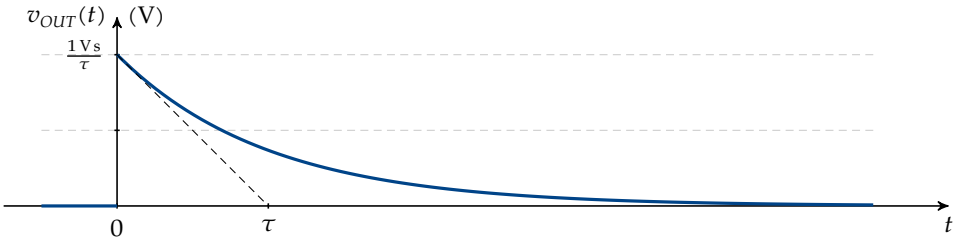
Let's excite our first-order system using a Dirac impulse and calculate the output v_{OUT}.



The analysis is straightforward:

$$
\begin{aligned}
V_{OUT}(s) &= \frac{1/\tau}{s+1/\tau} V_{IN}(s) \\
&\quad \Big\downarrow v_{IN}(t) = 1\,\mathrm{V}\cdot\delta(t) \quad \xrightarrow{\mathcal{L}} \quad V_{IN}(s) = 1\,\mathrm{V\,s} \\
V_{OUT}(s) &= 1\,\mathrm{V\,s}\cdot\frac{1}{\tau}\frac{1}{s+1/\tau} \\
&\quad \Big\downarrow \mathcal{L}^{-1} \\
v_{OUT}(t) &= 1\,\mathrm{V\,s}\cdot\frac{1}{\tau}\,\mathrm{u}(t)\;\mathrm{e}^{-t/\tau}
\end{aligned}
$$

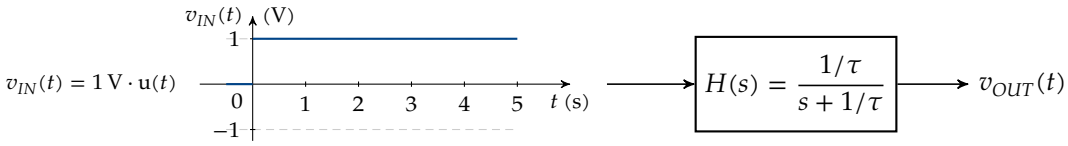
Graphically:



One can prove that the tangential line to the graph for $t = 0$, intersects with the t-axis at $t = \tau$. At that instant of time $t = \tau$, the curve has decayed to about 1/3 of its value. This can help you in drawing a correct graph.

7.3.3 Step response

Next, let's excite our first-order system using a step input and calculate the output v_{OUT}.

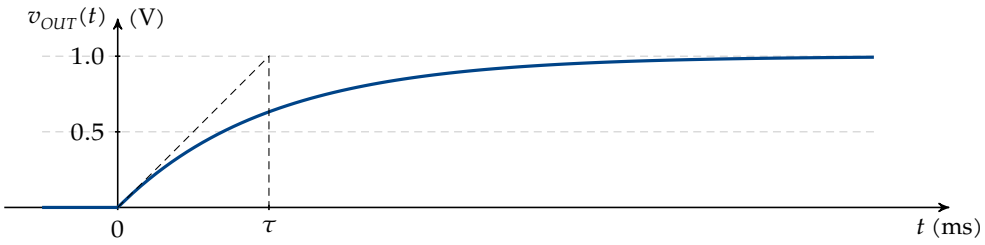


The analysis is again straightforward:

$$
\begin{aligned}
V_{OUT}(s) &= \frac{1/\tau}{s+1/\tau} V_{IN}(s) \\
&\quad \Big\downarrow v_{IN}(t) = 1\,\mathrm{V}\cdot\mathrm{u}(t) \quad \xrightarrow{\mathcal{L}} \quad V_{IN}(s) = 1\,\mathrm{V}\cdot\frac{1}{s} \\
V_{OUT}(s) &= 1\,\mathrm{V}\cdot\frac{1/\tau}{s+1/\tau}\frac{1}{s} \\
&\overset{\text{(PFE)}}{=} 1\,\mathrm{V}\cdot\left(\frac{1}{s} - \frac{1}{s+1/\tau}\right) \\
&\quad \Big\downarrow \mathcal{L}^{-1} \\
v_{OUT}(t) &= 1\,\mathrm{V}\cdot\mathrm{u}(t)\;\left(1-\mathrm{e}^{-t/\tau}\right)
\end{aligned}
$$

in which we used *partial fraction expansion* (*PFE*) as one of the standard techniques to allow calculating the inverse Laplace transform.

Graphically:



Again, one can prove that the tangential line to the graph for $t = 0$, intersects with the horizontal line of the final value at $t = \tau$. At that instant of time $t = \tau$, the curve has reached about 2/3 of its final value. This can help you in drawing a correct graph.

7.4 Second-order systems

7.4.1 Description

Consider the following generic second-order system (with $\zeta, \omega_n \in \mathbb{R}$):

$$X(s) \rightarrow \boxed{H(s) = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}} \rightarrow Y(s)$$

This particular form (in which the coefficient of s^2 is one and $|H(0)| = 1$) is called the *normal form*. We call ω_n the *natural frequency* and ζ the *damping factor* of the second-order system. It will not suprise you that the unit of the first is radian per second. The second one is dimensionless.

The poles of this system are determined by setting the characteristic polynomial to zero and solving for s:

$$s^2 + 2\zeta\omega_n s + \omega_n^2 = 0$$

Therefore:

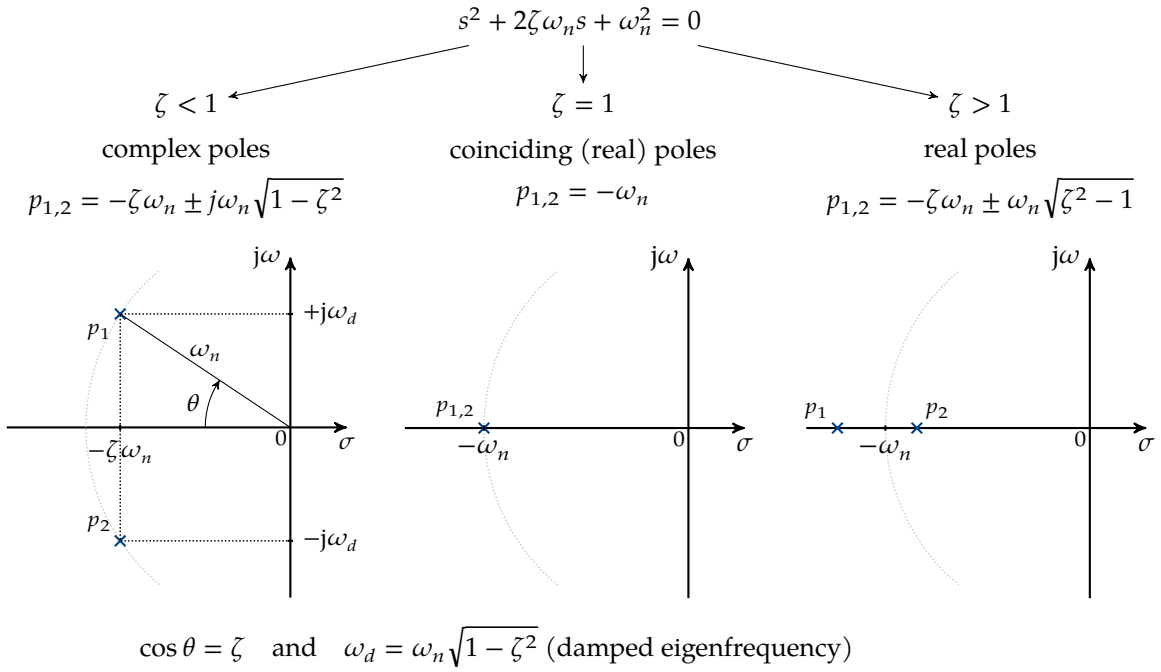
$$p_{1,2} = \frac{-2\zeta\omega_n \pm \sqrt{4\omega_n^2(\zeta^2 - 1)}}{2} = -\zeta\omega_n \pm \omega_n\sqrt{\zeta^2 - 1}$$

We can now distinguish three separate cases:

- $\zeta > 1$
- $\zeta = 1$
- $\zeta < 1$

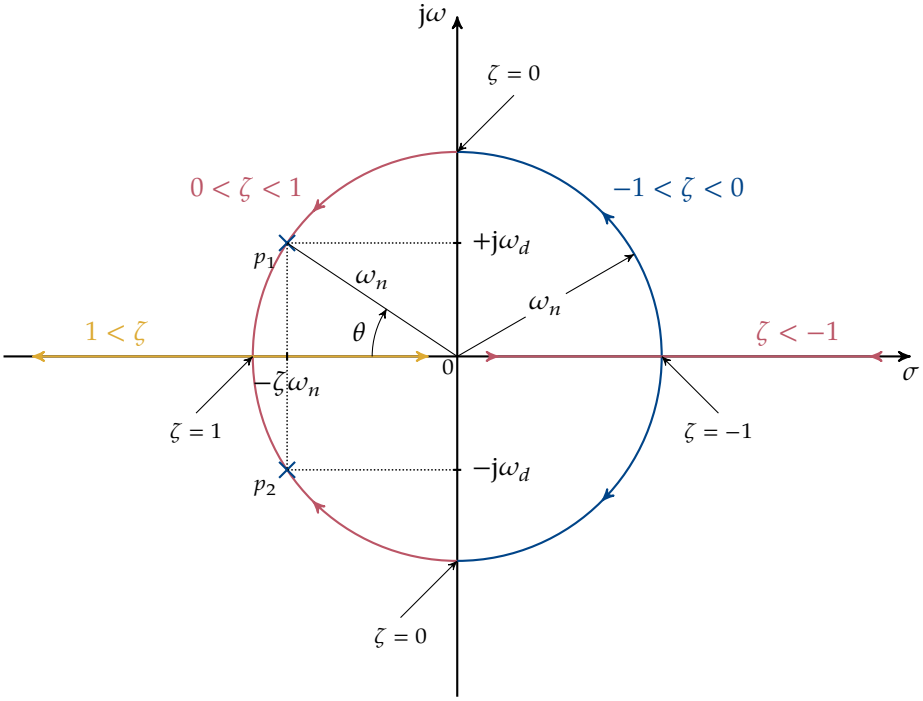
These cases have been elaborated in Figure 7.2a.

Note that we only considered values of $\zeta > 0$. Of course negative values are also possible. In fact the value of ζ causes the poles to migrate on what we call a root locus (see Figure 7.2b). For very negatively large values of ζ one pole starts in the origin, the other at $+\infty$. For increasing values of ζ they move closer together, until the meet at the real value ω_n (for $\zeta = -1$). Then they move on a circle with radius ω_n towards $\pm j\omega_n$ (for $\zeta = 0$, to continue on a mirrored path



$\cos\theta = \zeta$ and $\omega_d = \omega_n\sqrt{1-\zeta^2}$ (damped eigenfrequency)

(a) Typical cases for three positive values of ζ



(b) Root locus plot for any real ζ

Figure 7.2: Location of the complex conjugate poles of a second-order system in normal form, depending on the damping factor ζ; poles p_1 and p_2 have been drawn for $\zeta \approx 0.8$.

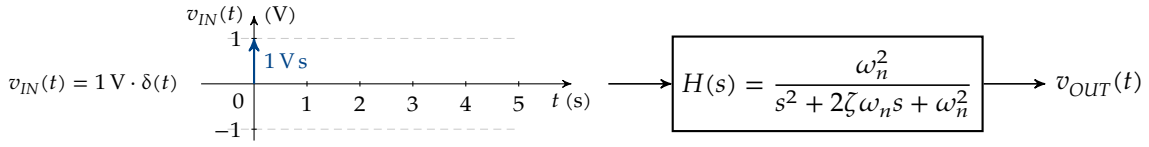
.

towards the real value $-\omega_n$ where they meet again for $\zeta = 1$. Afterwards they move apart (for larger values of ζ), one moving towards $-\infty$ and the other towards the origin.

For stable systems, the poles $p_{1,2}$ are located in the left-half plane. We will analyze the behavior of these stable systems in the experiments below.

7.4.2 Impulse response

Let's excite our second-order system using a Dirac impulse and calculate the output v_{OUT}.



The analysis is straightforward. Let's consider the three cases. You can find the graphs corresponding to the obtained results for the three cases in Figure 7.3.

- $\zeta > 1$:

$$
\begin{aligned}
V_{OUT}(s) &= \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} V_{IN}(s) \\
&\quad \Big\downarrow \; v_{IN}(t) = 1\,\mathrm{V} \cdot \delta(t) \quad \overset{\mathcal{L}}{\longrightarrow} \quad V_{IN}(s) = 1\,\mathrm{V\,s} \\
V_{OUT}(s) &= 1\,\mathrm{V\,s} \cdot \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} = 1\,\mathrm{V\,s} \cdot \frac{p_1 p_2}{(s - p_1)(s - p_2)} \\
&\overset{\text{(PFE)}}{=} 1\,\mathrm{V\,s} \cdot \frac{p_1 p_2}{p_1 - p_2} \left(\frac{1}{s - p_1} - \frac{1}{s - p_2} \right) \\
&\quad \Big\downarrow \; \mathcal{L}^{-1} \\
v_{OUT}(t) &= 1\,\mathrm{V\,s} \cdot \mathrm{u}(t) \; \frac{p_1 p_2}{p_1 - p_2} \left(\mathrm{e}^{p_1 t} - \mathrm{e}^{p_2 t} \right)
\end{aligned}
$$

- $\zeta = 1$:

$$
\begin{aligned}
V_{OUT}(s) &= \frac{\omega_n^2}{s^2 + 2\omega_n s + \omega_n^2} V_{IN}(s) \\
&\quad \Big\downarrow \; v_{IN}(t) = 1\,\mathrm{V} \cdot \delta(t) \quad \overset{\mathcal{L}}{\longrightarrow} \quad V_{IN}(s) = 1\,\mathrm{V\,s} \\
V_{OUT}(s) &= 1\,\mathrm{V\,s} \cdot \frac{\omega_n^2}{s^2 + 2\omega_n s + \omega_n^2} = 1\,\mathrm{V\,s} \cdot \frac{\omega_n^2}{(s + \omega_n)^2} \\
&\quad \Big\downarrow \; \mathcal{L}^{-1} \\
v_{OUT}(t) &= 1\,\mathrm{V\,s} \cdot \omega_n^2 \cdot \mathrm{u}(t) \cdot t \cdot \mathrm{e}^{-\omega_n t}
\end{aligned}
$$

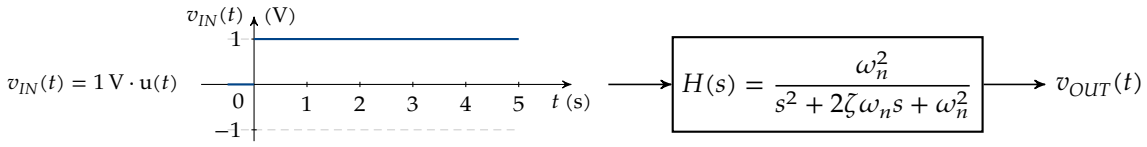
- $\zeta < 1$:

$$
\begin{aligned}
V_{OUT}(s) &= \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} V_{IN}(s) \\
&\quad \Big\downarrow v_{IN}(t) = 1\,\mathrm{V} \cdot \delta(t) \quad \overset{\mathcal{L}}{\longrightarrow} \quad V_{IN}(s) = 1\,\mathrm{V\,s} \\
V_{OUT}(s) &= 1\,\mathrm{V\,s} \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} = 1\,\mathrm{V\,s} \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \zeta^2\omega_n^2 - \zeta^2\omega_n^2 + \omega_n^2} \\
&= 1\,\mathrm{V\,s} \frac{\omega_n^2}{(s + \zeta\omega_n)^2 + (\underbrace{\omega_n\sqrt{1-\zeta^2}}_{\equiv\omega_d})^2} \\
&= 1\,\mathrm{V\,s} \cdot \frac{\omega_n}{\sqrt{1-\zeta^2}} \frac{\omega_d}{(s + \zeta\omega_n)^2 + \omega_d^2} \\
&\quad \Big\downarrow \mathcal{L}^{-1} \\
v_{OUT}(t) &= 1\,\mathrm{V\,s} \cdot \frac{\omega_n}{\sqrt{1-\zeta^2}}\ \mathrm{u}(t)\ \mathrm{e}^{-\zeta\omega_n t} \sin\omega_d t
\end{aligned}
$$

An overall view of the effect of ζ on the impulse response, can be found in Figure 7.4.

7.4.3 Step response

Next, let's excite our second-order system using a step input and calculate the output v_{OUT}.

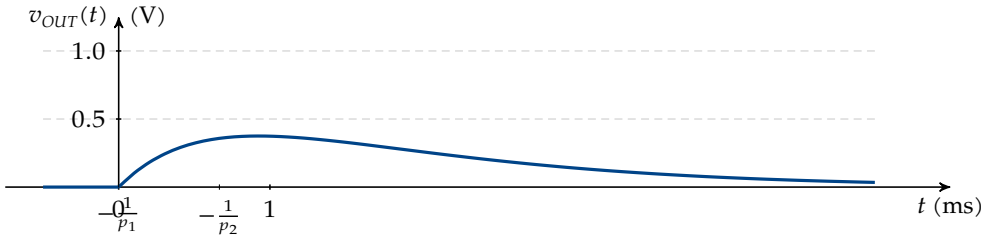


The analysis is straightforward. Let's consider the three cases. You can find the graphs corresponding to the obtained results for the three cases in Figure 7.5.

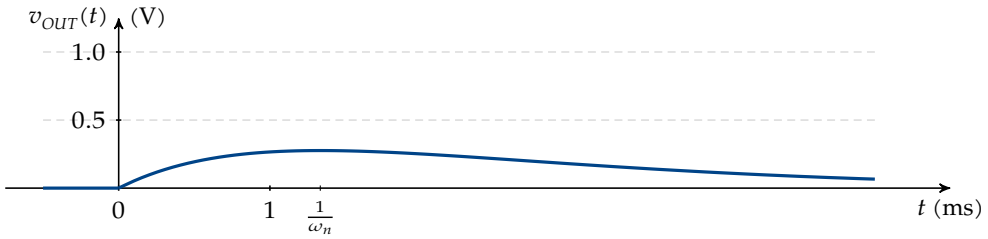
- $\zeta > 1$:

$$
\begin{aligned}
V_{OUT}(s) &= \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} V_{IN}(s) \\
&\quad \Big\downarrow v_{IN}(t) = 1\,\mathrm{V} \cdot \mathrm{u}(t) \quad \overset{\mathcal{L}}{\longrightarrow} \quad V_{IN}(s) = 1\,\mathrm{V} \cdot \frac{1}{s} \\
V_{OUT}(s) &= 1\,\mathrm{V} \cdot \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \cdot \frac{1}{s} = 1\,\mathrm{V} \cdot \frac{p_1 p_2}{(s - p_1)(s - p_2)} \cdot \frac{1}{s} \\
&\overset{\text{(PFE)}}{=} 1\,\mathrm{V} \cdot \left(\frac{1}{s} + \frac{\frac{p_2}{p_1 - p_2}}{s - p_1} - \frac{\frac{p_1}{p_1 - p_2}}{s - p_2} \right) \\
&\quad \Big\downarrow \mathcal{L}^{-1} \\
v_{OUT}(t) &= 1\,\mathrm{V} \cdot \mathrm{u}(t) \left(1 + \frac{p_2\,\mathrm{e}^{p_1 t} - p_1\,\mathrm{e}^{p_2 t}}{p_1 - p_2} \right)
\end{aligned}
$$

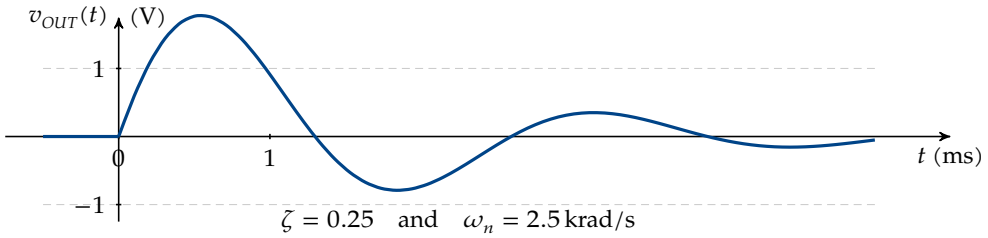
- $\zeta = 1$:



(a) $\zeta > 1$



(b) $\zeta = 1$



(c) $\zeta < 1$

Figure 7.3: Typical impulse responses of a stable second-order system (for three different values of ζ)

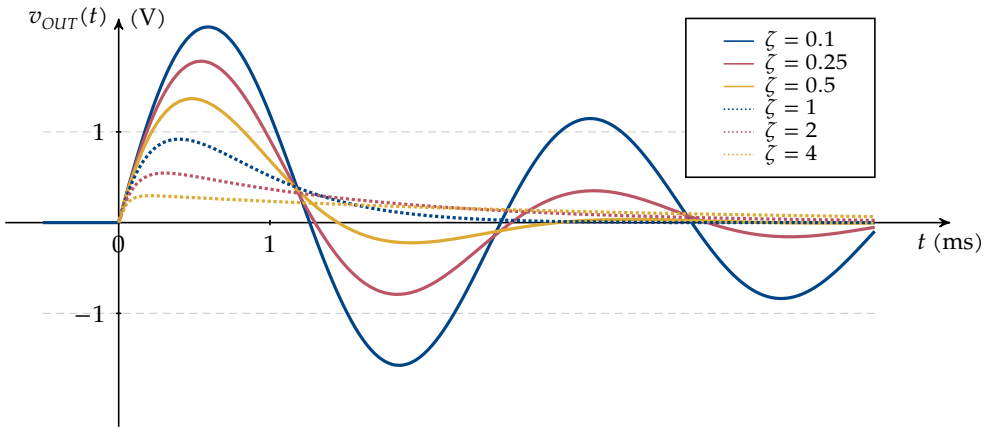


Figure 7.4: Impulse responses of a stable second-order system (for various values of ζ)

$$
\begin{aligned}
V_{OUT}(s) &= \frac{\omega_n^2}{s^2 + 2\omega_n s + \omega_n^2} V_{IN}(s) \\
&\Big\downarrow \; v_{IN}(t) = 1\,\mathrm{V} \cdot \mathrm{u}(t) \quad \xrightarrow{\mathcal{L}} \quad V_{IN}(s) = 1\,\mathrm{V} \cdot \frac{1}{s} \\
V_{OUT}(s) &= 1\,\mathrm{V} \cdot \frac{\omega_n^2}{s^2 + 2\omega_n s + \omega_n^2} \frac{1}{s} = 1\,\mathrm{V} \cdot \frac{\omega_n^2 \mathrm{V\,s}}{(s+\omega_n)^2} \frac{1}{s} \\
&\overset{\text{(PFE)}}{=} 1\,\mathrm{V} \left(\frac{1}{s} + \frac{-\omega_n}{(s+\omega_n)^2} + \frac{-1}{s+\omega_n} \right) \\
&\Big\downarrow \mathcal{L}^{-1} \\
v_{OUT}(t) &= 1\,\mathrm{V} \cdot \mathrm{u}(t) \left[1 - (1 + \omega_n t)\, \mathrm{e}^{-\omega_n t} \right]
\end{aligned}
$$

- $\zeta < 1$:

$$
\begin{aligned}
V_{OUT}(s) &= \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} V_{IN}(s) \\
&\Big\downarrow \; v_{IN}(t) = 1\,\mathrm{V} \cdot \mathrm{u}(t) \quad \xrightarrow{\mathcal{L}} \quad V_{IN}(s) = 1\,\mathrm{V} \cdot \frac{1}{s} \\
V_{OUT}(s) &= 1\,\mathrm{V} \cdot \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \frac{1}{s} \overset{\text{(PFE)}}{=} 1\,\mathrm{V} \left(\frac{1}{s} - \frac{s + 2\zeta\omega_n}{s^2 + 2\zeta\omega_n s + \omega_n^2} \right) \\
&= 1\,\mathrm{V} \cdot \left(\frac{1}{s} - \frac{s + \zeta\omega_n}{(s+\zeta\omega_n)^2 + \omega_d^2} - \frac{\zeta}{\sqrt{1-\zeta^2}} \frac{\omega_d}{(s+\zeta\omega_n)^2 + \omega_d^2} \right) \\
&\Big\downarrow \mathcal{L}^{-1} \\
v_{OUT}(t) &= 1\,\mathrm{V} \cdot \mathrm{u}(t) \left[1 - \left(\cos\omega_d t + \frac{\zeta}{\sqrt{1-\zeta^2}} \sin\omega_d t \right) \mathrm{e}^{-\zeta\omega_n t} \right]
\end{aligned}
$$

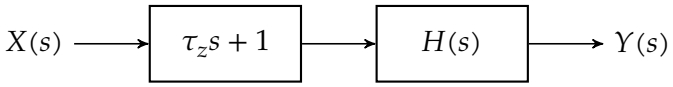
with $\omega_d = \omega_n \sqrt{1-\zeta^2}$.

An overall view of the effect of ζ on the step response, can be found in Figure 7.6.

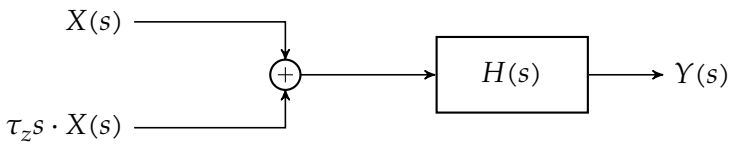
7.5 The effect of zeros

7.5.1 Description

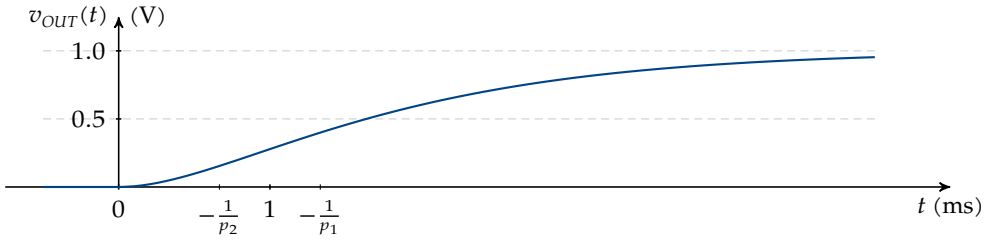
So far we only studied all-pole systems. How does a zero affect the way the system responds? Adding a zero to the systems above corresponds to cascading a block with a single zero in front (or after) the system:



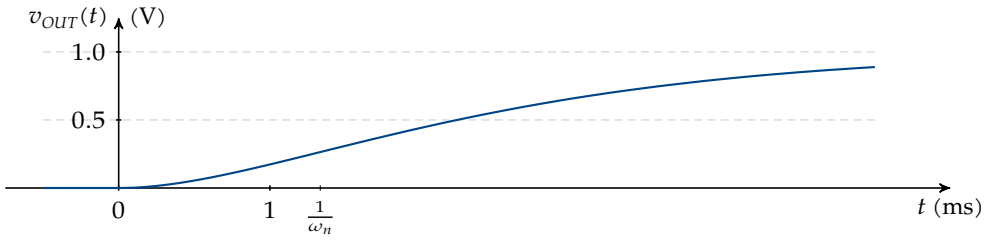
This is equivalent with:



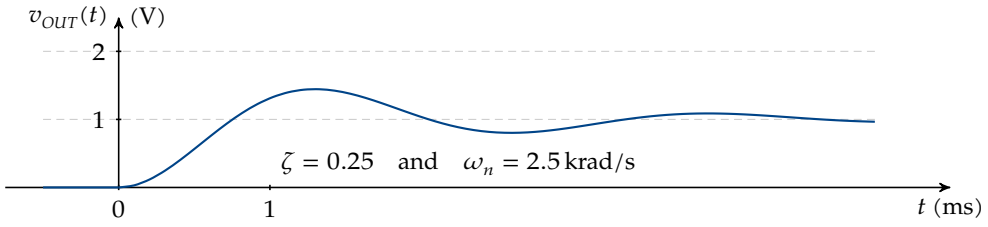
Now, let's assume that we limit ourselves to a step input, i.e. $X(s) = 1/s$. This results in:



(a) $\zeta > 1$



(b) $\zeta = 1$



(c) $\zeta < 1$

Figure 7.5: Typical step responses of a stable second-order system (for three different values of ζ); note that all these graphs have a horizontal asymptote for $t \to \infty$, i.e. $v_{OUT} = 1$

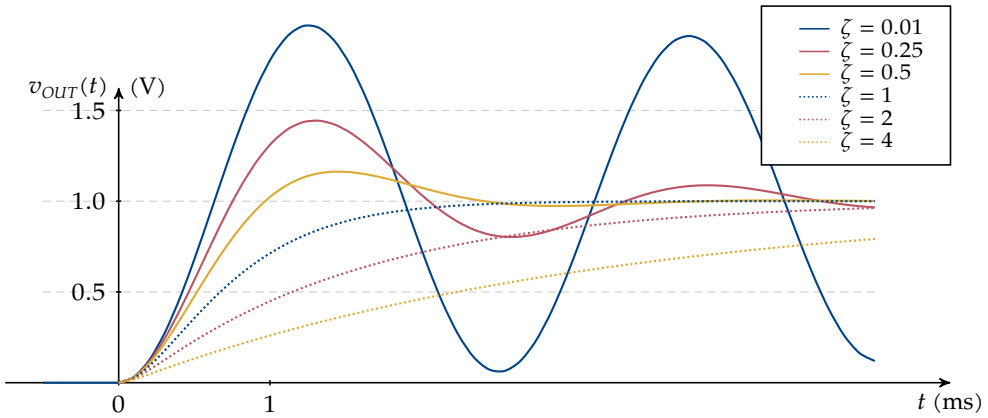
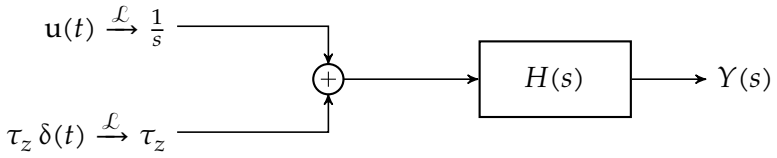


Figure 7.6: Step responses of a stable second-order system (for various values of ζ)

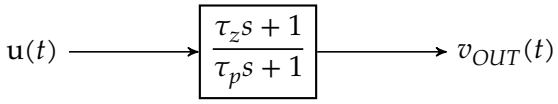


Conclusion: the step response of an LTI system with a zero, corresponds to a weighted sum of the step response $g(t)$ and the impulse response $h(t)$ of the system without the zero:

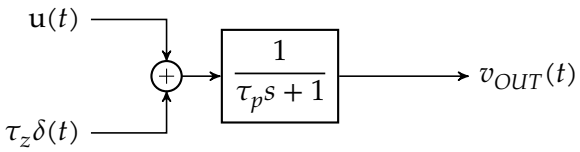
$$y(t) = g(t) + \tau_z h(t)$$

7.5.2 First-order systems with a zero

Consider as an example a first-order system with a zero:



This is equivalent with:



We calculated the impulse response $h(t)$ and the step response $g(t)$ of the first-order system without a zero before:

$$h(t) = \mathrm{u}(t) \cdot \frac{1}{\tau_p} \mathrm{e}^{-\frac{t}{\tau_p}}$$

$$g(t) = \mathrm{u}(t) \cdot \left(1 - \mathrm{e}^{-\frac{t}{\tau_p}}\right)$$

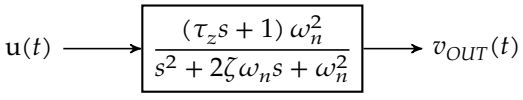
Therefore:

$$v_{OUT}(t) = \mathrm{u}(t) \left(1 - \mathrm{e}^{-t/\tau_p} + \frac{\tau_z}{\tau_p} \mathrm{e}^{-t/\tau_p}\right)$$

The location of the zero compared to the location of the pole, determines the impact of the zero. This has been illustrated in Figure 7.7. If the zero is dominant, a significant overshoot will occur.

7.5.3 Second-order systems with a zero

The case of a second-order system with a zero is no different. We only consider the case for which $\zeta < 1$ as the other cases are in fact multiple first-order systems. We will treat higher-order systems later.



This can be seen to be equivalent with:

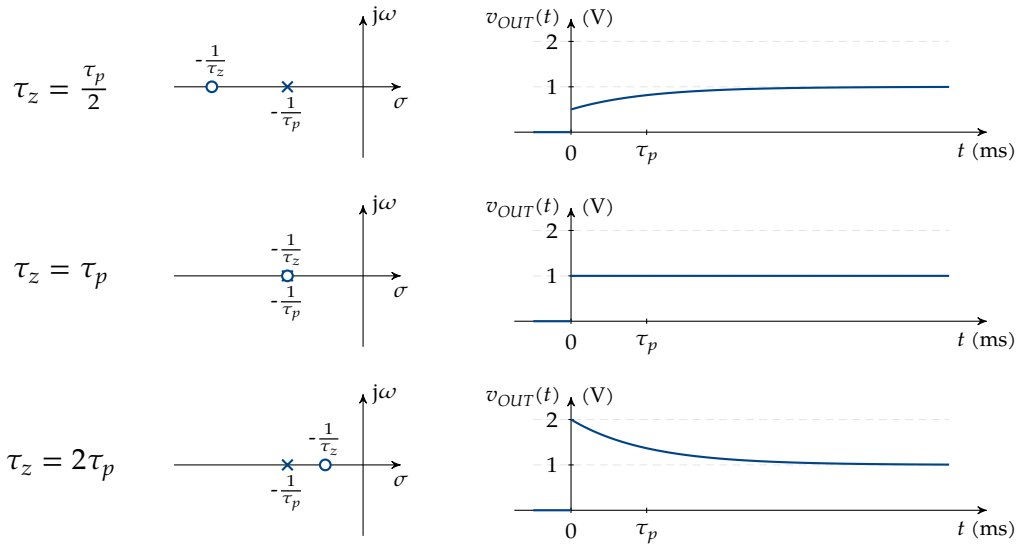
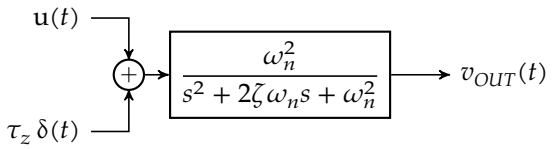


Figure 7.7: Impact of the location of a zero on the step input of a first-order system with a zero (nondominant zero at the top, a canceling zero in the middle and a dominant zero at the bottom)



We calculated the impulse response $h(t)$ and the step response $g(t)$ of the second-order system without a zero before:

$$h(t) = \mathrm{u}(t)\,\frac{\omega_n}{\sqrt{1-\zeta^2}}\,\mathrm{e}^{-\zeta\omega_n t}\sin\omega_d t$$

$$g(t) = \mathrm{u}(t)\;\mathrm{e}^{-\zeta\omega_n t}\left[1 - \left(\cos\omega_d t + \frac{\zeta}{\sqrt{1-\zeta^2}}\sin\omega_d t\right)\right]$$

Therefore:

$$v_{OUT}(t) = \mathrm{u}(t)\left[1 - \mathrm{e}^{-\zeta\omega_n t}\left(\cos\omega_d t + \frac{\zeta - \tau_z\omega_n}{\sqrt{1-\zeta^2}}\sin\omega_d t\right)\right]$$

Again, the location of the zero compared to the location of the poles, determines the impact of the zero. This has been illustrated in Figure 7.8. If the zero is dominant, a significant overshoot will occur.

7.5.4 Higher-order systems with zeros

Consider a higher-order transfer function of an LTI system with m zeros and n poles:

$$H(s) = \frac{b_m s^m + b_{m-1}s^{m-1} + \cdots + b_2 s^2 + b_1 s + b_0}{a_n s^n + a_{n-1}s^{n-1} + \cdots + a_3 s^3 + a_2 s^2 + a_1 s + a_0} = \frac{\sum_{j=0}^{m} b_j s^j}{\sum_{i=0}^{n} a_i s^i}$$

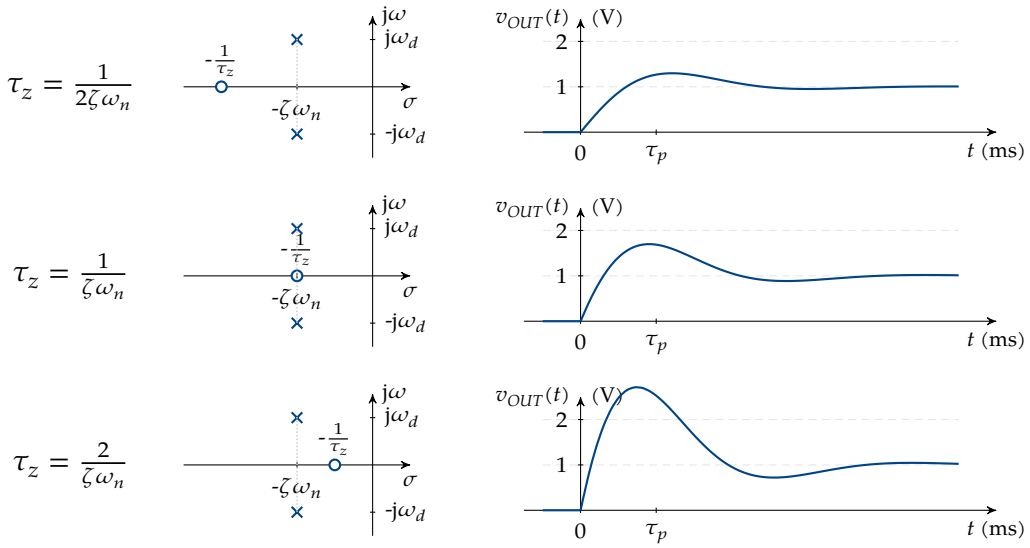


Figure 7.8: Impact of the location of a zero on the step input of a second-order system with a zero (nondominant zero at the top, a zero with the same real part as the poles in the middle and a dominant zero at the bottom)

We know that it is always possible to factorize such a transfer function into factors consisting of a single real pole/zero or pairs of complex conjugate poles and zeros:

$$H(s) = K \cdot \frac{\prod_{j=1}^{m_1}(s - z_j)}{\prod_{i=1}^{n_1}(s - p_i)} \cdot \frac{\prod_{j=1}^{m_2}(s - (\sigma_{z,j} + j\omega_{z,j}))(s - (\sigma_{z,j} - j\omega_{z,j}))}{\prod_{i=1}^{n_2}(s - (\sigma_{p,i} + j\omega_{p,i}))(s - (\sigma_{p,i} - j\omega_{p,i}))}$$

In practice this means that one can always see a higher-order system as cascade of first- and second-order systems with zeros.

If we consider the step response of such a system, we can make the following observation. If the system $H(s)$ is feasible, this means that $m \leq n$. So, for $Y(s) = H(s) \cdot \frac{1}{s}$, it must hold that the degree of the denominator of $Y(s)$ is strictly higher than the degree of the numerator.

Under these circumstances, we'd able to prove that we can always write it in the following form:

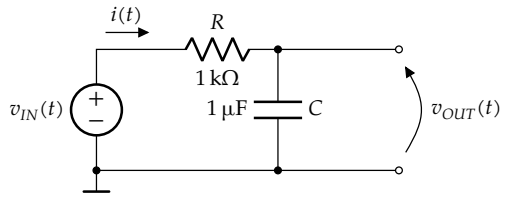
$$y(t) = \mathrm{u}(t)\left(K + \sum_{i=1}^{n_1} A_i\,\mathrm{e}^{p_i t} + \sum_{i=1}^{n_2} \mathrm{e}^{\sigma_{p,i}t}\left(B_j \cos\omega_{d,i}t + C_j \sin\omega_{d,i}t\right)\right)$$

The factor K is the gain of the overall system. The damping factor p_i and $\sigma_{p,i}$, and the damped eigenfrequencies $\omega_{d,i}$ are deteremind by the location of the poles. The coefficients A_i, B_j and C_j are determined by the location of the zeros.

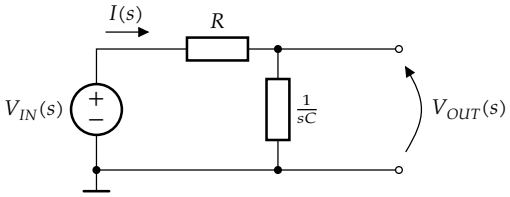
7.6 Examples

7.6.1 An RC filter

The RC filter is the posterchild example of a first-order system. Consider the following network:



Let's try to calculate the output $v_{OUT}(t)$ for a number of input signals $v_{IN}(t)$. However, let's first calculate the transfer function. We start by converting the model to the Laplace domain:



For this circuit it is easy to write the output voltage:

$$V_{OUT}(s) = \frac{\frac{1}{sC}}{R + \frac{1}{sC}} V_{IN}(s) = \frac{1}{RCs + 1} V_{IN}(s)$$

and the output current:

$$I(s) = \frac{1}{R + \frac{1}{sC}} V_{IN}(s) = \frac{Cs}{RCs + 1} V_{IN}(s)$$

Note that in all calculations, the best strategy is to work fully symbolic and to only substitute the symbols with real values at the very end, when making a graph.

A Dirac impulse

Let's calculate the impulse response for the output voltage:

$$\begin{aligned}
V_{OUT}(s) &= \frac{1}{RCs + 1} V_{IN}(s) \\
&\quad \downarrow \; v_{IN}(t) = 1\,\mathrm{V} \cdot \delta(t) \quad \xrightarrow{\mathcal{L}} \quad V_{IN}(s) = 1\,\mathrm{V\,s} \\
V_{OUT}(s) &= \frac{1}{RCs + 1} \cdot 1\,\mathrm{V\,s} = \frac{1\,\mathrm{V\,s}}{\tau} \frac{1}{s + 1/\tau} \qquad \text{with } \tau = RC = 1 \times 10^{-3}\,\mathrm{s} \\
&\quad \downarrow \; \mathcal{L}^{-1} \\
v_{OUT}(t) &= \frac{1\,\mathrm{V\,s}}{\tau}\, \mathrm{u}(t)\, \mathrm{e}^{-t/\tau}
\end{aligned}$$

Let's also calculate the impulse response for the output current:

$$
\begin{aligned}
I(s) &= \frac{Cs}{RCs+1} V_{IN}(s) \\
&\quad \Big\downarrow v_{IN}(t) = 1\,\mathrm{V} \cdot \delta(t) \quad \xrightarrow{\mathcal{L}} \quad V_{IN}(s) = 1\,\mathrm{V\,s} \\
I(s) &= \frac{Cs}{RCs+1} \cdot 1\,\mathrm{V\,s} \\
&= \frac{1\,\mathrm{V\,s}}{R} \frac{RCs+1-1}{RCs+1} = \frac{1\,\mathrm{V\,s}}{R}\left(1 - \frac{1}{RCs+1}\right) = \frac{1\,\mathrm{V\,s}}{R}\left(1 - \frac{\frac{1}{\tau}}{s+\frac{1}{\tau}}\right) \quad \text{with } \tau = RC \\
&\quad \Big\downarrow \mathcal{L}^{-1} \\
i(t) &= \frac{1\,\mathrm{V}}{R}\,\delta(t) - \frac{1\,\mathrm{V\,s}}{R^2 C}\,\mathrm{u}(t)\ \mathrm{e}^{-t/\tau} \\
&= 1\,\mathrm{mA}\ \delta(t) - 1\,\mathrm{A}\ \mathrm{u}(t)\ \mathrm{e}^{-t/\tau}
\end{aligned}
$$

You can find these responses in Figure 7.9a.

A unit step

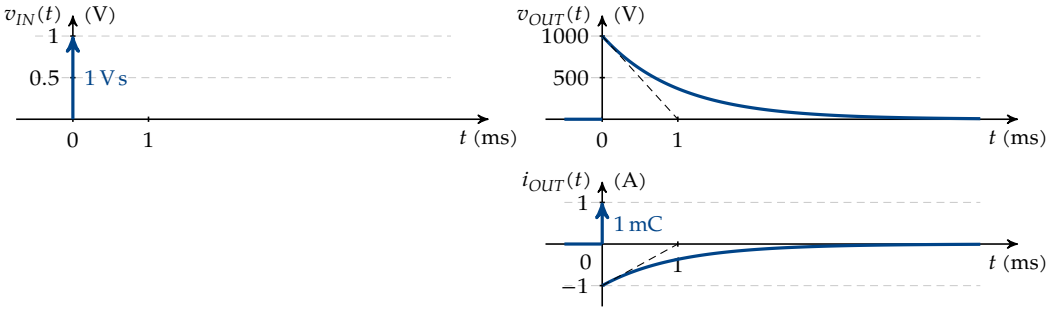
The step response of the output voltage can be easily calculated as follows:

$$
\begin{aligned}
V_{OUT}(s) &= \frac{1}{RCs+1} V_{IN}(s) \\
&\quad \Big\downarrow v_{IN}(t) = 1\,\mathrm{V} \cdot \mathrm{u}(t) \quad \xrightarrow{\mathcal{L}} \quad V_{IN}(s) = \frac{1\,\mathrm{V}}{s} \\
V_{OUT}(s) &= \frac{1}{RCs+1} \cdot \frac{1\,\mathrm{V}}{s} = \frac{1/\tau}{s+1/\tau} \cdot \frac{1\,\mathrm{V}}{s} \overset{\text{(PFE)}}{=} 1\,\mathrm{V}\left(\frac{1}{s} - \frac{1}{s+1/\tau}\right) \quad \text{with } \tau = RC \\
&\quad \Big\downarrow \mathcal{L}^{-1} \\
v_{OUT}(t) &= 1\,\mathrm{V}\ \mathrm{u}(t)\ \left(1 - \mathrm{e}^{-t/\tau}\right)
\end{aligned}
$$

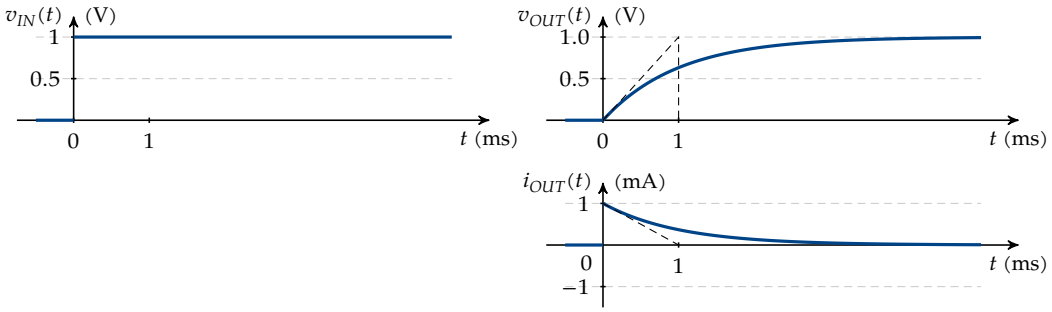
Let's do the same for the output current:

$$
\begin{aligned}
I(s) &= \frac{Cs}{RCs+1} V_{IN}(s) \\
&\quad \Big\downarrow v_{IN}(t) = 1\,\mathrm{V} \cdot \mathrm{u}(t) \quad \xrightarrow{\mathcal{L}} \quad V_{IN}(s) = \frac{1\,\mathrm{V}}{s} \\
I(s) &= \frac{Cs}{RCs+1} \cdot \frac{1\,\mathrm{V}}{s} = 1\,\mathrm{V}\frac{1}{R}\frac{1}{s+1/\tau} \quad \text{with } \tau = RC \\
&\quad \Big\downarrow \mathcal{L}^{-1} \\
i(t) &= 1\,\mathrm{mA}\ \mathrm{u}(t)\ \mathrm{e}^{-t/\tau}
\end{aligned}
$$

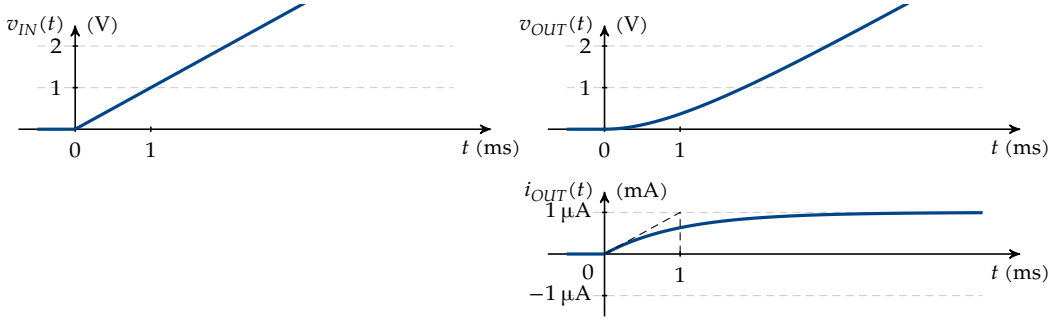
You can find this response in Figure 7.9b.



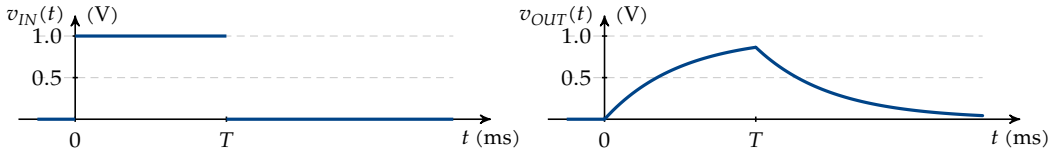
(a) Impulse response



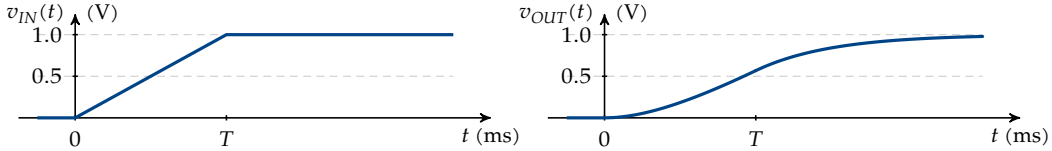
(b) Step response



(c) Unit ramp response



(d) Square pulse response



(e) Real step response

Figure 7.9: Response of the output voltage of an RC filter: input left, output(s) right

A unit ramp

The unit ramp response of the output voltage can be calculated as follows:

$$
\begin{aligned}
V_{OUT}(s) &= \frac{1}{RCs+1} V_{IN}(s) \\
&\quad \Big| \; v_{IN}(t) = 1\,\mathrm{V/s} \cdot \mathrm{u}(t) \cdot t \quad \xrightarrow{\mathcal{L}} \quad V_{IN}(s) = \frac{1\,\mathrm{V/s}}{s^2} \\
V_{OUT}(s) &= \frac{1}{RCs+1} \cdot \frac{1\,\mathrm{V/s}}{s^2} = \frac{1/\tau}{s+1/\tau} \cdot \frac{1\,\mathrm{V/s}}{s^2} \overset{\text{(PFE)}}{=} 1\,\mathrm{V/s}\left(\frac{1}{s^2} - \frac{\tau}{s} + \frac{\tau}{s+1/\tau}\right) \quad \text{with } \tau = RC \\
&\quad \downarrow \mathcal{L}^{-1} \\
v_{OUT}(t) &= 1\,\mathrm{V/s}\left(\mathrm{u}(t)\; t + \tau\left(\mathrm{e}^{-t/\tau} - 1\right)\right)
\end{aligned}
$$

The current response can be calculated as follows:

$$
\begin{aligned}
I(s) &= \frac{Cs}{RCs+1} V_{IN}(s) \\
&\quad \Big| \; v_{IN}(t) = 1\,\mathrm{V/s} \cdot \mathrm{u}(t) \cdot t \quad \xrightarrow{\mathcal{L}} \quad V_{IN}(s) = \frac{1\,\mathrm{V/s}}{s^2} \\
I(s) &= \frac{Cs}{RCs+1} \cdot \frac{1\,\mathrm{V/s}}{s^2} = 1\,\mathrm{V/s}\frac{1}{R}\frac{1}{s+1/\tau} \cdot \frac{1}{s} \quad \text{with } \tau = RC \\
&\quad \downarrow \mathcal{L}^{-1} \\
i(t) &= 1\,\mathrm{V/s}\; \mathrm{u}(t)\; C\left(1 - \mathrm{e}^{-t/\tau}\right)
\end{aligned}
$$

You can find this response in Figure 7.9c.

A square pulse

The square pulse response can be calculated as follows:

$$
\begin{aligned}
V_{OUT}(s) &= \frac{1}{RCs+1} V_{IN}(s) \\
&\quad \Big| \; v_{IN}(t) = 1\,\mathrm{V}(\mathrm{u}(t) - \mathrm{u}(t-T)) \quad \xrightarrow{\mathcal{L}} \quad V_{IN}(s) = 1\,\mathrm{V}\left(\frac{1}{s} - \frac{1}{s}\mathrm{e}^{-sT}\right) \\
V_{OUT}(s) &= \frac{1}{RCs+1} \cdot 1\,\mathrm{V}\left(\frac{1}{s} - \frac{1}{s}\mathrm{e}^{-sT}\right) = \frac{1/\tau}{s+1/\tau} \cdot 1\,\mathrm{V}\left(\frac{1}{s} - \frac{1}{s}\mathrm{e}^{-sT}\right) \quad \text{with } \tau = RC \\
&\quad \downarrow \mathcal{L}^{-1} \\
v_{OUT}(t) &= 1\,\mathrm{V}\left[\mathrm{u}(t)\; \left(1 - \mathrm{e}^{-t/\tau}\right) - \mathrm{u}(t-T)\left(1 - \mathrm{e}^{-(t-T)/\tau}\right)\right]
\end{aligned}
$$

You can find this response in Figure 7.9d.

A real step

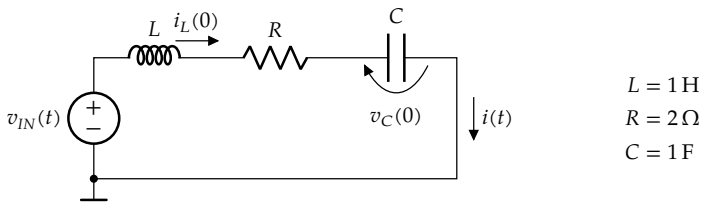
Finally, a real step with a finite slope:

$$
\begin{aligned}
V_{OUT}(s) &= \frac{1}{RCs+1} V_{IN}(s) \\
&\quad \Big| \; v_{IN}(t) = \frac{1\,\mathrm{V}}{T}\left(\mathrm{u}(t) \cdot t - \mathrm{u}(t-T) \cdot (t-T)\right) \quad \xrightarrow{\mathcal{L}} \quad V_{IN}(s) = \frac{1\,\mathrm{V}}{T}\left(\frac{1}{s^2} - \frac{1}{s^2}\mathrm{e}^{-sT}\right) \\
V_{OUT}(s) &= \frac{1}{RCs+1} \cdot \frac{1\,\mathrm{V}}{T}\left(\frac{1}{s^2} - \frac{1}{s^2}\mathrm{e}^{-sT}\right) = \frac{1/\tau}{s+1/\tau} \cdot \frac{1\,\mathrm{V}}{T}\left(\frac{1}{s^2} - \frac{1}{s^2}\mathrm{e}^{-sT}\right) \quad \text{with } \tau = RC \\
&\quad \downarrow \mathcal{L}^{-1} \\
v_{OUT}(t) &= \frac{1\,\mathrm{V}}{T}\left[\mathrm{u}(t)\left(t + \tau\left(\mathrm{e}^{-t/\tau} - 1\right)\right) - \mathrm{u}(t-T)\left((t-T) + \tau\left(\mathrm{e}^{-\frac{t-T}{\tau}} - 1\right)\right)\right]
\end{aligned}
$$

You can find this response in Figure 7.9e.

7.6.2 An RLC circuit

Consider the RLC circuit below:



Assume the following initial conditions:

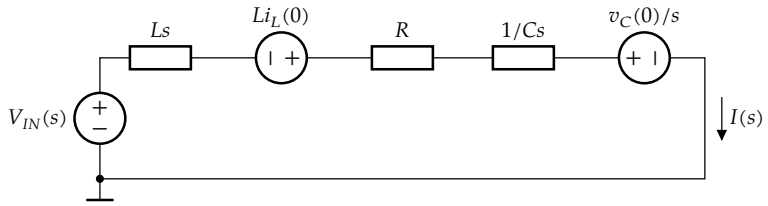
$$i_L(0) = 1\,\mathrm{A} \qquad\qquad v_C(0) = 1\,\mathrm{V}$$

and the following input signal:

$$v_{IN}(t) = 8\,\mathrm{u}(t)\ \mathrm{e}^{t} \quad \xrightarrow{\mathcal{L}} \quad V_{IN}(s) = \frac{8}{s-1}$$

Let's analyze the current $i(t)$.

The first step is to convert the entire schematic to the Laplace domain, taking into account the initial conditions:

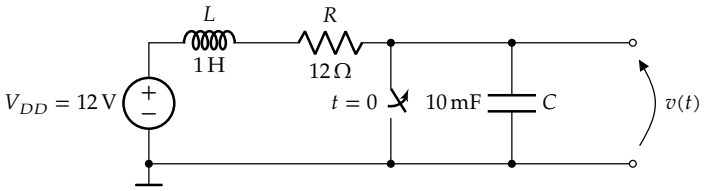


We've chosen for the series equivalents, as this keeps the calculations simple in this case. By simple inspection of the circuit, we can write:

$$\begin{aligned}
I(s) &= \frac{V_{IN}(s) + Li_L(0) - \frac{v_C(0)}{s}}{Ls + R + \frac{1}{Cs}} \\
&= \frac{\frac{8}{s-1} + 1 - \frac{1}{s}}{s + 2 + \frac{1}{s}} \\
&= \frac{8s + s(s-1) - (s-1)}{(s^2 + 2s + 1)(s-1)} = \frac{s^2 + 6s + 1}{(s+1)^2(s-1)} \\
&\overset{\text{(PFE)}}{=} \frac{2}{(s+1)^2} + \frac{-1}{s+1} + \frac{2}{s-1} \\
&\downarrow \mathcal{L}^{-1} \\
i(t) &= \mathrm{u}(t)\left(2t\,\mathrm{e}^{-t} - \mathrm{e}^{-t} + 2\,\mathrm{e}^{t}\right)
\end{aligned}$$

7.6.3 Calculating switching transients in a spark plug voltage generator

Consider the following circuit:



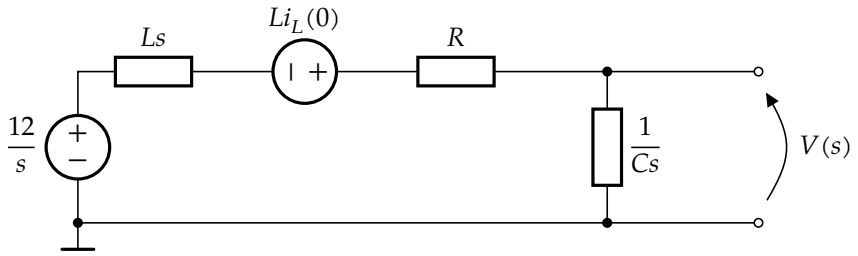
Calculate $v(t)$ for $t \geq 0$ when the switch opens at $t = 0$.

We will treat this problem in two phases: (1) $t < 0$, and (2) $t \geq 0$.
The first phase is focused on finding the initial conditions that will apply in the second phase. As the first phase lasted 'forever', we may assume that currents and voltages have become constants and therefore, inductors have become a short circuit and capacitors an open circuit.

Phase 1: $t < 0$
Obviously the switch short-circuited the capacitor, therefore $v_C(0) = 0\,\text{V}$.
As the inductor is a short circuit in the first phase, we may assume that it conducts at the end of the phase a current of

$$i_L(0) = \frac{12\,\text{V}}{12\,\Omega} = 1\,\text{A}$$

Phase 2: $t \geq 0$
Knowing the initial conditions, we can now transform the circuit to the Laplace domain:



Solving this network for $V(s)$ yields:

$$\begin{aligned} V(s) &= \left(V_{DD}\frac{1}{s} + Li_L(0) \right) \frac{1/Cs}{Ls + R + 1/Cs} \\ &= \left(V_{DD}\frac{1}{s} + Li_L(0) \right) \frac{1/LC}{s^2 + \frac{R}{L}s + 1/LC} \end{aligned}$$

Note that we have written the second-order factor in the expression above in its normal form. By equating coefficients, we can determine the classical second-order parameters:

$$\frac{1/LC}{s^2 + \frac{R}{L}s + 1/LC} \quad \Leftrightarrow \quad \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$$

Therefore:

$$\omega_n = \frac{1}{\sqrt{LC}} = \frac{1}{\sqrt{1\,\mathrm{H}\cdot 10\,\mathrm{mF}}} = 10\,\mathrm{rad/s}$$
$$\zeta = \frac{R}{2}\sqrt{\frac{C}{L}} = \frac{12\,\Omega}{2}\sqrt{\frac{10\,\mathrm{mF}}{1\,\mathrm{H}}} = 0.6$$

and a parameter we will need later: $\omega_d = \omega_n\sqrt{1-\zeta^2} = 8\,\mathrm{rad/s}$.

Let's continue the hard work:

$$\begin{aligned}
V(s) &= \left(V_{DD}\frac{1}{s} + Li_L(0)\right)\frac{1/LC}{s^2 + \frac{R}{L}s + 1/LC} = 1\,\mathrm{V}\left(\frac{12}{s}+1\right)\frac{100}{s^2+12s+100}\\
&= 1\,\mathrm{V}\frac{100s+1200}{s(s^2+12s+100)}\\
&\overset{\text{(PFE)}}{=} 1\,\mathrm{V}\left(\frac{12}{s} - \frac{12s+44}{s^2+12s+100}\right)\\
&\downarrow d = b^2-4ac = 12^2 - 4\cdot 100 < 0\\
&= 1\,\mathrm{V}\left(\frac{12}{s} - \frac{12s+44}{s^2+12s+36-36+100}\right)\\
&= 1\,\mathrm{V}\left(\frac{12}{s} - \frac{12s+44}{(s+6)^2+64}\right)
\end{aligned}$$

The table of common transform pairs only contain entries like:

$$\frac{s+a}{(s+a)^2+\omega^2} \qquad \text{and} \qquad \frac{\omega}{(s+a)^2+\omega^2}$$

We therefore try to rework our last line to these forms:

$$\begin{aligned}
V(s) &= 1\,\mathrm{V}\left[\frac{12}{s} - \left(12\frac{s+6}{(s+6)^2+8^2} + \frac{44-72}{(s+6)^2+8^2}\right)\right]\\
&= 1\,\mathrm{V}\left[\frac{12}{s} - \left(12\frac{s+6}{(s+6)^2+8^2} - \frac{28}{8}\frac{8}{(s+6)^2+8^2}\right)\right]\\
&\downarrow \mathcal{L}^{-1}\\
v(t) &= 1\,\mathrm{V}\,\mathrm{u}(t)\left[12 - \left(12\cos 8t - \frac{7}{2}\sin 8t\right)\mathrm{e}^{-6t}\right]
\end{aligned}$$

Note that in this case, we have used the numerical values quite soon in the process. The more hardened reader, might try the full symbolic math and will obtain:

$$V(s) = \frac{V_{DD}}{s} - \left(V_{DD}\frac{s+\zeta\omega_n}{(s+\zeta\omega_n)^2+\omega_d^2} + \frac{Li_L(0)\omega_n^2 - \zeta\omega_n V_{DD}}{\omega_d}\frac{\omega_d}{(s+\zeta\omega_n)^2+\omega_d^2}\right)$$

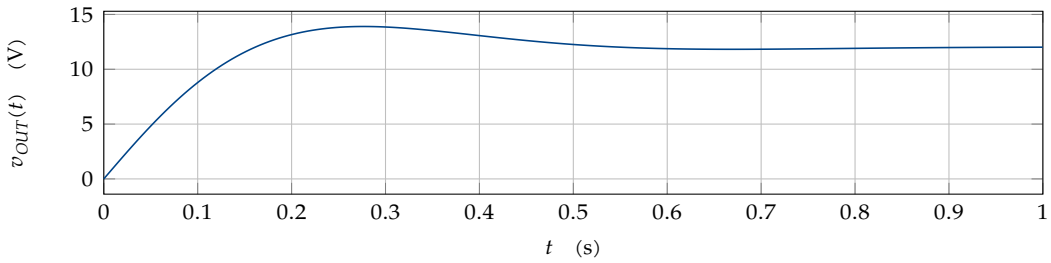
which can be inversely transformed to the time domain as:

$$v(t) = \mathrm{u}(t)\left[V_{DD} - \mathrm{e}^{-\zeta\omega_n t}\left(V_{DD}\cos\omega_d t + \frac{Li_L(0)\omega_n^2 - \zeta\omega_n V_{DD}}{\omega_d}\sin\omega_d t\right)\right] \tag{7.2}$$

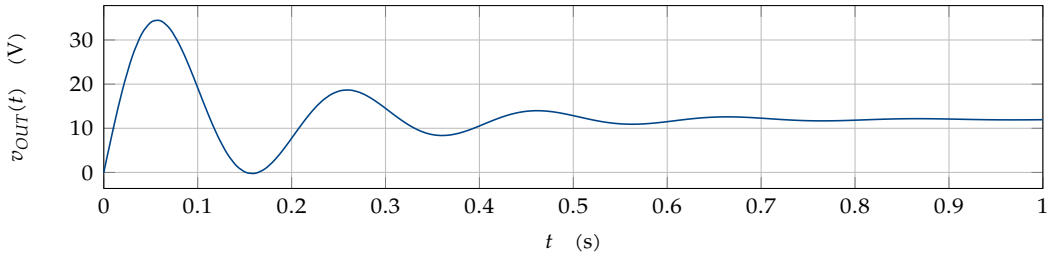
with:

$$\omega_n = \frac{1}{\sqrt{LC}} \qquad \zeta = \frac{R}{2}\sqrt{\frac{C}{L}} \qquad \omega_d = \omega_n\sqrt{1-\zeta^2}$$

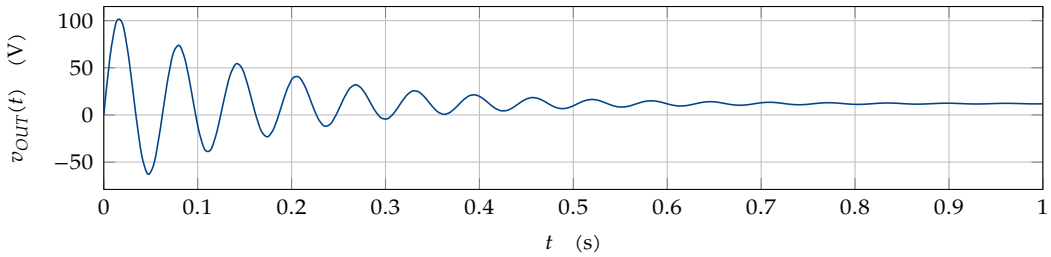
Let's graph this switching transient for multiple values of the capacitor. We did so for the values 10 mF down to 1 µF in steps of 10. The results can be found in Figure 7.10.



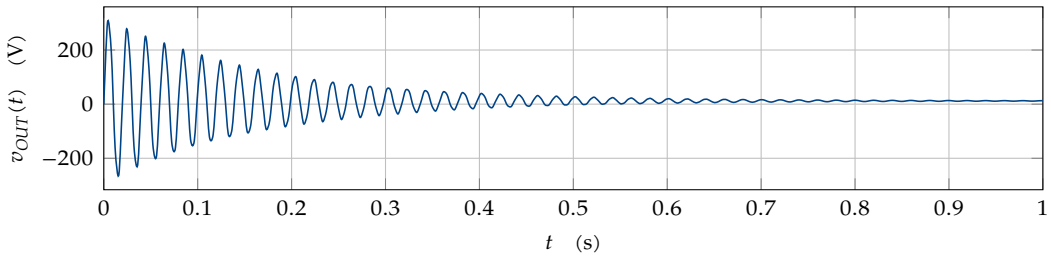
(a) $C = 10\,\text{mF}$



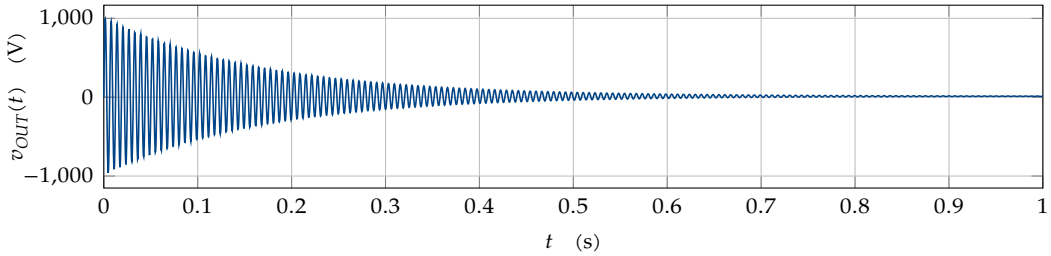
(b) $C = 1\,\text{mF}$



(c) $C = 100\,\mu\text{F}$



(d) $C = 10\,\mu\text{F}$



(e) $C = 1\,\mu\text{F}$

Figure 7.10: Transient phenomenon of the spark plug high voltage generator for multiple values of the capacitor

Remarks

- The circuit is a resonance LC-tank, and it constantly converts back and forth electromagnetic energy (in the inductor = $Li^2/2$) into electrostatic energy (in the capacitor = $Cv^2/2$).
- The smaller the capacitor the higher the peak voltage, and the higher the eigenfrequency.
- The damping time constant is not a function of the size of the capacitor, therefore you see the same decay in all the graphs.
- An approximate value of the maximal peak equals

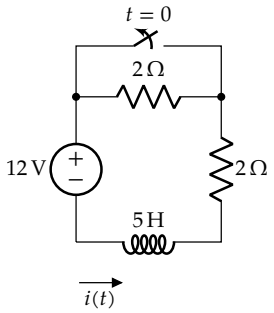
$$v_{\max} = i_L(0)\sqrt{\frac{L}{C}}$$

- This circuit is used to generate high voltages for the ignition of the spark plugs in a car.

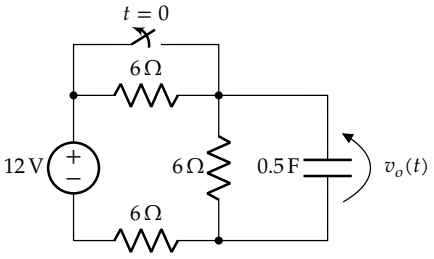
Final question: take a look at (7.2) again, and try to predict what happens if you increase the value of L!

Exercises

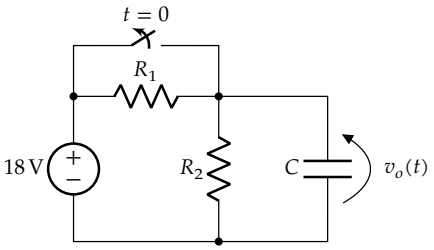
Exercise 7.6.3-1:
Consider the circuit below. Determine $i(t)$ for $t > 0$.



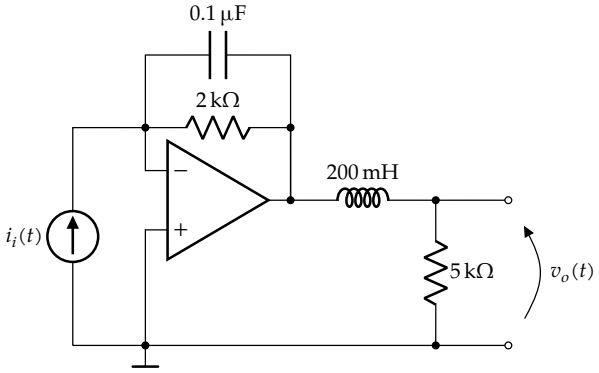
Exercise 7.6.3-2:
Consider the circuit below. Determine $v_o(t)$ for $t > 0$.



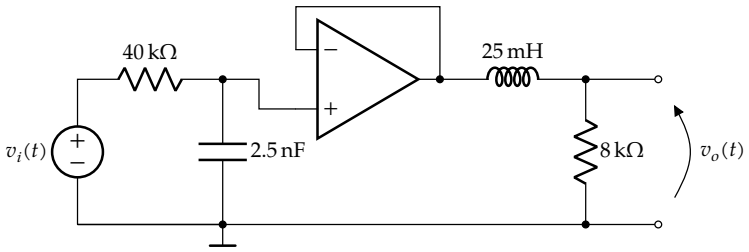
Exercise 7.6.3-3:
In the circuit below, the voltage across the capacitor is given by $v_o(t) = 6 + 12\,e^{-2t}$ when $t > 0$. Determine the value of the capacitance C and the value of the resistance R_1, knowing that $R_2 = 3\,\Omega$.



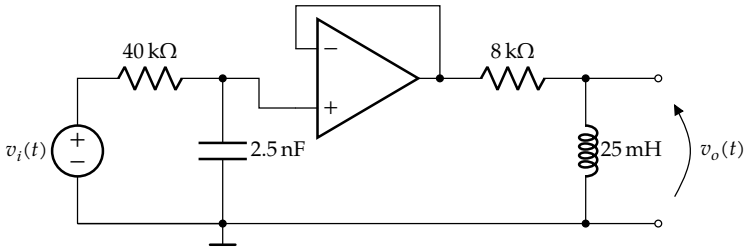
Exercise 7.6.3-4:
Determine the impulse response of the circuit below with $i_i(t)$ as input and $v_o(t)$ as output.



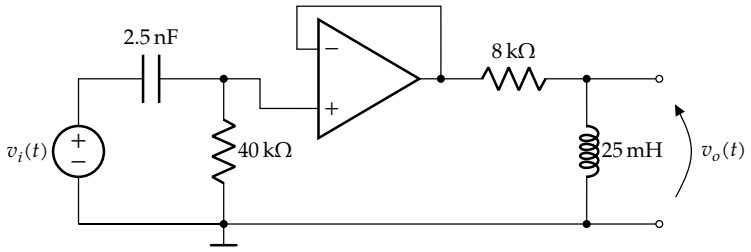
Exercise 7.6.3-5:
Determine the impulse response of the circuit below with $v_i(t)$ as input and $v_o(t)$ as output.



Exercise 7.6.3-6:
Determine the impulse response of the circuit below with $v_i(t)$ as input and $v_o(t)$ as output.



Exercise 7.6.3-7:
Determine the impulse response of the circuit below with $v_i(t)$ as input and $v_o(t)$ as output.



7.7 Conclusion

In this chapter, we have studied first-order and second-order systems in the time-domain. Knowing these basic systems, we can have at least some insight in the time-domain behavior of real-world (higher-order) systems.

Besides this basic knowledge, we have also seen how we can use the Laplace transform as one of the blades of our Swiss army knife to solve any time-domain question about LTI-systems. Make sure you master the trade! Or the blade?

Chapter 8

Analyzing LTI systems — frequency domain

In this chapter, you will learn about:

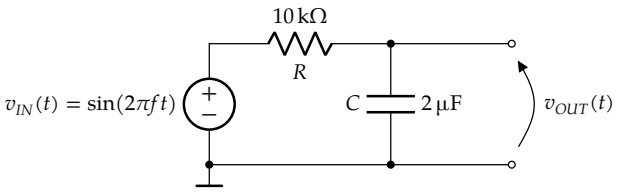
- transfer functions, how we can derive them analytically and graphically,
- how we can draw them as asymptotic Bode diagrams.

After having read/studied this chapter, you are expected to be able to

- derive transfer functions of electronic circuits, and
- draw asymptotic Bode diagrams for any possible LTI transfer function.

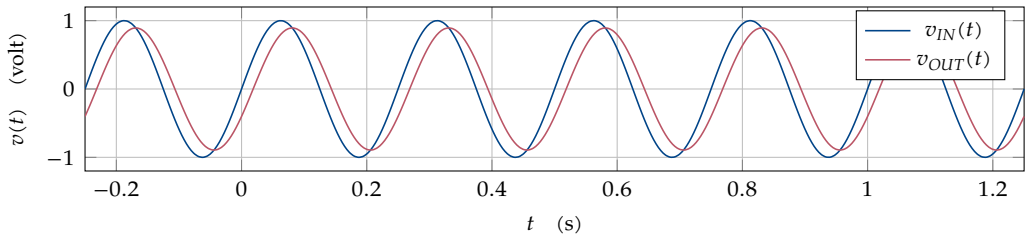
8.1 Introduction

Consider the following test case, a simple RC filter. Our goal is to learn more about its frequency behavior.

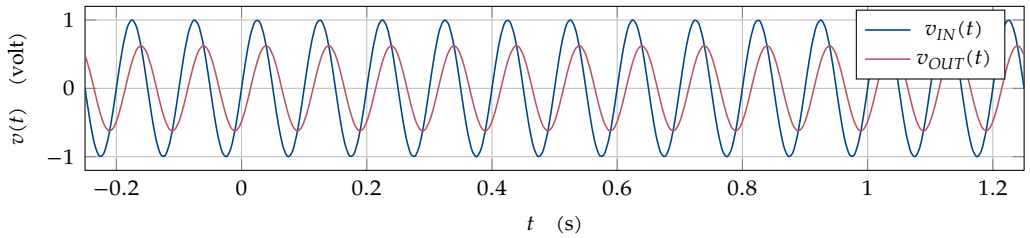


We know that sine waves are eigenfunctions of linear systems, so let's study this sine-wave behavior. To this end, we will excite this filter using a sine-wave generator and measure the output using an oscilloscope. Remember, the oscilloscope makes a measurement in time. You can find the result for multiple frequencies f in Figure 8.1. We can clearly see that the frequency of the output is the same as the input frequency (we expect that from linear systems), the amplitude of the output is decaying for increasing frequency and the phase lag is also increasing with increasing frequency. It's the combo of amplitude and phase delay that we want to study as a function of frequency. However, we don't want to be restricted to measuring, we also want to calculate its behavior.

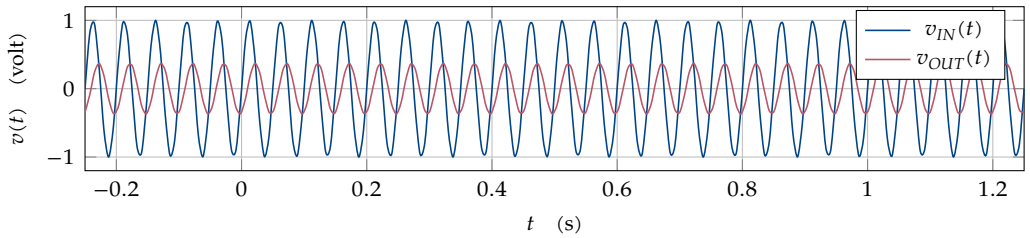
Therefore, we will study LTI systems in the frequency domain in this chapter. We will use the Laplace transform once more as our go-to Swiss army knife blade, as we can use it as a convenient way of performing a Fourier transform, required to draw frequency-domain graphs



(a) $f = 4\,\text{Hz}$



(b) $f = 10\,\text{Hz}$



(c) $f = 20\,\text{Hz}$

Figure 8.1: Input and output waveforms of the RC filter when exciting it with $\sin(2\pi f t)$ for various frequencies f

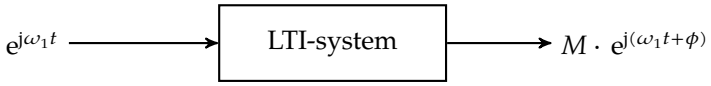
(spectra). To this end, we assume our problem to be a fully linear one (so no initial conditions), such that we can use the Laplace transform as a proxy for the Fourier transform.

We will also address the art of drawing spectra as Bode plots. This turns out to be a lot like playing with LEGO bricks. If you know the Bode plot of small dedicated blocks (of LTI systems like integrators, differentiators, first-order and second-order systems), you can use these smaller blocks to draw the spectrum (Bode plot) of any LTI-system you can imagine.

We will finish the chapter with a number of examples that illustrate the technique that we have learnt throughout this chapter.

8.2 Composing frequency transfer functions

Remember that sinors are eigenfunctions of LTI systems:



Our goal is to find the gain and the delay a sinor experiences on its way throug the system.

$$|H| = M$$
$$\angle H = \phi$$

Note that we use the typical engineering notation $\angle H$ to denote $\arg H$.
This relationship is in its essence the *frequency transfer function* of the system:

$$H(\omega) = \frac{Y(\omega)}{X(\omega)}$$

How can we calculate the frequency transfer function?

8.2.1 Calculating the frequency transfer function

Directly

We can calculate the frequency transfer function of a system with input x and output y in a very straightforward manner by calculating the Fourier transform of the impulse response. This assumes we calculate the impulse response first:

- Compose the system's differential equations
- Solve these equations for $x(t) = \delta(t)$ to obtain $y(t) = h(t)$
- Calculate $h(t) \xrightarrow{\mathcal{F}} H(\omega)$

Though this is a respectable route, it is in most cases a long one and therefore not the one we prefer. We rather calculate the frequency transfer function indirectly.

Indirectly

This assumes we execute the following procedure:

- Compose the system's equations in the Fourier domain
- Solve for $H(\omega)$

This will be the way we prefer. However, there is one more peculiarity and that is (but keep it quiet): we prefer Laplace over Fourier.

Using the Laplace transform as our proxy

The frequency transfer function is the Fourier transform of the system's impulse response. Given the fact that for almost all impulse responses that we need in practice, the Laplace transform will also converge, we can use the Laplace transform as our 'proxy' to calculate the Fourier transform:

$$H(\omega) = H(s)|_{s=j\omega}$$

Why do we take that detour? Three reasons:

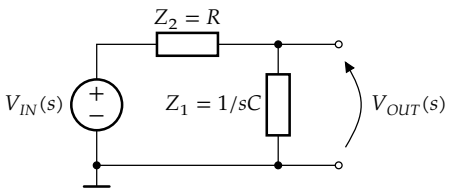
- because we are much more familiar with the Laplace transform than with the Fourier transform, and
- because writing s is shorter than writing $j\omega$ (it also looks less complex).
- because in the complex Laplace domain, we can have higher-level insights that aren't possible in the Fourier domain.

Composing the system equations in the Laplace/Fourier domain can be done for every scientific domain that is governed by linear laws. It is based on transforming all composing blocks or laws to the Fourier/Laplace domain. We already have seen how it works for electronics:

- Kirchhoff's laws translate into exactly the same form.
- The branch equations of the linear elements contain derivatives and therefore need some work to be transformed.

The result of this transformation process has been summarized in Figure 8.2.

Let's get back to our RC filter example and apply the technique. First we replace every network element by its Laplace equivalent. This leads to:



We recognize a potentiometric divider, and therefore can write by simple inspection:

$$H(s) \equiv \frac{V_{OUT}(s)}{V_{IN}(s)} = \frac{1/Cs}{1/Cs + R} = \frac{1}{1 + RCs}$$

$$\downarrow s = j\omega$$

$$H(j\omega) = \frac{1}{1 + jRC\omega}$$

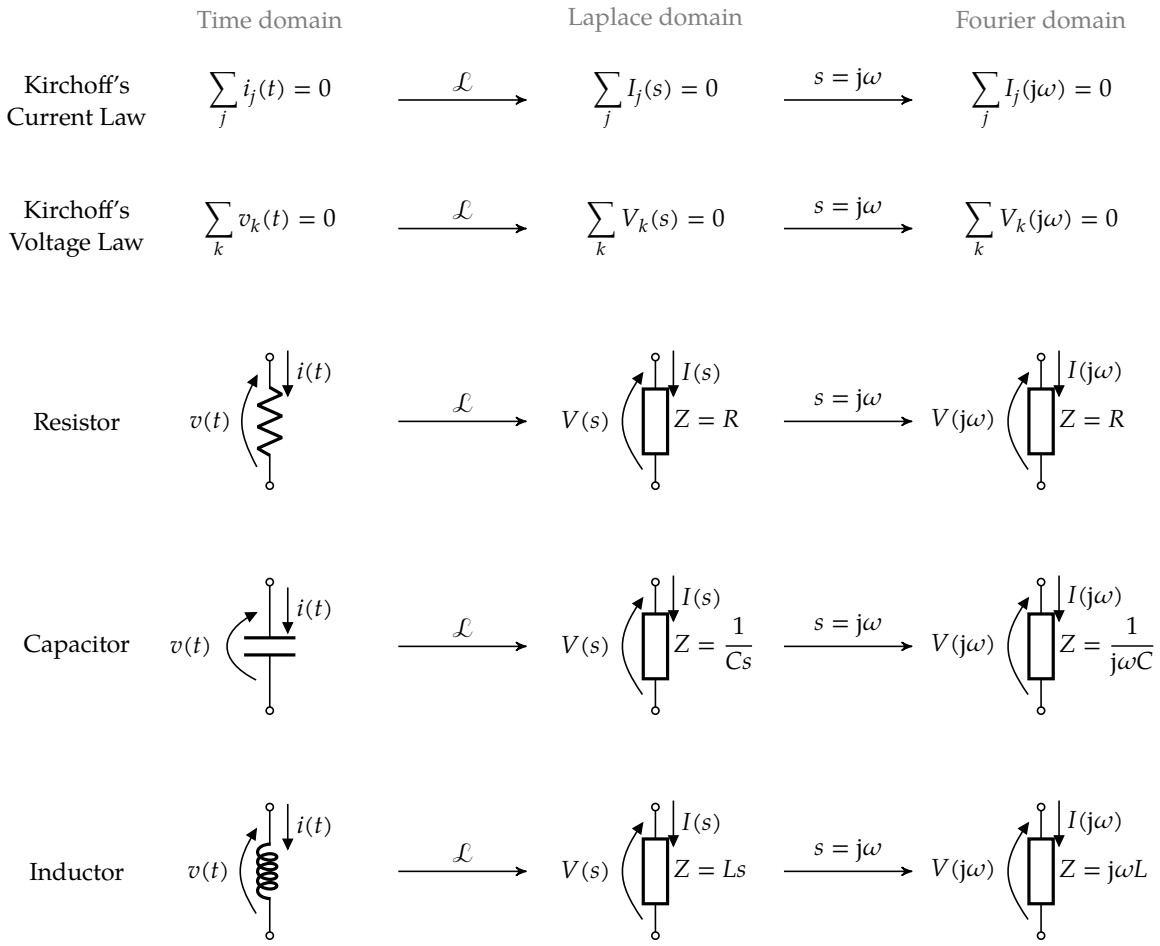


Figure 8.2: The transformation of the laws of electronics to the Fourier/Laplace domain

Starting from there, we can calculate the

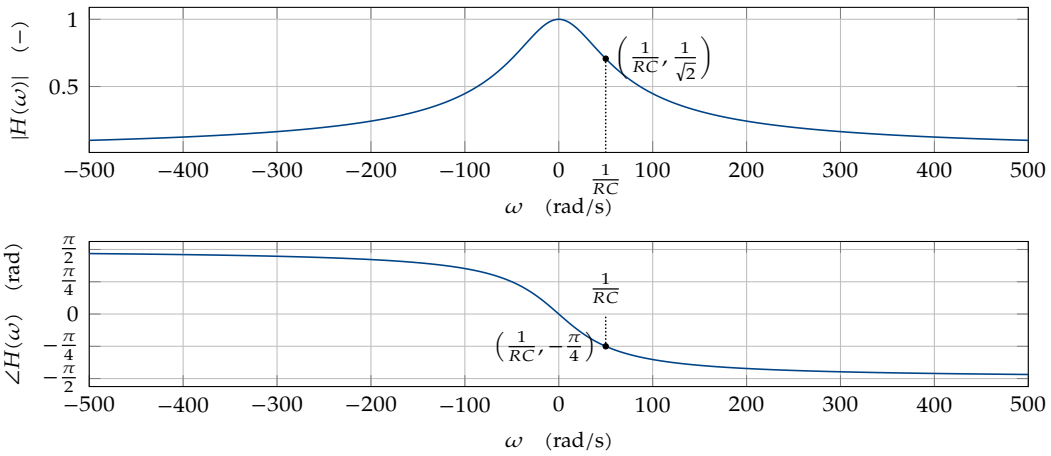
- magnitude:

$$|H(\omega)| = \left|\frac{1}{1+jRC\omega}\right| = \frac{|1|}{|1+jRC\omega|} = \frac{1}{\sqrt{1+(RC\omega)^2}}$$

- phase:

$$\angle H(\omega) = \angle\left(\frac{1}{1+jRC\omega}\right) = \angle(1) - \angle(1+jRC\omega) = 0 - \operatorname{atan}\left(\frac{RC\omega}{1}\right)$$

And based on those equations, we can make a plot of the spectrum, using any program capable of plotting functions. Try your favorite tool to generate the magnitude and phase plot below:



We can clearly see the low-pass nature of the filter, we even indicated the -3 dB point at which the gain has lowered by 3 dB i.e. lowered by a factor of $\sqrt{2}$.

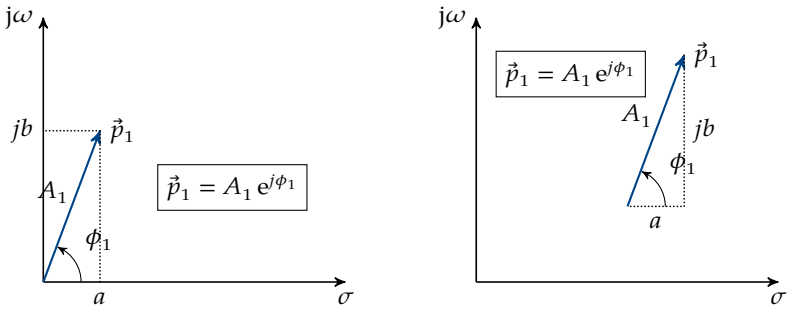
8.3 Analysis in the Argand plane

There is an alternative way to compose a magnitude and phase plot that starts from drawing the situation in the Argand plane and making a geometric reasoning in that plane. To understand this method, we need to make sure we remember some things well about the complex plane.

8.3.1 Some things to remember

Complex numbers are vectors with a magnitude (or length) and an angle

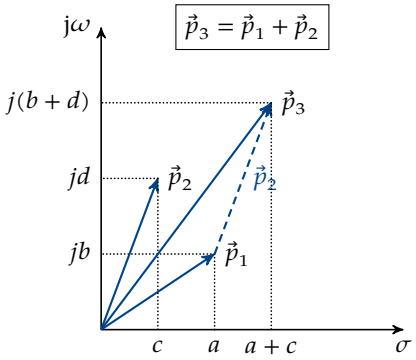
Any complex number $a + jb$ can be considered to be a vector $\vec{p}_1$ from the origin to the location of the complex number. We call such a vector a *bound vector* (see below left). However, the same number can also represent a *free vector* $\vec{p}_1$, as indicated below right:



Both vectors have a magnitude and an angle (that indicates the direction of the vector). We can use the complex number representation $a + jb$ and the vector representation $\vec{p}_1$ as equivalent.

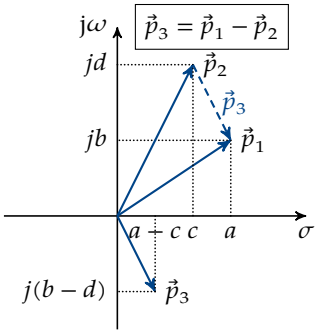
Addition is making a chain of vectors

Adding two complex numbers corresponds with creating a chain of their vectors (i.e. making the second free and moving it's origin to the end-point of the first one):



Subtraction is making a vector from one complex number to another

Subtracting two complex numbers represented by $\vec{p}_1 - \vec{p}_2$, corresponds to creating a new vector starting from the tip of the second vector $\vec{p}_2$ and pointing its tip to the tip of the first vector $\vec{p}_1$:



So far the things to remember. Now let's introduce something new.

8.3.2 Recognizing the vectors in a factorized transfer function

Consider the following generic transfer function of an LTI system:

$$H(s) = \frac{\sum_{j=0}^{m} b_j s^j}{\sum_{i=0}^{n} a_i s^i}$$

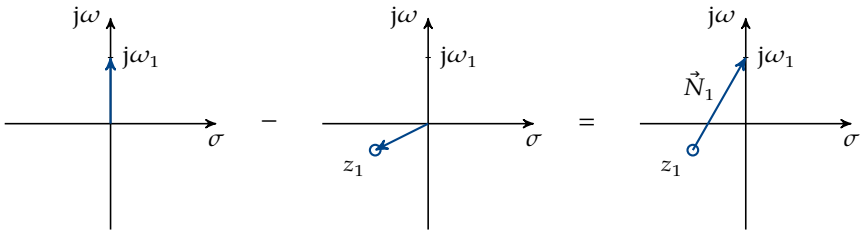
We can always factorize this transfer function into

$$H(s) = K \frac{(s - z_1)(s - z_2) \cdots (s - z_m)}{(s - p_1)(s - p_2) \cdots (s - p_n)}$$

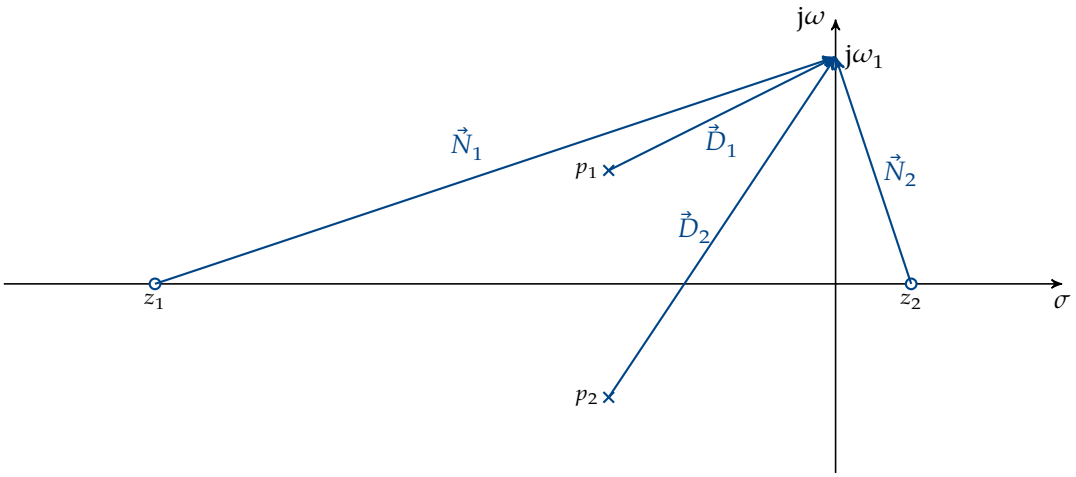
with $K = b_m / a_n$. To migrate from the Laplace to the Fourier domain, we need to substitute $s = \mathrm{j}\omega$:

$$H(\mathrm{j}\omega) = K \frac{\overbrace{(\mathrm{j}\omega - z_1)}^{\vec{N}_1} \overbrace{(\mathrm{j}\omega - z_2)}^{\vec{N}_2} \cdots \overbrace{(\mathrm{j}\omega - z_m)}^{\vec{N}_m}}{\underbrace{(\mathrm{j}\omega - p_1)}_{\vec{D}_1} \underbrace{(\mathrm{j}\omega - p_2)}_{\vec{D}_2} \cdots \underbrace{(\mathrm{j}\omega - p_n)}_{\vec{D}_n}} \tag{8.1}$$

The above expression shows that it consists of a product of factors each of which can be seen as a vector, as we can write it as a difference of two complex numbers, e.g. $\vec{N}_1 = \mathrm{j}\omega - z_1$. How this vector is formed is illustrated below:



This insight allows us to make a drawing of all these vectors, given a certain configuration of poles p_i and zeros z_j and a specific ω_1 that we want to consider. For example:



8.3.3 Using the vectors in a factorized transfer function

Now, how are those vectors useful? Well, the mathematical operations on these vectors (given where they appear mathematically in (8.1)) happen to be really simple.

Let's first focus on the magnitude:

$$|H(\mathrm{j}\omega)| = |K| \cdot \left| \frac{\vec{N}_1 \cdot \vec{N}_2 \cdots \vec{N}_m}{\vec{D}_1 \cdot \vec{D}_2 \cdots \vec{D}_n} \right| = |K| \cdot \frac{|\vec{N}_1| \cdot |\vec{N}_2| \cdots |\vec{N}_m|}{|\vec{D}_1| \cdot |\vec{D}_2| \cdots |\vec{D}_n|}$$

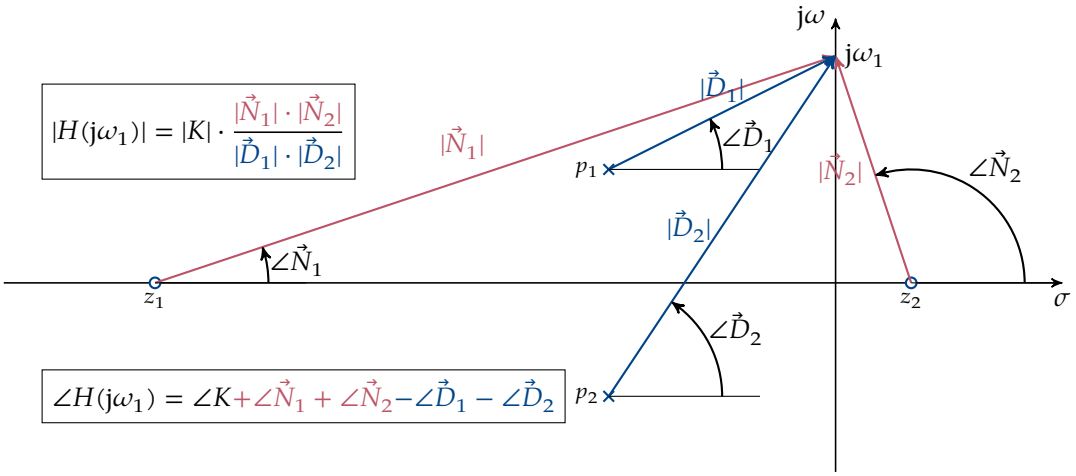
This means that we can calculate the magnitude by multiplying $|K|$ with all the lengths of the numerator vectors and dividing that by the product of the lengths of all denominator vectors.

Next, let's focus on the phase:

$$\begin{aligned} \angle\left(H(\mathrm{j}\omega)\right) &= \angle\left(K \cdot \frac{\vec{N}_1 \cdot \vec{N}_2 \cdots \vec{N}_m}{\vec{D}_1 \cdot \vec{D}_2 \cdots \vec{D}_n}\right) \\ &= \angle K + \angle\vec{N}_1 + \angle\vec{N}_2 + \cdots + \angle\vec{N}_m - \angle\vec{D}_1 - \angle\vec{D}_2 - \cdots - \angle\vec{D}_n \end{aligned}$$

This means that we can calculate the phase by summing all angles of the numerator vectors to the angle of K and then subtracting from that the sum of all angles of the denominator vectors. If you wonder what the angle of K is: if K is positive, its angle is zero degrees; if K is negative, the angle is 180° or π rad.

If we apply this to the example above, this leads to the following overall view:



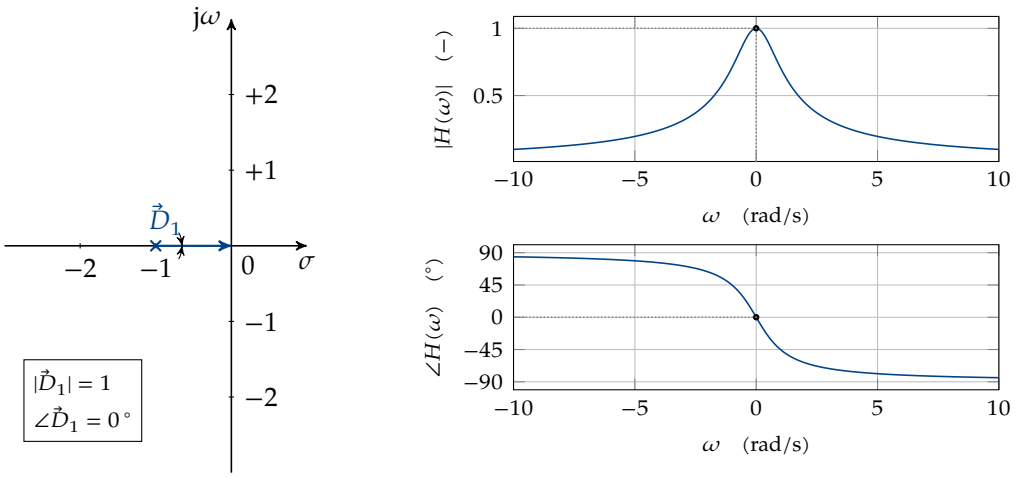
Now let's apply this technique to analyze first-order, second-order and all-pass systems.

8.3.4 First-order systems

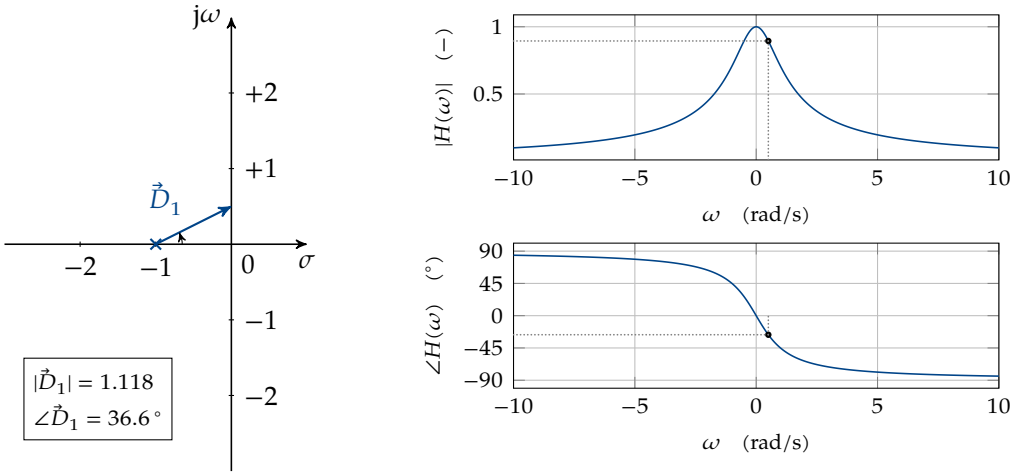
Let's apply this technique to a first-order system. As an example, we consider:

$$H(s) = \frac{1}{s+1}$$

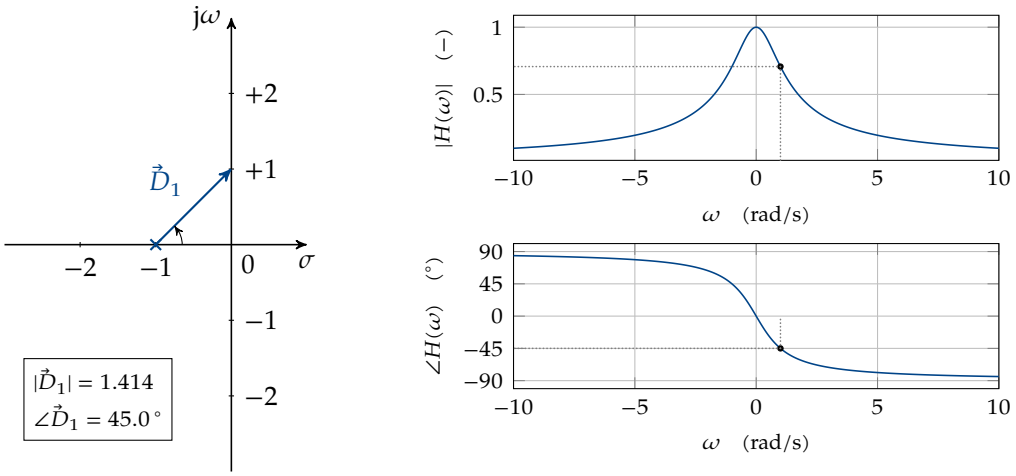
In Figure 8.3 and 8.4 we have illustrated the calculation of the magnitude and phase plot for several values of ω. The relevant 'measurements' (of lengths and angles) have been indicated on the pole-zero plot on the left-hand side. Considering this construction, based on measuring the length and the angle of D_1, helps you to understand why the phase starts at 90° and evolves to −90°. It also shows very clearly that when the phase reaches ±45° the magnitude decayed by $\sqrt{2}$ which corresponds with −3 dB.



(a) $\omega = 0\,\text{rad/s}$

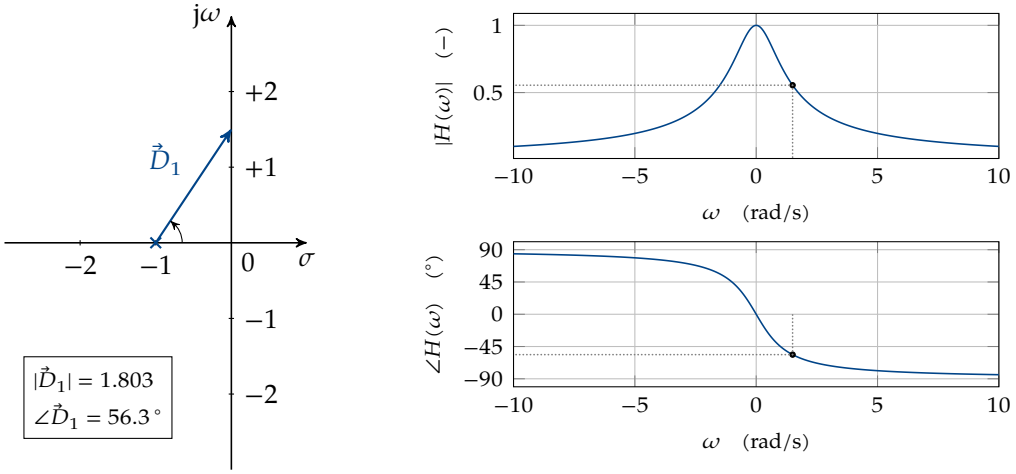


(b) $\omega = 0.5\,\text{rad/s}$

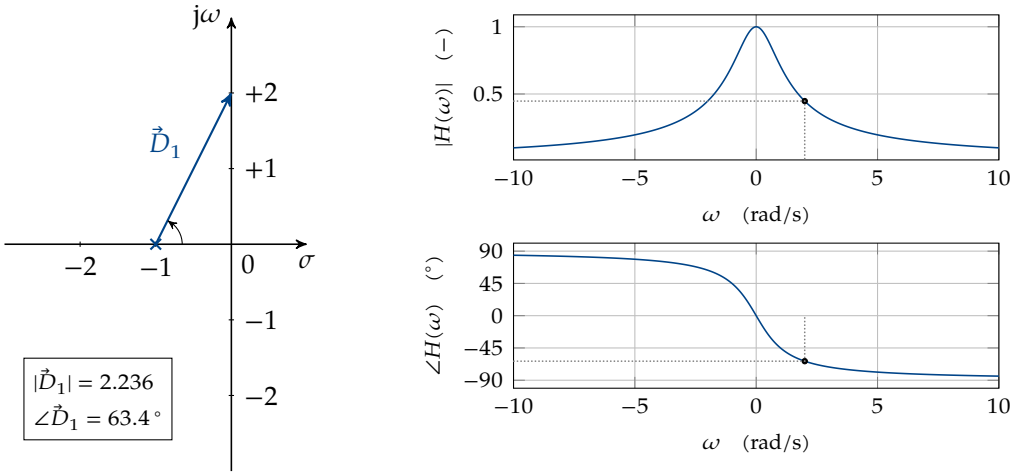


(c) $\omega = 1.0\,\text{rad/s}$

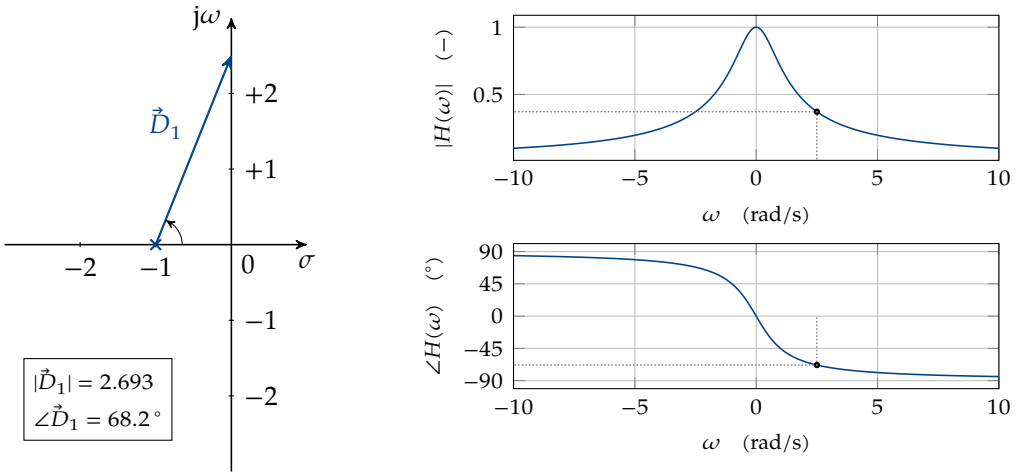
Figure 8.3: Illustration of the computation of the magnitude and phase plot of a first-order transfer function, with on the left-hand side the Argand plane and on the right-hand side the magnitude and phase plot on which the calculated point has been indicated; several values of ω are considered (first half — see facing page for the second half).



(a) $\omega = 1.5\,\text{rad/s}$



(b) $\omega = 2.0\,\text{rad/s}$



(c) $\omega = 2.5\,\text{rad/s}$

Figure 8.4: Illustration of the computation of the magnitude and phase plot of a first-order transfer function, with on the left-hand side the Argand plane and on the right-hand side the magnitude and phase plot on which the calculated point has been indicated; several values of ω are considered (second half — see facing page for the first half).

8.3.5 Second-order systems

Let's apply the same technique to a second-order system. Let's use the following example system:

$$H(s) = \frac{1}{s^2 + 2s + 4.0625}$$

In Figure 8.5 and 8.6 we have illustrated the calculation of the magnitude and phase plot for several values of ω. The relevant 'measurements' (of lengths and angles) have been indicated on the pole-zero plot on the left-hand side. Considering this construction, based on measuring the length and the angle of D_1 and D_2, helps you to understand why the phase starts at 180° and evolves to −180°. It is also striking to see that there is a peak in the magnitude plot (a so called frequency resonance).

Resonance — analytically

Let's investigate this resonance from an analytical point of view. But let's not only do it for the example, but let's investigate it in general, i.e. we start from the following transfer function:

$$H(s) = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$$

We move over to the Fourier domain by setting $s = \mathrm{j}\omega$:

$$\begin{aligned} H(\mathrm{j}\omega) &= \frac{\omega_n^2}{-\omega^2 + 2\zeta\omega_n \mathrm{j}\omega + \omega_n^2} \\ &= \frac{\omega_n^2}{\omega_n^2 - \omega^2 + j2\zeta\omega_n\omega} \end{aligned}$$

Given the fact that we'd like to investigate the magnitude, let's focus on that aspect alone:

$$|H(\mathrm{j}\omega)| = \frac{\omega_n^2}{\sqrt{(\omega_n^2 - \omega^2)^2 + (2\zeta\omega_n\omega)^2}} = \omega_n^2 \left((\omega_n^2 - \omega^2)^2 + (2\zeta\omega_n\omega)^2\right)^{-\frac{1}{2}} \tag{8.2}$$

The frequency resonance is a maximum in the magnitude plot. Let's therefore determine the extrema, which should be located in the stationary points of the graph, i.e. we're looking for values of ω that fulfill

$$\frac{\mathrm{d}|H(\mathrm{j}\omega)|}{\mathrm{d}\omega} = 0$$

Let's for the sake of organizing our calculations smoothly, switch the two parts of the equation:

$$\begin{aligned} 0 &= \frac{\mathrm{d}|H(\mathrm{j}\omega)|}{\mathrm{d}\omega} \\ &= \frac{\mathrm{d}}{\mathrm{d}\omega}\left[\omega_n^2 \left((\omega_n^2 - \omega^2)^2 + (2\zeta\omega_n\omega)^2\right)^{-\frac{1}{2}}\right] \\ &= -\frac{1}{2}\omega_n^2 \left((\omega_n^2 - \omega^2)^2 + (2\zeta\omega_n\omega)^2\right)^{-\frac{3}{2}} \cdot \left(2(\omega_n^2 - \omega^2)\cdot(-2\omega) + 8\zeta^2\omega_n^2\omega\right) \\ &= -\frac{1}{2}\omega_n^2 \left((\omega_n^2 - \omega^2)^2 + (2\zeta\omega_n\omega)^2\right)^{-\frac{3}{2}} \cdot \left(-4(\omega_n^2 - \omega^2)\omega + 8\zeta^2\omega_n^2\omega\right) \\ &= -\frac{1}{2}\omega_n^2 \frac{\left(-4(\omega_n^2 - \omega^2)\omega + 8\zeta^2\omega_n^2\omega\right)}{\sqrt{\left((\omega_n^2 - \omega^2)^2 + (2\zeta\omega_n\omega)^2\right)^3}} \end{aligned}$$

This expression can only be zero if the numerator is zero, i.e.:

$$\begin{aligned} & 0 = -4(\omega_n^2 - \omega^2)\omega + 8\zeta^2\omega_n^2\omega \\ \Leftrightarrow \quad & 0 = \omega \cdot \left(-(\omega_n^2 - \omega^2) + 2\zeta^2\omega_n^2\right) \end{aligned}$$

Based on the final equation, we conclude that there are stationary points under two conditions, leading to

1. the flat center point of the magnitude graph:

$$\omega = 0$$

2. the resonance frequency:

$$-(\omega_n^2 - \omega_r^2) + 2\zeta^2\omega_n^2 = 0 \tag{8.3a}$$

$$\Leftrightarrow \qquad \omega_r = \omega_n\sqrt{1 - 2\zeta^2} \tag{8.3b}$$

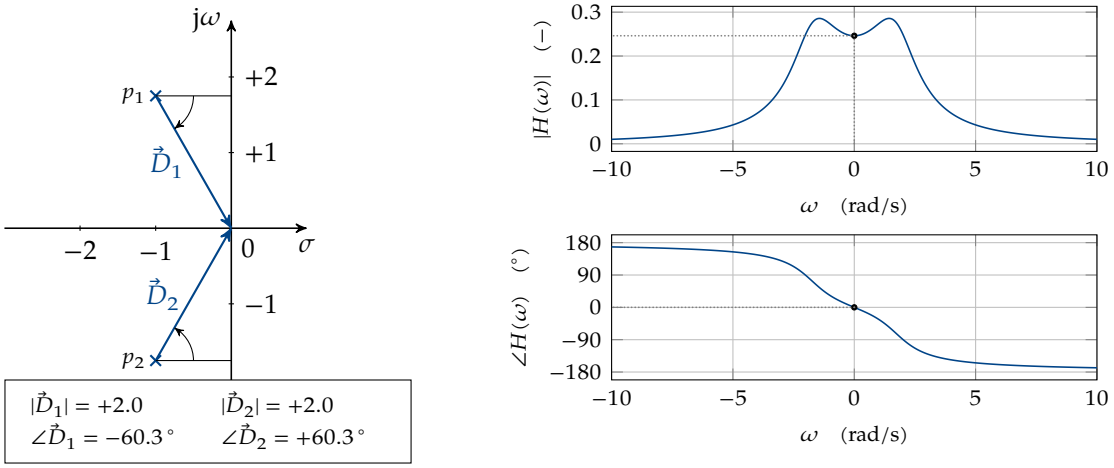
That final expression is of course only valid when $1 - 2\zeta^2 \geq 0$, i.e.

$$\begin{aligned} & \zeta^2 \leq \frac{1}{2} \\ \Leftrightarrow \quad & |\zeta| \leq \frac{1}{\sqrt{2}} = 0.7071 \end{aligned}$$

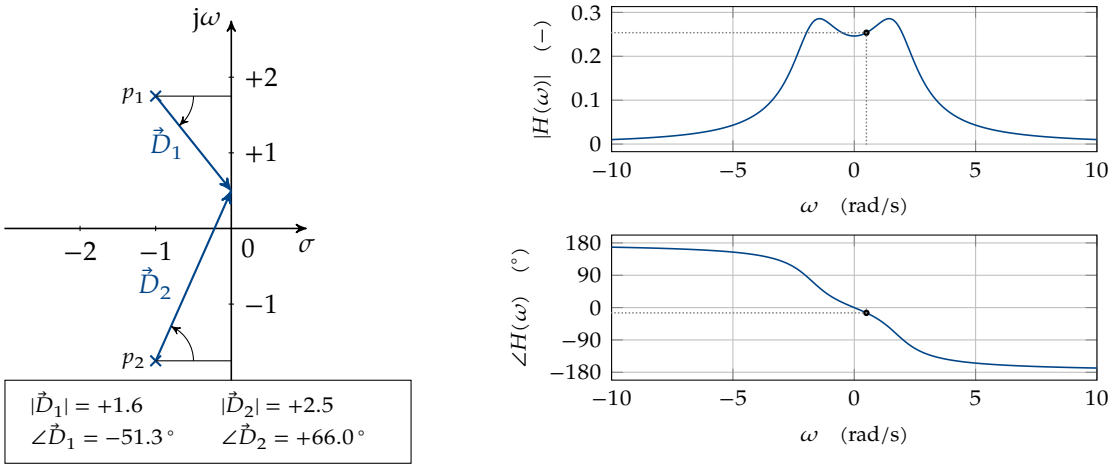
To obtain the height of the resonance peak, we plug (8.3a) in the first term and (8.3b) in the second term of the denominator of (8.2):

$$\begin{aligned} |H(\mathrm{j}\omega_r)| &= \frac{\omega_n^2}{\sqrt{(\omega_n^2 - \omega_r^2)^2 + (2\zeta\omega_n\omega_r)^2}} \\ &= \frac{\omega_n^2}{\sqrt{4\zeta^4\omega_n^4 + 4\zeta^2\omega_n^4\left(1 - 2\zeta^2\right)}} \\ &= \frac{1}{4\zeta^4 + 4\zeta^2 - 8\zeta^4} \\ &= \frac{1}{\sqrt{4\zeta^2 - 4\zeta^4}} \\ &= \frac{1}{2\zeta\sqrt{1 - \zeta^2}} \end{aligned}$$

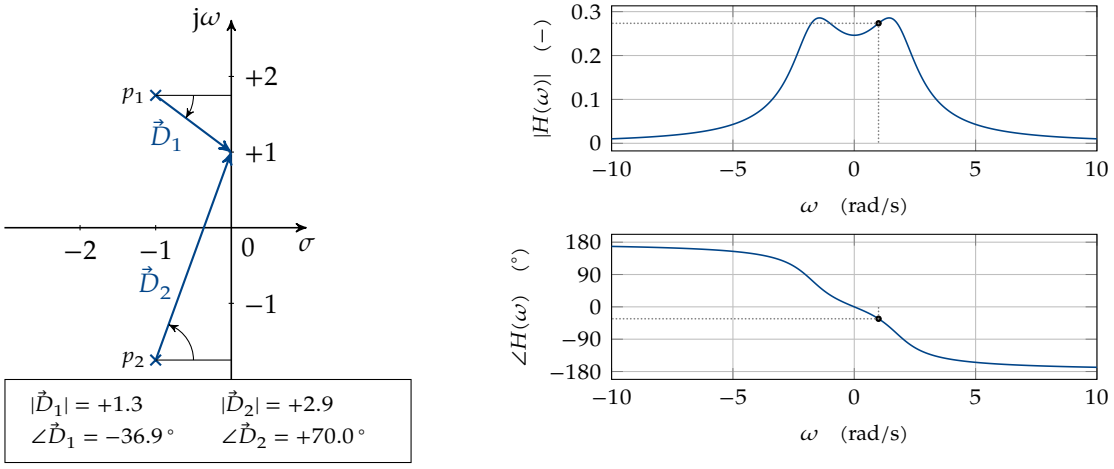
Note that the height of the peak is not dependent on the location of the peak (ω_r)! Remember, this equation only holds if $|\zeta| \leq \frac{1}{\sqrt{2}}$.



(a) $\omega = 0.0\,\text{rad/s}$

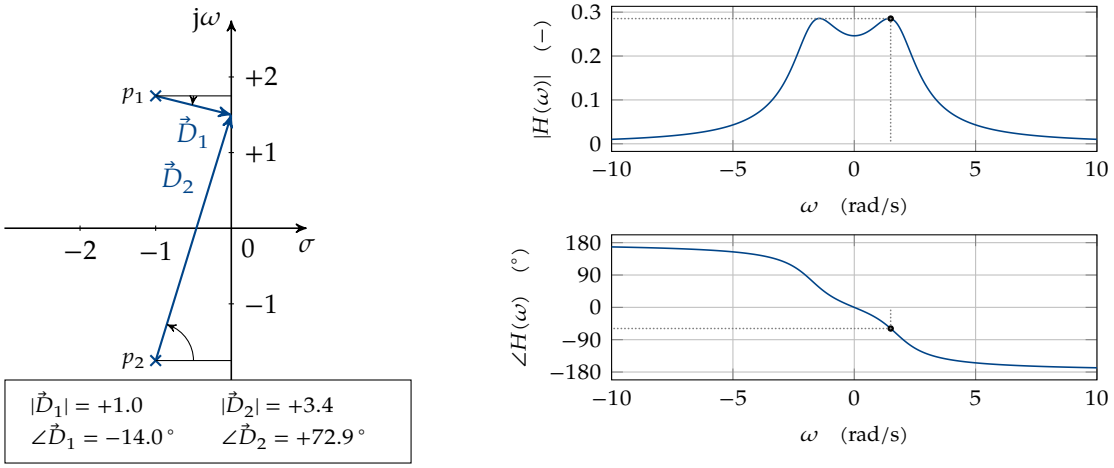


(b) $\omega = 0.5\,\text{rad/s}$

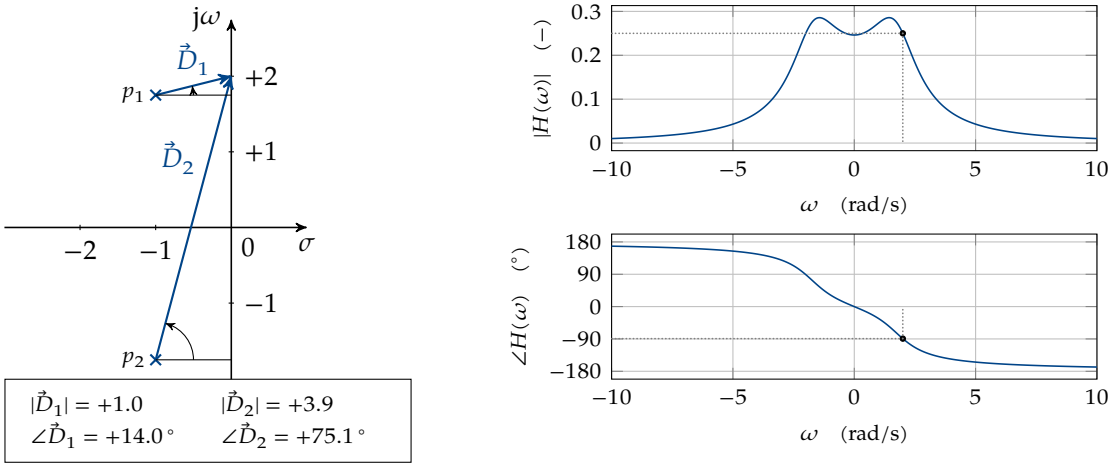


(c) $\omega = 1.0\,\text{rad/s}$

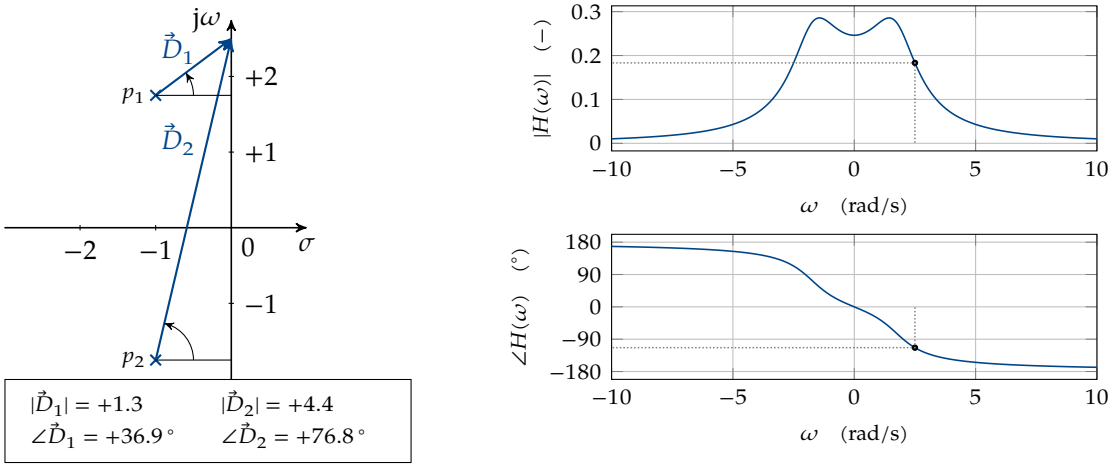
Figure 8.5: Illustration of the computation of the magnitude and phase plot of a second-order transfer function, with on the left-hand side the Argand plane and on the right-hand side the magnitude and phase plot on which the calculated point has been indicated; several values of ω are considered (first half — see facing page for the second half).



(a) $\omega = 1.5\,\text{rad/s}$



(b) $\omega = 2.0\,\text{rad/s}$

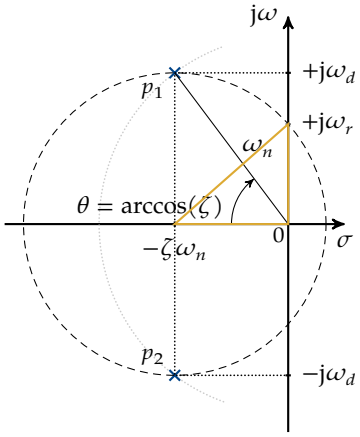


(c) $\omega = 2.5\,\text{rad/s}$

Figure 8.6: Illustration of the computation of the magnitude and phase plot of a second-order transfer function, with on the left-hand side the Argand plane and on the right-hand side the magnitude and phase plot on which the calculated point has been indicated; several values of ω are considered (second half — see facing page for the first half).

Resonance — graphically

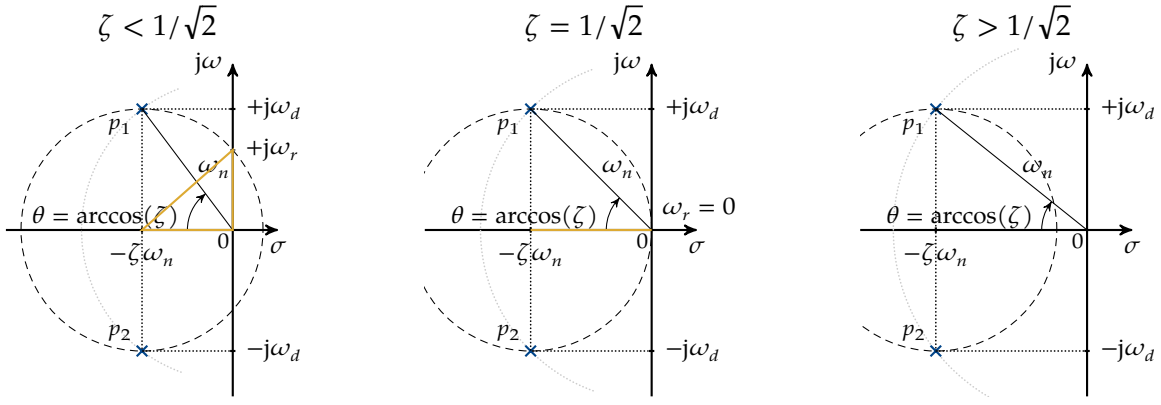
Let's also investigate this resonance from a graphical point of view. Some smart-ass mathematicians were able to pinpoint the resonance frequency ω_r onto the pole-plot of a second-order system (in the Argand plane). This is what they found: if you draw a circle centered on the real axis on the real part of the pole locations and you give it ω_d as radius, then ω_r is the crossing of that circle with the imaginary axis. This wonderful idea has been illustrated in the drawing below:



And indeed, the orange triangle allows applying the Pythagorean theorem: its vertical side is the square of the hypothenuse minus the horizontal side. Let's label the vertical side ω_r and check if we arrive at the proper equation:

$$\begin{aligned}\omega_r &= \sqrt{\omega_d^2 - (\zeta\omega_n)^2} \\ &= \sqrt{(\omega_n\sqrt{1-\zeta^2})^2 - (\zeta\omega_n)^2} \\ &= \sqrt{\omega_n^2(1-\zeta^2) - \omega_n^2\zeta^2} \\ &= \omega_n\sqrt{1-2\zeta^2}\end{aligned}$$

Therefore the claim has been proven. This also clarifies why for some values of ζ there is no frequency resonance. The drawing below compares three situations:



$\omega_r = \omega_n\sqrt{1-2\zeta^2}$ $\quad$ $\omega_r = \omega_n\sqrt{1-2\zeta^2} = 0$ $\quad$ no frequency resonance

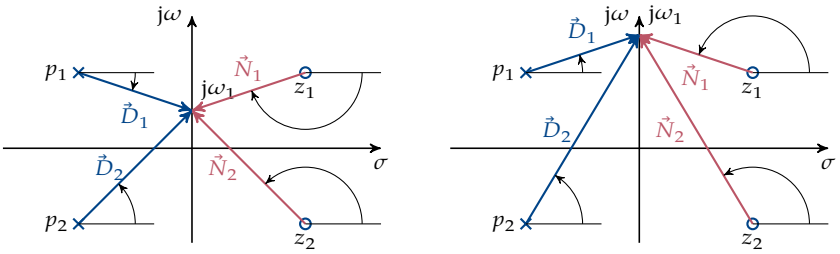
In the left-hand drawing you can clearly see an intersection of the dashed circle with the imaginary axis. When ζ increases to $1/\sqrt{2}$ the intersection moves downwards until $\omega_r = 0$. As of that

situation (and increasing ζ) there is no more resonance as there is no intersection.

The magnitude and phase plot of a second-order LTI system for various values of ζ can be found in Figure 8.7. Note how in the magnitude plot the peak (for tiny values of ζ) starts at $\omega_r \approx \omega_n$ and then lowers in frequency as the peak comes down (for increasing values of ζ). As soon as $\zeta = 1/\sqrt{2}$ there is no more frequency resonance — the magnitude curve goes down monotonically. The phase changes over the midway point (90° for the negative frequencies and −90° for the positive frequencies, always exactly at $\omega = \omega_n$! The lower ζ, the sharper the transition.

8.3.6 All-pass systems

An all-pass system is a system that has a flat magnitude response. Its only purpose is to perform a phase change. The way we build this, is by putting poles and zeros on pairs. Every pole is accompanied by a zero mirrored w.r.t. the imaginary axis. Below, we illustrate this for a second-order system:



The transfer function that goes with this configuration is:

$$H(s) = K\frac{s^2 - 2\zeta\omega_n s + \omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$$

The conclusion is obvious, given the fact that every vector in the numerator is balanced by a vector in the denominator (in the example $|\vec{N}_1| = |\vec{D}_1|$ and $|\vec{N}_2| = |\vec{D}_2|$). This results in:

$$|H(j\omega)| = K\frac{|\vec{N}_1| \cdot |\vec{N}_2|}{|\vec{D}_1| \cdot |\vec{D}_2|} = K$$

Therefore, except for the gain K, only a phase change occurs as a function of ω.

8.3.7 Minimum vs. non-minimum phase systems

Based on this insight, we can also understand how we can make a filter with a peculiar magnitude spectrum in two versions. One with the zeros in the left half plane and one with the zeros in the right half plane. We illustrate this with the example below:

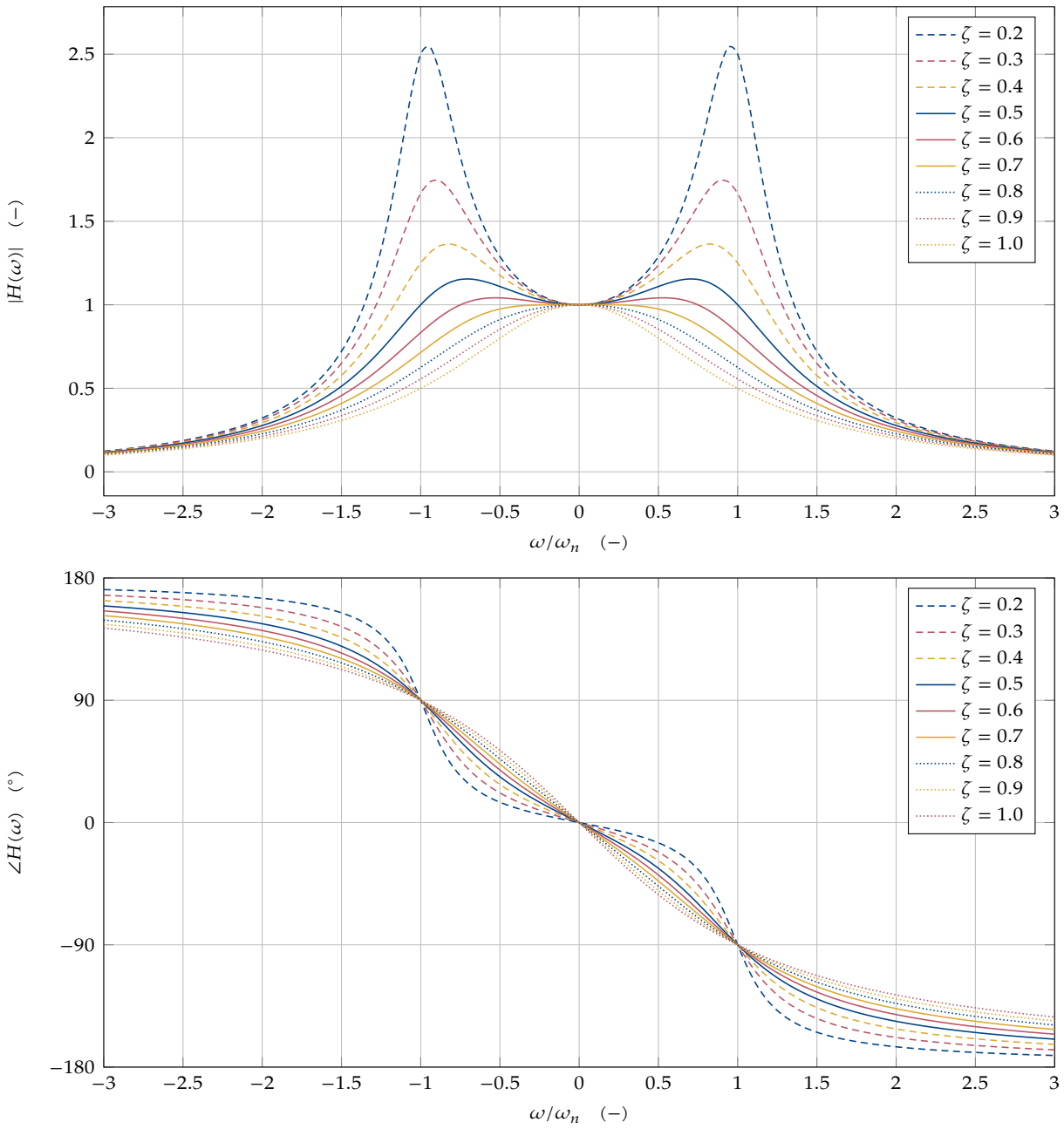
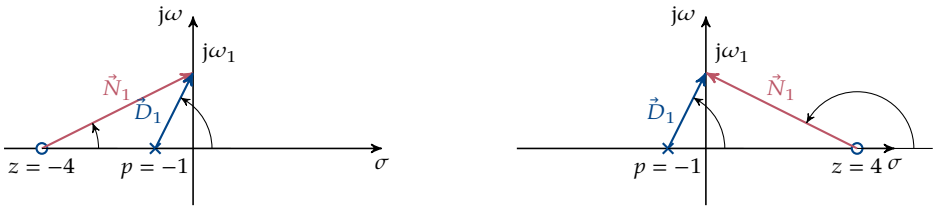


Figure 8.7: The magnitude and phase spectrum of a second-order LTI system (without zeros) with the damping factor ζ as a parameter



From the equations below, and the drawings above, it is clear that both filters have the same magnitude spectrum, but a different phase spectrum:

$$|H(\mathrm{j}\omega)| = K\frac{|\vec{N}_1|}{|\vec{D}_1|} \qquad \angle H(\mathrm{j}\omega) = \angle\vec{N}_1 - \angle\vec{D}_1$$

When the zeros are located in the left half plane, we say that the filter is a *minimum-phase filter*. If the zeros are not (all) located in the left half plane, we say that the filter is a *non-minimum-phase filter*. We elaborated the two situations side-by-side for our example below:

Minimum phase

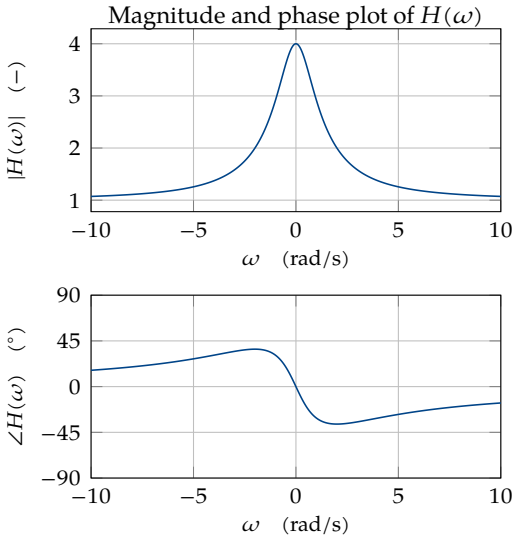
$$H(s) = \frac{s+4}{s+1}$$
$$H(\mathrm{j}\omega) = \frac{4+\mathrm{j}\omega}{1+\mathrm{j}\omega}$$
$$|H(\mathrm{j}\omega)| = \frac{\sqrt{16+\omega^2}}{\sqrt{1+\omega^2}}$$
$$\angle H(\mathrm{j}\omega) = \operatorname{atan}\left(\frac{\omega}{4}\right) - \operatorname{atan}\left(\frac{\omega}{1}\right)$$

Non-minimum phase

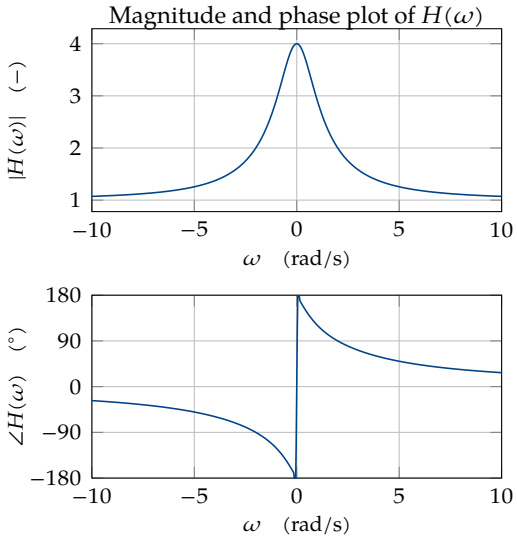
$$H(s) = \frac{s-4}{s+1}$$
$$H(\mathrm{j}\omega) = \frac{-4+\mathrm{j}\omega}{1+\mathrm{j}\omega}$$
$$|H(\mathrm{j}\omega)| = \frac{\sqrt{16+\omega^2}}{\sqrt{1+\omega^2}}$$
$$\angle H(\mathrm{j}\omega) = \pi - \operatorname{atan}\left(\frac{\omega}{4}\right) - \operatorname{atan}\left(\frac{\omega}{1}\right)$$

You can find the plots of the magnitude and the phase below. As you can clearly see, the magnitude plot is the same, and the phase plot achieves smaller values for the minimum phase version than for the non-minimal phase version.

Minimum phase



Non-minimum phase



Note how in the graph on the bottom right, the curve makes a jump from $-180°$ to $-180°$, while this discontinuity is not there in the corresponding equation of the phase. This is due to the fact that we often plot phase graphs in the range of $-180°$ tot $-180°$. When the curve at $\omega = 0\,\text{rad/s}$ goes below $-180°$ we 'phase-wrap' it to become its value plus $360°$.

8.4 Bode plots — in theory

8.4.1 Genesis

Magnitude and phase plots in their purest form, i.e. with a linear frequency scale, and a linear magnitude and phase scale, exhibit a number of problems:

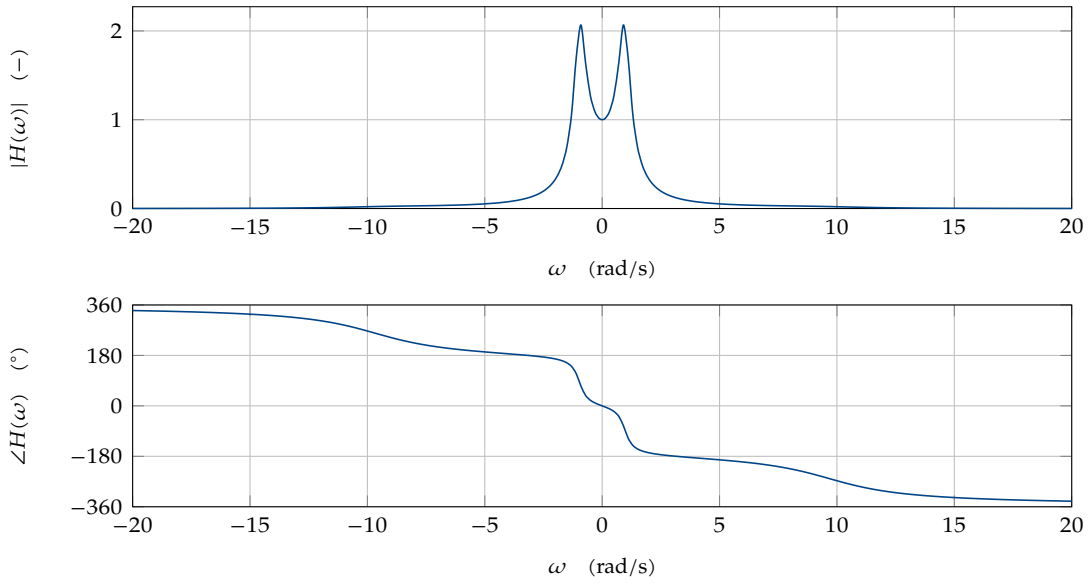
- they hide higher-order / smaller effects,
- their transition effects are short in the frequency domain for small frequencies and wider in the frequency domain for large frequencies.

Let's first observe these two effects using the following example:

$$H(s) = \frac{1}{\underbrace{s^2 + 0.5s + 1}_{\zeta=0.25 \text{ and } \omega_n=1\,\text{rad/s}}} \cdot \frac{100}{\underbrace{s^2 + 5s + 100}_{\zeta=0.25 \text{ and } \omega_n=10\,\text{rad/s}}}$$

Note that this is a cascade of two second-order systems, one active at $\omega_n = 1\,\text{rad/s}$, the second active at $\omega_n = 10\,\text{rad/s}$. Both should show a resonance peak, as $\zeta < 1/\sqrt{2}$, peaking with a relative amplitude of a factor 2.

We graphed the magnitude and phase plot below:



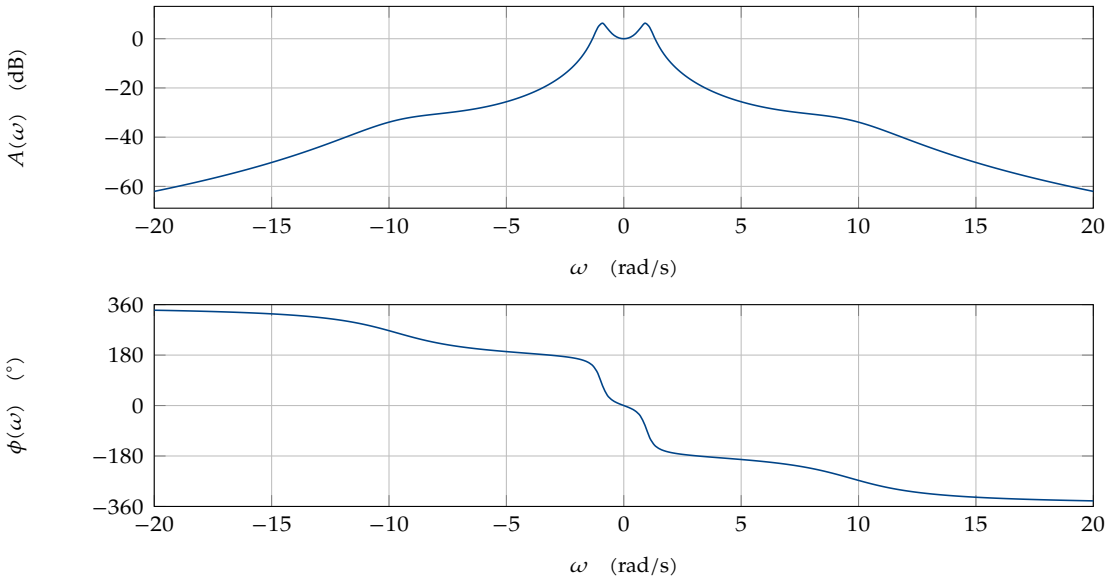
Note how the first resonsance peak (at $\omega = 1\,\text{rad/s}$) is clearly visible, while the second one (near $\omega = 10\,\text{rad/s}$) drowns in the lack of resolution for small magnitudes. However, we can clearly see it in the phase plot.

The solution is straightforward: make the magnitude scale logarithmic! Instead of plotting $|H(\omega)|$, we will be plotting:

$$A(\omega) = 20\log_{10}(|H(\mathrm{j}\omega)|) \qquad \text{(dB)}$$

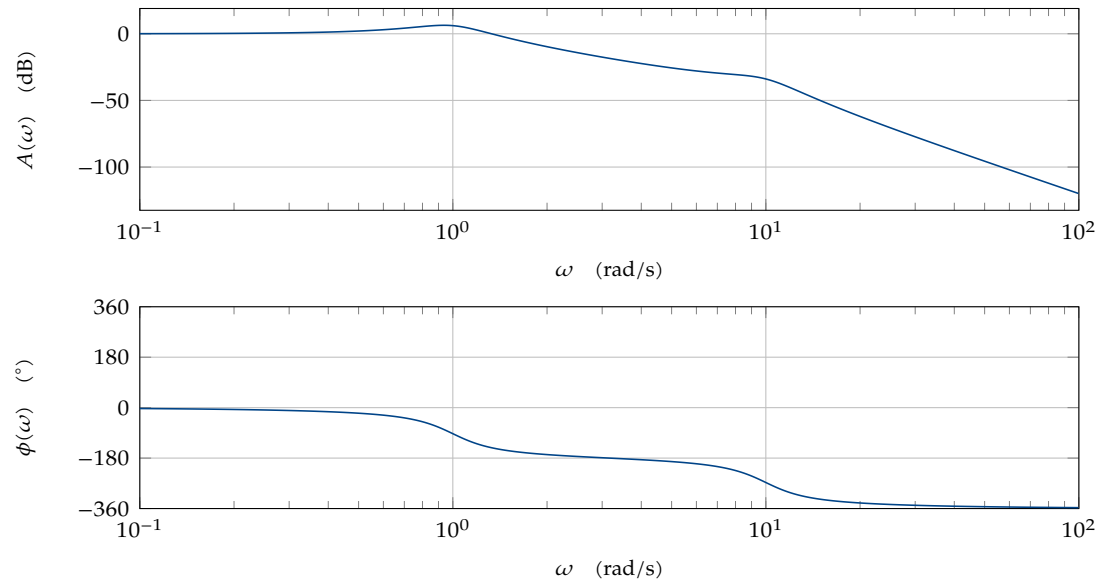
We often use the the abbreviation $\phi(\omega)$ as a synonym for $\angle H(\omega)$.

The result of the magnitude scale improvement can be observed in the magnitude and phase plot below:



The second resonance peak kan now be observed in the magnitude plot around $\omega = 10\,\mathrm{rad/s}$. Still, one problem remains: the length of the phase transition is very different for the two second-order systems that make up the cascade.

The solution for this problem is to use a logarithmic frequency scale. Instead of using ω on the horizontal axis, we will use $\log\omega$. This presupposes that we abandon the two-sided plot because the negative axis is not part of the domain of the logarithmic function. However, the result looks spectacularly good:



Why the enthusiasm? Well...next to the equalization of the phase transitions, also in the magnitude domain the transitions look much more alike. The curve even almost simplified itself to a concatenation of straight lines!

The combo of a logarithmic magnitude scale and a logarithmic frequency scale caught the eye of Hendrik Wade Bode in the 1930s (see Figure 8.8), leading to his insight that an approximation with straight line segments would allow control engineers to make insightful diagrams without the need for intricate calculations. His groundbreaking work still determines our view on LTI systems today.

Figure 8.8: Hendrik Wade Bode (*1905–†1982)

8.4.2 The logarithmic magnitude axis — two worlds

Making the magnitude axis logarithmic has another consequence that greatly simplifies working with Bode plots. Remember the equation below, that revealed that the magnitude can be seen as multiplication of factors:

$$|H(\mathrm{j}\omega)| = |K| \cdot \left| \frac{\vec{N}_1 \cdot \vec{N}_2 \cdots \vec{N}_m}{\vec{D}_1 \cdot \vec{D}_2 \cdots \vec{D}_n} \right| = |K| \cdot \frac{|\vec{N}_1| \cdot |\vec{N}_2| \cdots |\vec{N}_m|}{|\vec{D}_1| \cdot |\vec{D}_2| \cdots |\vec{D}_n|}$$

Given the fact that we will not plot $|H(\mathrm{j}\omega)|$ but $20\log_{10}|H(\mathrm{j}\omega)|$, this yields:

$$\begin{aligned} 20\log_{10}|H(\mathrm{j}\omega)| &= 20\log_{10}\left(|K| \frac{|\vec{N}_1| \cdot |\vec{N}_2| \cdots |\vec{N}_m|}{|\vec{D}_1| \cdot |\vec{D}_2| \cdots |\vec{D}_n|} \right) \\ &= 20\log_{10}|K| + 20\log_{10}|\vec{N}_1| + 20\log_{10}|\vec{N}_2| + \cdots + 20\log_{10}|\vec{N}_m| \\ &\quad - 20\log_{10}|\vec{D}_1| - 20\log_{10}|\vec{D}_2| - \cdots - 20\log_{10}|\vec{D}_n| \end{aligned}$$

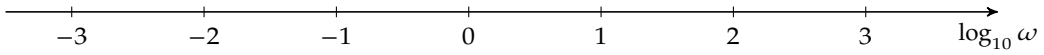
Therefore the difficult graphical operation of multiplication reduces itself to the easier graphical operation of addition. This joins the two worlds of magnitude and phase graphs as they both become sums of contributions.

You might also want to review 'the decibel' in Appendix A to fully understand how the new unit of the logarithmic magnitude axis relates to the original linear values.

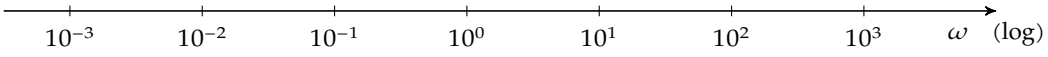
8.4.3 The logarithmic frequency axis — land of confusion

A logarithmic axis is a tricky thing, especially while we have the habit of not displaying $\log x$ as values on the axis, but x-values on the locations of $\log x$. As an example, let's consider the logarithmic frequency axis.

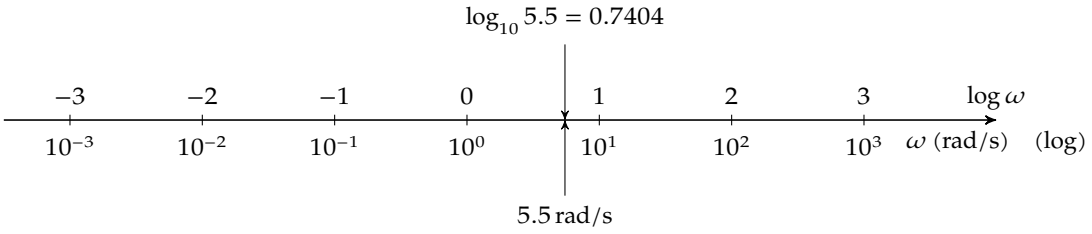
First, consider the pure $\log\omega$-axis below:



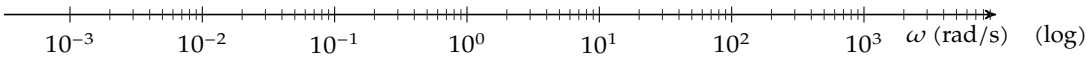
When $\omega = 100$, this means $\log\omega = 2$ and therefore in the drawing above, the position indicated on the scale by the label 2, means $\log\omega = 2$ or $\omega = 10^2$. In this way, we can draw a different axis that is still logarithmic, but contains the original values as labels. We stress the fact that the scale is logarithmic by indicating a '(log)' next to the axis label ω.[1]



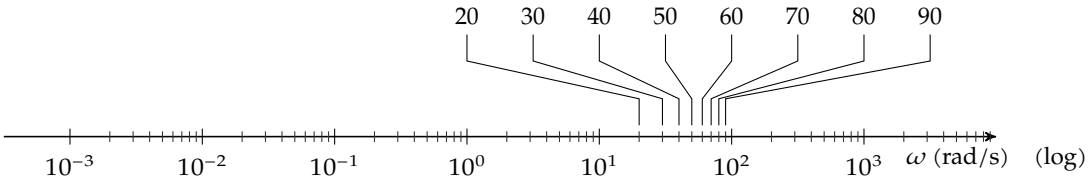
An obvious question is: where is $\omega = 5.5\,\text{rad/s}$ on this axis? Is its value halfway in between $\omega = 1$ and $\omega = 10$, therefore straight in the middle of both? No. The only proper way is to think again about the fact that the underlying axis is logarithmic, so we need to think of 5.5 as $\log 5.5 = 0.7404$. Therefore, it is almost three quarters on the stretch from 10^0 to 10^1.



To help the user of the axis, often a linear grid is imposed on the logarithmic scale:



Familiarize yourself with finding values on the axis! To help you, the values in between 10^1 and 10^2 have been indicated on the scale below.



8.4.4 The combination of both — sledgehammer!

There is great power in the combination of a logarithmic frequency axis and a logarithmic magnitude axis. We will illustrate that with the magnitude plot of a first-order system:

$$H(s) = \frac{1/\tau}{s + 1/\tau} \xrightarrow{s=j\omega} A(\omega) = |H(\mathrm{j}\omega)|_{\mathrm{dB}} = 20\log_{10}|H(\mathrm{j}\omega)| = 20\log_{10}\frac{1/\tau}{\sqrt{\omega^2 + 1/\tau^2}}$$

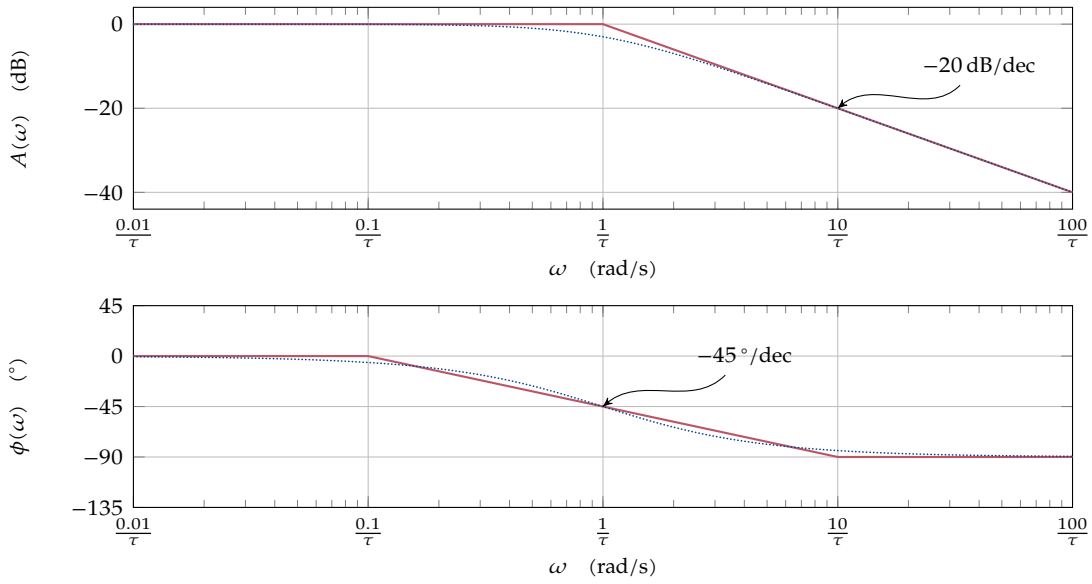
[1] Remember, when we write log in fact we mean $\log_{10}$, i.e. the Briggsian logarithms.

We can simplify this graph into two segments:

- $\omega \ll \dfrac{1}{\tau}$: $\quad A(\omega) = 20\log_{10}\dfrac{1/\tau}{\sqrt{\cancel{\omega^2}+1/\tau^2}} = 0\,\text{dB}$
- $\omega \gg \dfrac{1}{\tau}$: $\quad A(\omega) = 20\log_{10}\dfrac{1/\tau}{\sqrt{\omega^2+\cancel{1/\tau^2}}} = 20\log_{10}\frac{1}{\tau} - 20\log_{10}\omega$

Both segments are straight lines! We call these asymptotes. It is as if all the curves have been made flat by a sledgehammer.

The situation is illustrated in the graph below. The dashed red curves indicate the asymptotic approximation and the blue (dotted) curve the actual unapproximated graphs.



Note that also for the phase graph a decent *polylinear approximation* is possible! The errors are quite acceptable. The maximal error occurs for the magnitude at the location of the pole, and can be calculated to be:

$$A\left(\frac{1}{\tau}\right) = 20\log_{10}\frac{1}{\sqrt{1+1}} \approx -3.01\,\text{dB}$$

when compared to the horizontal 0 dB asymptote.

The maximal phase error occurs for $\omega = 0.1/\tau$ and can be calculated to be:

$$\phi\left(\frac{0.1}{\tau}\right) = -\operatorname{atan}\left(\frac{0.1}{\tau}\right) \approx -5.71^\circ$$

when compared to 0° asymptote. The situation at $\omega = 10/\tau$ is similar.

Exercises

Exercise 8.4.4-1:
Convert the linear voltage gains A to the corresponding logarithmic voltage gains A_{dB} and vice versa, and the frequencies f (in Hz) to frequencies in decades f_{dec} and vice versa (i.e. fill out the dots):

$A(-)$	$A_{\text{dB}}(\text{dB})$
15	...
-70	...
0.02	...
-0.1	...
...	70
...	122
...	-20
...	0

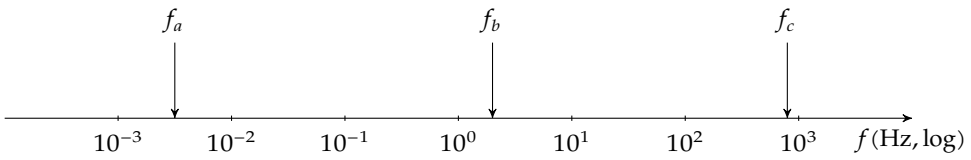
$f(\text{Hz})$	$f_{\text{dec}}(\text{decade})$
15×10^3	...
-700	...
0.02	...
-0.1	...
...	3.2
...	10.5
...	-4
...	0

The converted values are filled out in the tables below. Note that logarithmic values are only defined on strictly positive values (hence the $|\cdot|$).

$A(-)$	$A_{\text{dB}}(\text{dB})$
15	23.52
$\lvert -70\rvert$	36.90
0.02	-33.98
$\lvert -0.1\rvert$	-20
3162	70
1.259×10^6	122
0.1	-20
1	0

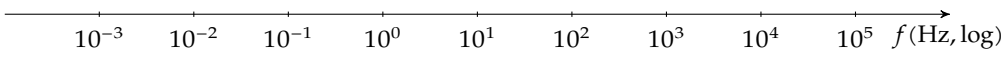
$f(\text{Hz})$	$f_{\text{dec}}(\text{decade})$
15×10^3	4.18
$\lvert -700\rvert$	2.85
0.02	-1.70
$\lvert -0.1\rvert$	-1
1.585×10^3	3.2
31.623×10^9	10.5
1×10^{-4}	-4
1	0

Exercise 8.4.4-2:
Consider the logarithmic scale below. Determine the values of f_a, f_b and f_c:



Exercise 8.4.4-3:
Indicate the following frequencies on the scale below:

$$f_a = 300\,\text{Hz} \qquad f_b = 0.05\,\text{Hz} \qquad f_c = 80\,\text{kHz}$$



8.5 Bode plots — in practice

8.5.1 The bricks

The basis for drawing asymptotic Bode plots is the observation is that the transfer function of any LTI system can be factorized into first- and second-order contributions:

$$H(s) = K \cdot \frac{\prod_{j=1}^{m_1}(s - z_j) \cdot \prod_{j=1}^{m_2}(s^2 + 2\zeta_{z,j}\omega_{n,z,j}s + \omega_{n,z,j}^2)}{\prod_{i=1}^{n_1}(s - p_i) \cdot \prod_{i=1}^{n_2}(s^2 + 2\zeta_{p,i}\omega_{n,p,i}s + \omega_{n,p,i}^2)}$$

with all $z_j, p_i, \zeta_{z,j}, \omega_{n,z,j}, \zeta_{p,i}$ and $\omega_{n,p,i}$ real, we can determine the logarithmic magnitude graph to be:

$$|H(s)|_{\text{dB}} = 20\log_{10}|K| + \sum_{j=1}^{m_1} 20\log_{10}\left|s - z_j\right| + \sum_{j=1}^{m_2} 20\log_{10}\left|s^2 + 2\zeta_{z,j}\omega_{n,z,j}s + \omega_{n,z,j}^2\right|$$
$$- \sum_{i=1}^{n_1} 20\log_{10}|s - p_i| - \sum_{i=1}^{n_2} 20\log_{10}\left|s^2 + 2\zeta_{p,i}\omega_{n,p,i}s + \omega_{n,p,i}^2\right|$$

in which we substitute $s = \text{j}\omega$. Similarly, we can write for the phase graph:

$$\angle H(s) = \angle K + \sum_{j=1}^{m_1} \angle\left(s - z_j\right) + \sum_{j=1}^{m_2} \angle\left(s^2 + 2\zeta_{z,j}\omega_{n,z,j}s + \omega_{n,z,j}^2\right)$$
$$- \sum_{i=1}^{n_1} \angle\left(s - p_i\right) - \sum_{i=1}^{n_2} \angle\left(s^2 + 2\zeta_{p,i}\omega_{n,p,i}s + \omega_{n,p,i}^2\right)$$

The conclusion is that both graphs can be compsed by calculating a sum of contributions. It is as if each of these contributions is a Lego brick. If we know all the basic Lego bricks in detail, we can analyze any LTI system, because it is just a sum of those bricks.

Let's start by considering an example, an arbitrary rational transfer function that we factorized.

$$H(s) = \frac{100s^3 - 100s^2 - 100s - 1500}{s^4 + 15s^3 + 84s^2 + 170s} = 100\frac{(s^2 + 2s + 5)(s - 3)}{s(s + 5)(s^2 + 10s + 34)} \tag{8.15}$$

2nd-degree factor in numerator
constant factor
1st-degree factor in numerator
factor s in denominator
2nd-degree factor in denominator
1st-degree factor in denominator

So, in fact, the only building blocks we need to study as 'Lego bricks' are:

- a constant factor
- a factor s in the numerator (not present in the example above) and the denominator
- a first-degree factor in the numerator and the denominator
- a second-degree factor in the numerator and the denominator

8.5.2 Making the bricks 'standard size'

Before we start, we need to make sure that the different factors as we listed them above are always in the same form. We call this *writing the transfer function in its normal form*. The principle is very simple: make sure that every polynomial has a low-frequency gain of 1. Of course, this does not work for factors s or powers of s. They are considered to be normalized as they are.

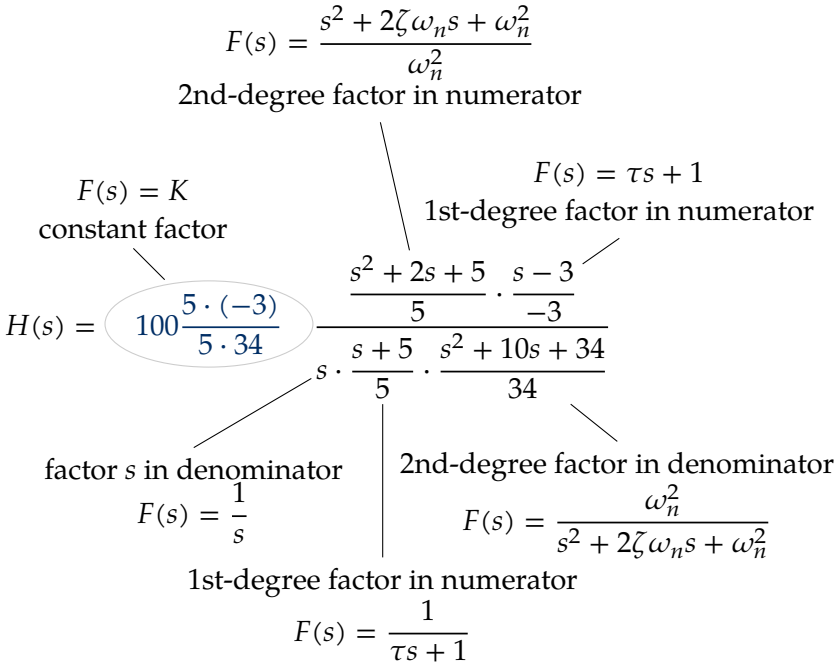
Considering the factors in the example we discussed earlier (See (8.5.1)), the following table shows how to write each polynomial appearing in the equation. The normalization to get the LF-gain equal to 1, asks for a correction factor to compensate the rewrite.

		Rewrite		**Correction factor**
$s+5$	$\rightarrow$	$\frac{s+5}{5}$	$\Rightarrow$	5
$s-3$	$\rightarrow$	$\frac{s-3}{-3}$	$\Rightarrow$	-3
s^2+2s+5	$\rightarrow$	$\frac{s^2+2s+5}{5}$	$\Rightarrow$	5
$s^2+10s+34$	$\rightarrow$	$\frac{s^2+10s+34}{34}$	$\Rightarrow$	34

How do these normalizations and the correction factors fit in the original example? Well, here they are. Note that we gathered the correction factors up front.

$$H(s) = 100\frac{(s^2+2s+5)(s-3)}{s(s+5)(s^2+10s+34)} = \underbrace{100\frac{5\cdot(-3)}{5\cdot 34}}_{\text{correction factors}} \underbrace{\frac{\frac{s^2+2s+5}{5}\cdot\frac{s-3}{-3}}{s\cdot\frac{s+5}{5}\cdot\frac{s^2+10s+34}{34}}}_{\text{normalized factors}}$$

The schematic below shows us the normal forms of each and every factor:



Now that the standard blocks are clear, we can study them one by one.

8.5.3 Analyzing the standard-size bricks

Constant factor

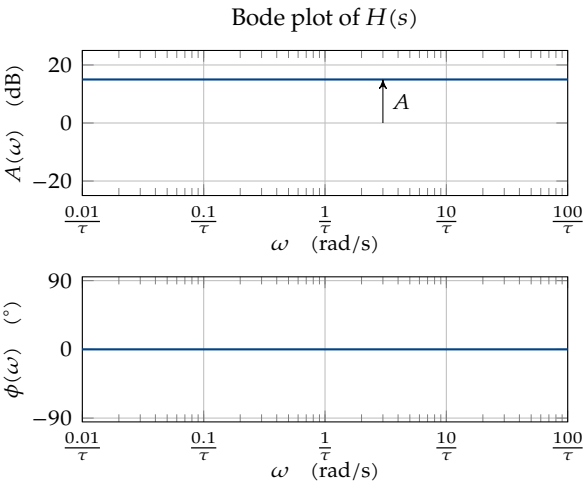
Normalized transfer function:

$$H(s) = K$$

shifts the function up ($K > 1$) or down ($K < 1$) with an amount

$$A = 20 \log_{10} |K|$$

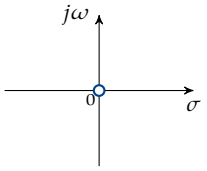
If K is negative, an extra phase change of $\pm 180°$ occurs.



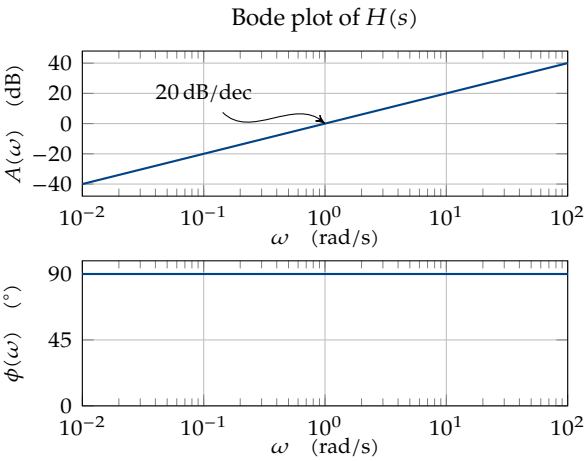
Factor s in the numerator (a differentiator)

Normalized transfer function:

$$H(s) = s$$



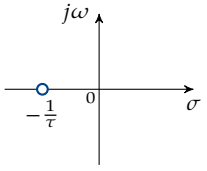
Note: the ('silent') factor 1 before s has the unit 's'.



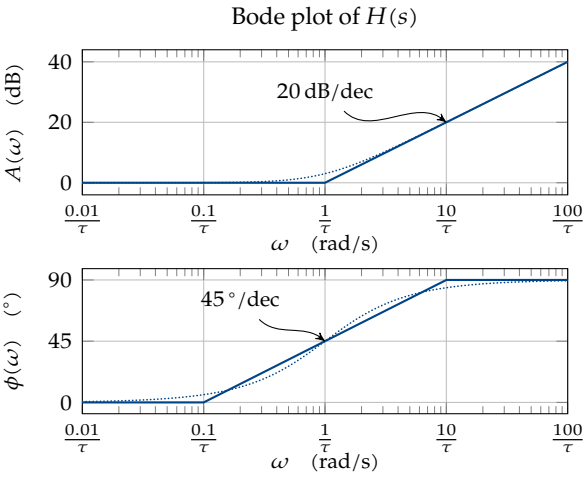
First-degree factor in the numerator (zero in LHP)

Normalized transfer function ($\tau > 0$):

$$H(s) = \tau s + 1$$



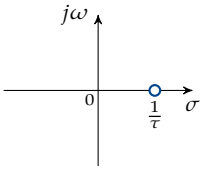
For low frequencies it behaves as a constant factor, for high frequencies as a factor s in the numerator.
Note that for a LHP zero the phase increases together with the magnitude.



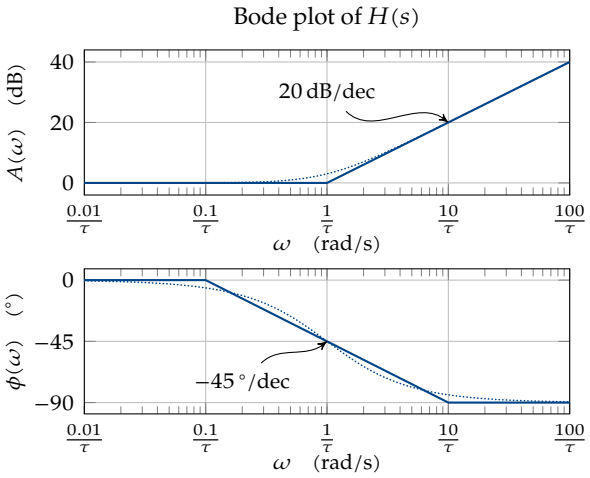
First-degree factor in the numerator (zero in RHP)

Normalized transfer function ($\tau > 0$):

$$H(s) = -\tau s + 1$$



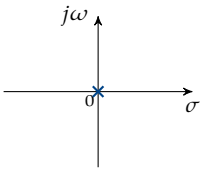
For low frequencies it behaves as a constant factor, for high frequencies as a factor s in the numerator and an extra $\pm 180°$ phase change. Note that for a RHP zero the phase decreases opposite to the magnitude.



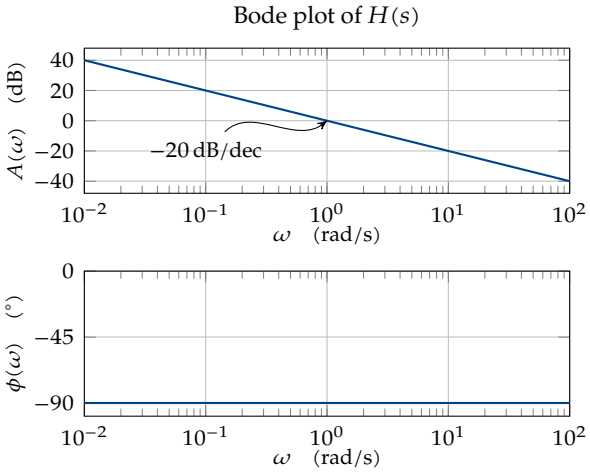
Factor s in the denominator (an integrator)

Normalized transfer function:

$$H(s) = \frac{1}{s}$$



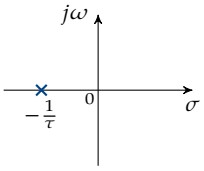
Note: the ('silent') factor 1 before s has the unit 's'.



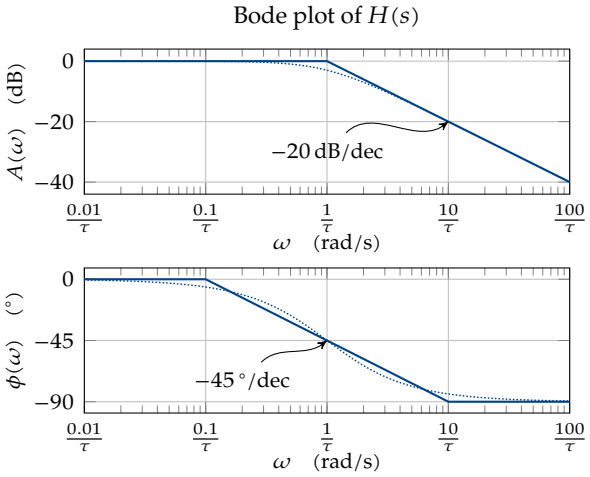
First-degree factor in the denominator (pole in LHP)

Normalized transfer function ($\tau > 0$):

$$H(s) = \frac{1}{\tau s + 1}$$



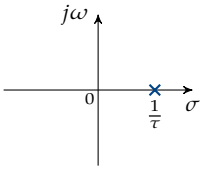
For low frequencies, it behaves as a constant factor, for high frequencies as a factor s in the denominator.
Note that for a LHP pole, the phase decreases together with the magnitude.



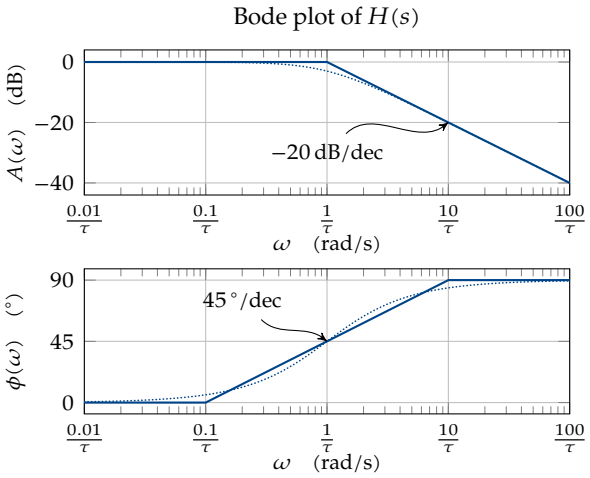
First-degree factor in the denominator (pole in RHP)

Normalized transfer function ($\tau > 0$):

$$H(s) = \frac{1}{-\tau s + 1}$$



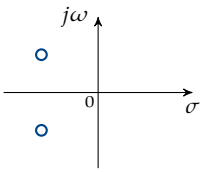
Note: this corresponds to an unstable block! For low frequencies, it behaves as a constant factor, for high frequencies as a factor s in the denominator with an extra $\pm 180°$ phase change.
Note that for a RHP pole, the phase increases opposite to the magnitude.



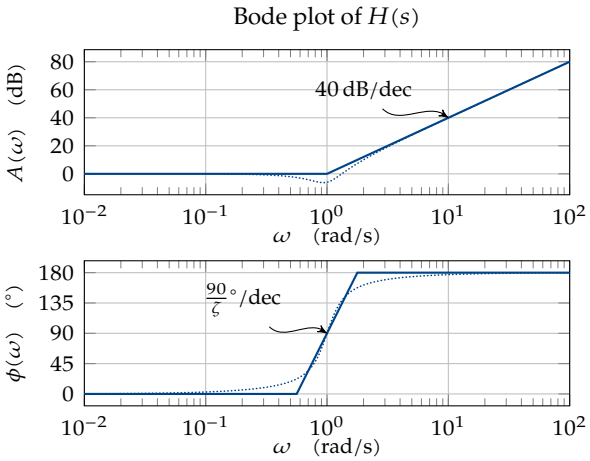
Second-degree factor in the numerator (zeros in LHP)

Normalized transfer function ($0 \leq \zeta \leq 1$):

$$H(s) = \frac{s^2 + 2\zeta\omega_n s + \omega_n^2}{\omega_n^2}$$



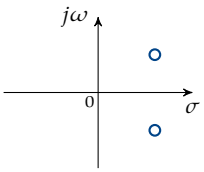
Note that for LHP zeros, the phase increases together with to the magnitude. Depending on the value of ζ, a resonance peak may occur.



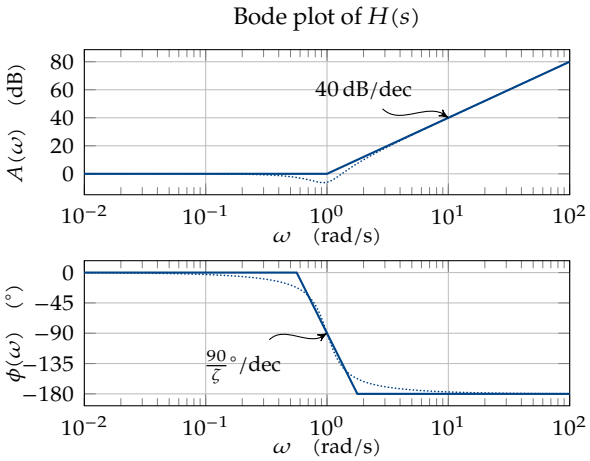
Second-degree factor in the numerator (zeros in RHP)

Normalized transfer function ($0 \geq \zeta \geq -1$):

$$H(s) = \frac{s^2 + 2\zeta\omega_n s + \omega_n^2}{\omega_n^2}$$



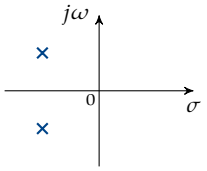
Note that for RHP zeros, the phase decreases opposite to the magnitude. Depending on the value of ζ, a resonance peak may occur.



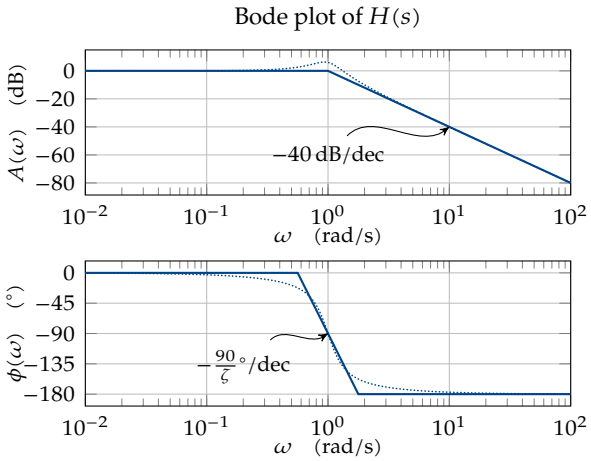
Second-degree factor in the denominator (poles in LHP)

Normalized transfer function ($0 \leq \zeta \leq 1$):

$$H(s) = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$$



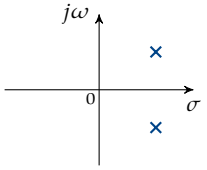
Note that for LHP poles, the phase decreases together with to the magnitude. Depending on the value of ζ, a resonance peak may occur.



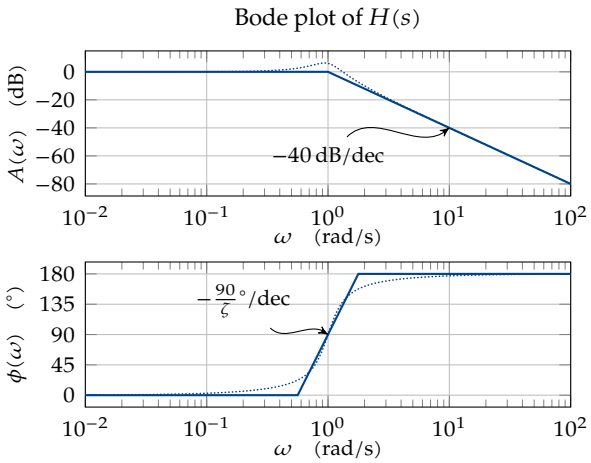
Second-degree factor in the denominator (poles in RHP)

Normalized transfer function ($0 \geq \zeta \geq -1$):

$$H(s) = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$$



Note that for RHP poles, the phase increases opposite to the magnitude. Depending on the value of ζ, a resonance peak may occur.



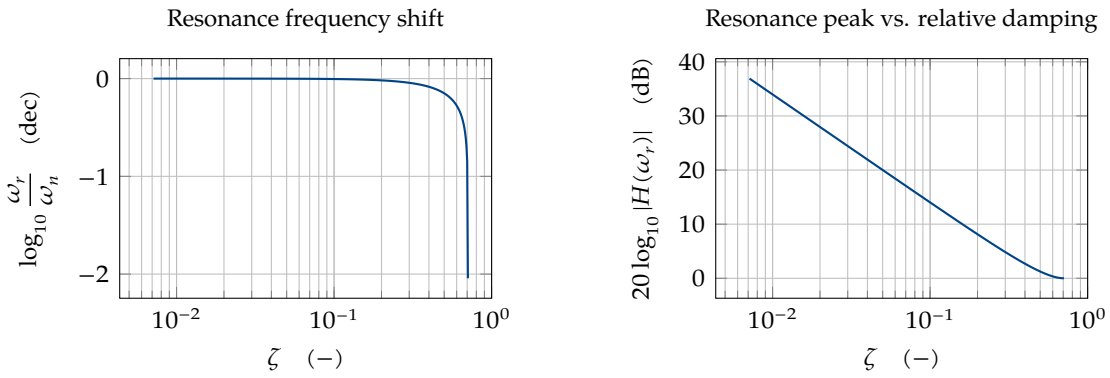
Remarks

- Every value of $1/\tau$ or ω_n is called a *break frequency*, as the magnitude curve that usually is a straight line is 'broken' there.
- A useful rule of thumb is that for LHP poles and zeros the phase has the same trend as the magnitude curve. For RHP poles and zeros, the opposite is true.

8.5.4 Resonance correction

Regarding resonance (in case of complex conjugated poles or zeros), we usually correct the graph by hand, estimating the resonance peak shift and the height of the peak for a normalized second-order system parameterized by ω_n and ζ. Whether it consists of poles or zeros, the following relationship is valid:

$$\omega_r = \omega_n\sqrt{1 - 2\zeta^2} \qquad |H(\omega_r)| = \frac{1}{2\zeta\sqrt{1 - \zeta^2}}$$



(a) Resonance frequency shift

(b) Resonance peak

Figure 8.9: Frequency resonance parameters as a function of the relative damping factor ζ

These expressions have been cast into convenient graphs that one can use when drawing Bode plots (see Figure 8.9).

8.5.5 Procedure

Now that we know all the building blocks, let's investigate how we can proceed to draw an asymptotic Bode diagram. The following procedure can be used:

Asymptotic Bode Plot Drawing Procedure

1. Compose the transfer function $H(s)$.
2. Write $H(s)$ in standard form.
3. Determine all individual building blocks:
   3.1. write down the break frequency related to the block
   3.2. make a tiny sketch of the magnitude and phase effect of the block
4. Order the break frequencies from small to large, and treat the blocks in that order; the frequency range of your Bode plot at least should cover all these break frequencies.
5. Draw the magnitude graph:
   - with each zero the slope increases with 20 dB/dec
   - with each pole the slope decreases with 20 dB/dec
   - for complex poles/zeros: 40 dB/dec (increase for zero, decrease for pole)
     $\omega_r = \omega_n\sqrt{1-2\zeta^2}$ $\quad |H(j\omega_r)| = \frac{1}{2\zeta\sqrt{1-\zeta^2}}$
6. Draw the phase graph: Phase shift $|\Delta\phi|$ of 90° per zero/pole
   - in the LHP: same trend as magnitude graph (zero = increase, pole = decrease)
   - in the RHP: opposite trend as magnitude graph (zero = decrease, pole = increase)
   - For complex poles/zeros: slope of $\frac{90}{\zeta}$°/dec (with similar directions)

8.6 Examples

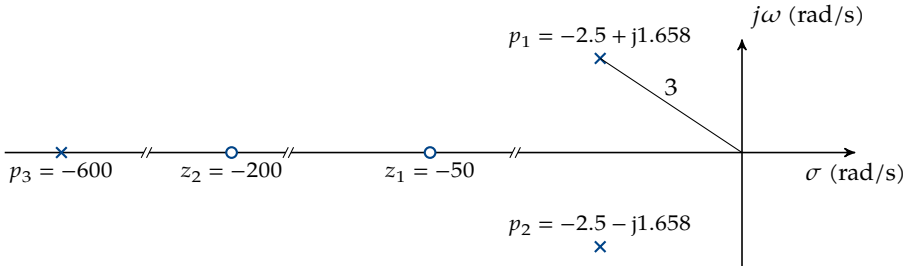
Let's apply this procedure to a few examples.

8.6.1 Example 1

Consider the following transfer function:

$$H(s) = \frac{13(s+50)(s+200)}{(s^2+5s+9)(s+600)}$$

Step 1: Compose the transfer function
This step has already been taken because the transfer function has been given. Though not necessary, let's draw the poles and zeros of this transfer function in pole-zero plot. Warning: the drawing is not to scale. The openings in the horizontal axis (delimited by the slanted lines) indicate this fact.



Step 2. Write $H(s)$ in standard form

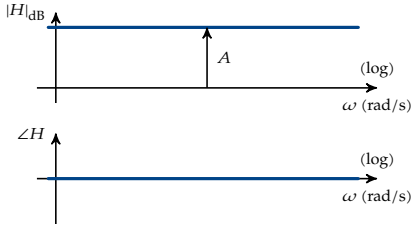
$$H(s) = \frac{13(s+50)(s+200)}{(s^2+5s+9)(s+600)} = \underbrace{\frac{13\cdot 50\cdot 200}{9\cdot 600}}_{K} \cdot \frac{s+50}{50} \cdot \frac{s+200}{200} \cdot \frac{9}{s^2+5s+9} \cdot \frac{600}{s+600}$$

Step 3. Determine all individual standard blocks
Conclusion, we have five standard blocks. Let's go over them. We'll make a quick sketch for each of them (on the right).

1. Constant factor

$$K = \frac{13\cdot 50\cdot 200}{9\cdot 600} = 24.074 \sim A = 27.63\,\text{dB}$$

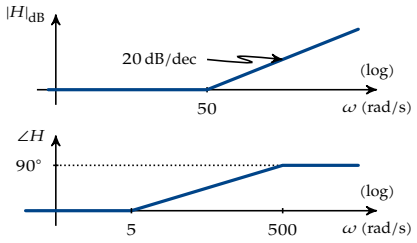
The factor is positive, so the phase change is zero.



2. Real zero

$$H(s) = \frac{s+50}{50}$$

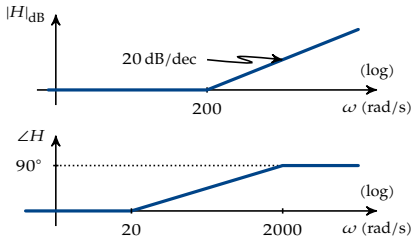
The zero is in the LHP, therefore the phase goes up together with the magnitude.



3. Real zero

$$H(s) = \frac{s+200}{200}$$

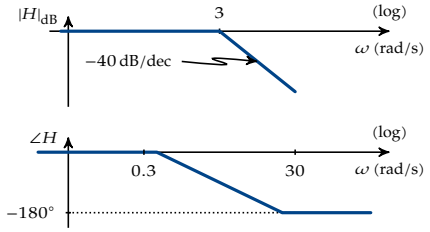
The zero is in the LHP, therefore the phase goes up together with the magnitude.



4. Complex conjugated pole pair

$$H(s) = \frac{9}{s^2+5s+9}$$

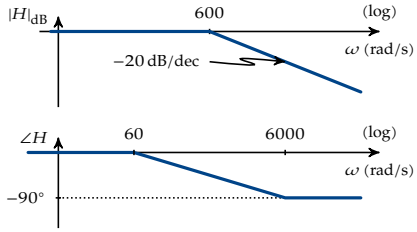
We can easily deduce: $\omega_n = 3$ and $\zeta = 0.8333$.
As $\zeta > 1/\sqrt{2}$ we won't have any resonance. The poles are in the LHP, therefore the phase goes down together with the magnitude. The phase transition takes 2ζ decades.



5. Real pole

$$H(s) = \frac{600}{s+600}$$

The pole is in the LHP, therefore the phase goes down together with the magnitude.



Step 4. Order the break frequencies from small to large

We found (see above): 3, 50, 200, 600 (all in rad/s). In general you round down the lowest value to the nearest decade boundary (a power of 10) and the highest value of the next decade. Then you extend the domain at both sides with an extra decade. In this case (assuming all values have units radians per seconde in the equations below):

$$\omega_{\text{min}} = 3 \xrightarrow{\text{round down}} 1 \xrightarrow{\text{extra decade}} 10^{-1}$$

$$\omega_{\text{max}} = 600 \xrightarrow{\text{round up}} 10^{3} \xrightarrow{\text{extra decade}} 10^{4}$$

We will therefore generate a plot from 0.1 rad/s to 1×10^4 rad/s.

Steps 5 and 6. Draw the magnitude and phase plot

There are two routes you can take to complete the drawing. They both start with plotting all the basic building blocks on the graph. This has been done in dashed red on Figure 8.10.

Then, you can take one of the two possible following approaches. Usually you have a preference for one of the two methods. Find out your favorite method and train yourself to bits on it.

method 1 — start from the low frequencies, calculate the slope of the magnitude graph and the starting phase of the phase graph and then for every break frequency you encounter, calculate the new slope and build the graph incrementally.

method 2 — for every break frequency determine the contributions of the building blocks and put this sum as a dot on the graph. Then connect all the dots.

This has been done in blue on Figure 8.10. For comparison's sake the exact curve has also been indicated in dotted blue line.

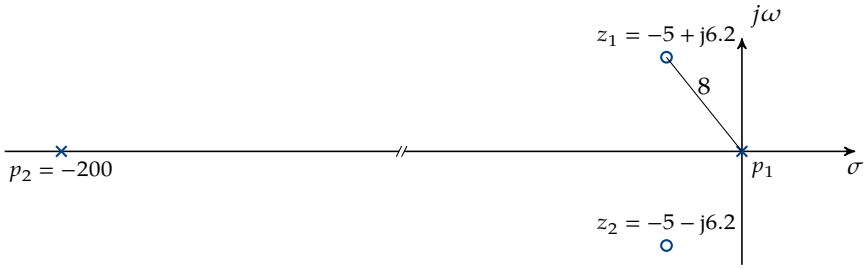
8.6.2 Example 2

Consider the following transfer function:

$$H(s) = \frac{5(s^2 + 10s + 64)}{s(s + 200)}$$

Step 1: Compose the transfer function

This step has already been taken because the transfer function has been given. Though not necessary, let's draw the poles and zeros of this transfer function in pole-zero plot. Warning: the drawing is not to scale.



Step 2. Write $H(s)$ in standard form

$$H(s) = \frac{5(s^2 + 10s + 64)}{s(s + 200)} = \frac{5 \cdot 64}{200 \cdot s} \cdot \frac{200}{s + 200} \frac{s^2 + 10s + 64}{64}$$

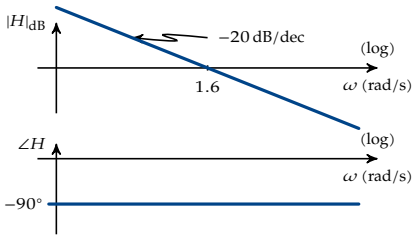
Step 3. Determine all individual standard blocks

Conclusion, we have three standard blocks. We could have considered a separate gain block and a separate integrator. However it makes sense to combine them, as this will put all individual graphs closer together.

Let's go over them. We'll make a quick sketch for each of them (on the right).

1. Integrator

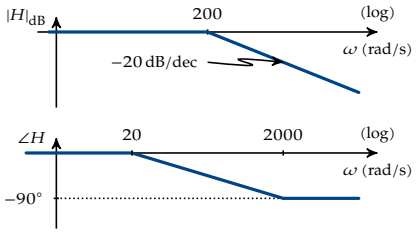
$$K = \frac{5 \cdot 64}{200s} = \frac{1.6}{s}$$



This gives a downward magnitude line with slope −20 dB/dec, that crosses the zero-db line when $\omega = 1.6$ rad/s. The phase is −90°.

2. Real pole

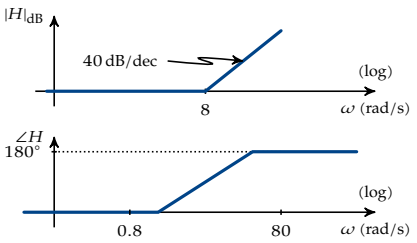
$$H(s) = \frac{200}{s + 200}$$



The pole is in the LHP, therefore the phase goes down together with the magnitude.

3. Complex conjugated zero pair

$$K = \frac{s^2 + 10s + 64}{64}$$



We can easily deduce: $\omega_n = 8$ and $\zeta = 0.625$.
The zeros are in the LHP, therefore the phase goes up together with the magnitude. The phase transition takes 2ζ decades. As $\zeta < 1/\sqrt{2}$ we will have resonance. From the graphs of Figure 8.9, we can read the following approximate values:

$$\log_{10} \frac{\omega_r}{\omega_n} \approx -0.2 \qquad 20 \log_{10} |H(\omega_r)| \approx 0.5\,\text{dB}$$

which is only a very tiny resonance peak. After drawing the Bode plot, you can add a little bump of about 0.5 dB with a peak bottoming out 0.2 decades before ω_n.

Step 4. Order the break frequencies from small to large

We found (see above): 8 and 200 (all in rad/s). We will therefore generate a plot that includes at least these frequencies, let's say from 0.1 rad/s to 1×10^4 rad/s.

Steps 5 and 6. Draw the magnitude and phase plot

You can find the contributing blocks in dashed red and the final asymptotic curve in blue on Figure 8.11. The dotted curve represents the exact curve.

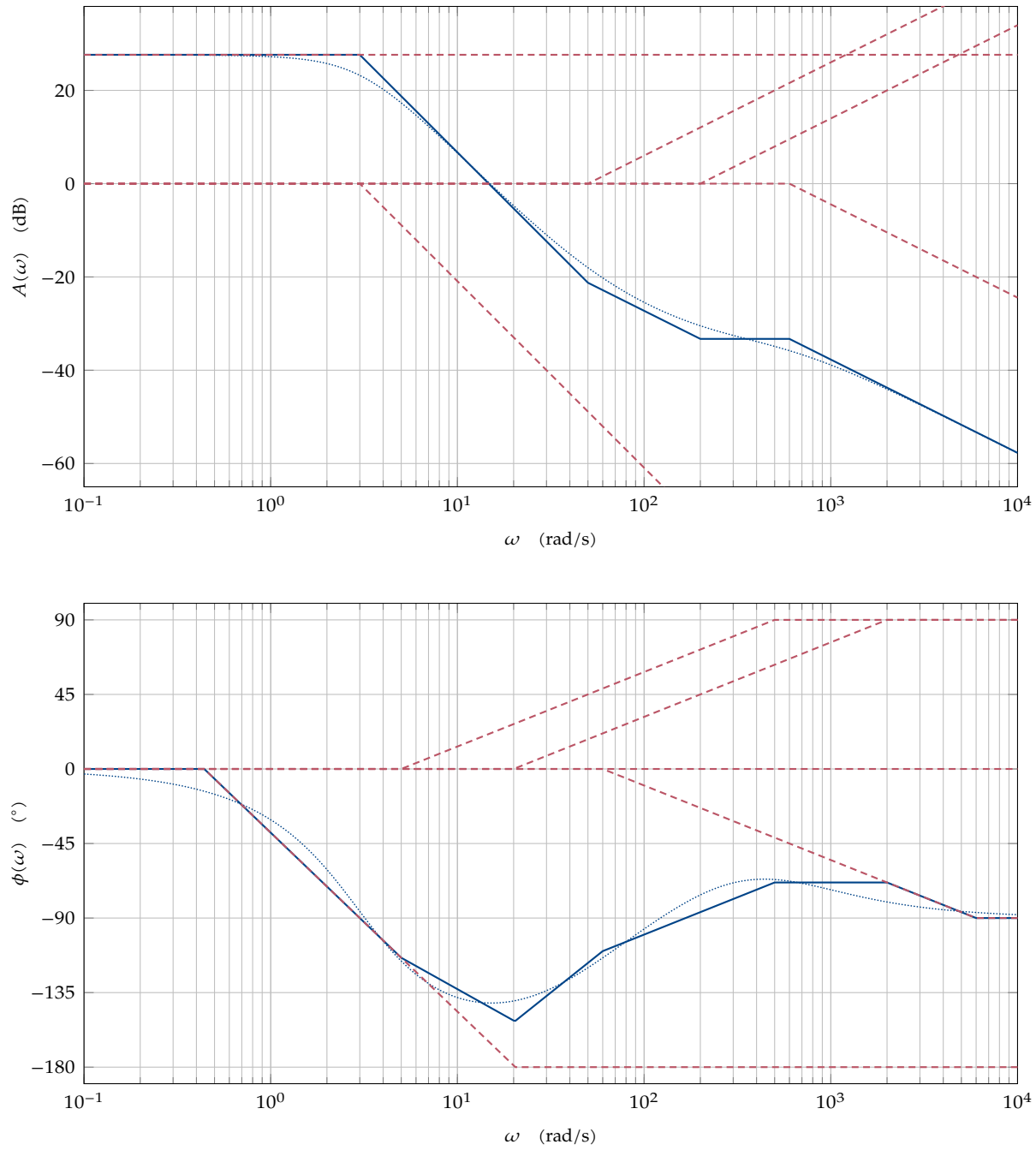


Figure 8.10: Bode plot of example 1 (see page 188); contributions of the standard blocks in dashed red lines, total asymptotic Bode plot in blue line, exact bode plot in dotted blue line

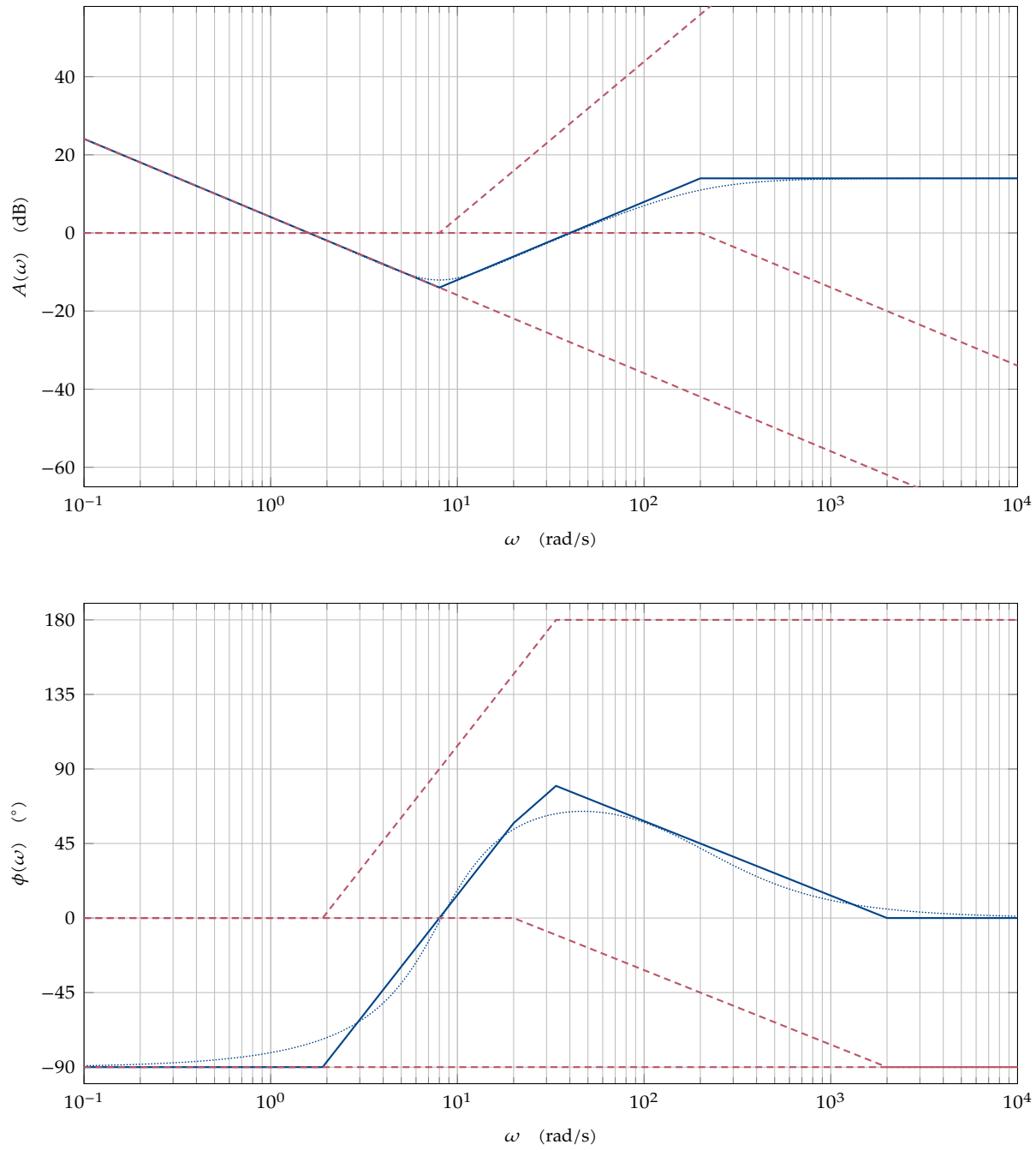


Figure 8.11: Bode plot of example 2 (see page 190); contributions of the standard blocks in dashed red lines, total asymptotic Bode plot in blue line, exact bode plot in dotted blue line

Exercises

Exercise 8.6.2-1:
Calculate the low and high frequency gain (magnitude and phase of the following transfer functions). Low means $\omega \to 0$ and high means $\omega \to +\infty$.

$$F(s) = \frac{10s^2 + 3s - 5}{s^2 + 5s + 10}$$
$$G(s) = \frac{s^3 + 5s^2 + 20}{s^2 + 100}$$
$$H(s) = \frac{3s^2 - 20s}{s^5 + 100s^3}$$

Exercise 8.6.2-2:
Draw the Bode diagram of the following transfer function and determine the frequency at the crossing with the 0 dB line and the phase at 10 rad/s:

$$H(s) = \frac{100}{s + 30}$$

Exercise 8.6.2-3:
Draw the Bode diagram of the following transfer function:

$$H(s) = 100\,\frac{s + 1}{s^2 + 110s + 1000}$$

Exercise 8.6.2-4:
Draw the Bode diagram of the following transfer function:

$$H(s) = 10\,\frac{s + 10}{s^2 + 3s}$$

Exercise 8.6.2-5:
Draw the Bode diagram of the following transfer function:

$$H(s) = -100\,\frac{s}{s^3 + 12s^2 + 21s + 10}$$

Exercise 8.6.2-6:
Draw the Bode diagram of the following transfer function:

$$H(s) = 10\,\frac{s + 300}{s^2 + 3s + 50}$$

Exercise 8.6.2-7:
Draw the Bode diagram of the following transfer function:

$$H(s) = 4\,\frac{s^2 + s + 25}{s^3 + 100s^2}$$

Exercise 8.6.2-8:
Draw the Bode diagram of the following transfer function and determine the phase at 10 rad/s:

$$H(s) = -20\frac{0.25s + 1}{s(0.01s - 1)}$$

Exercise 8.6.2-9:
Draw the Bode diagram of the following transfer function and determine the phase at 10 rad/s:

$$H(s) = 200\frac{s - 4}{s^2(s + 10)}$$

Exercise 8.6.2-10:
Draw the Bode diagram of the following transfer function and determine the phase at 20 rad/s:

$$H(s) = \frac{80s^2}{(s - 10)(s - 100)}$$

Exercise 8.6.2-11:
Draw the Bode diagram of the following transfer function and determine the frequency at the magnitude crossing with 0 dB:

$$H(s) = -1500\,\frac{s^2 + 10s + 64}{(s + 2)(s - 100)(s + 600)}$$

Exercise 8.6.2-12:
Draw the Bode diagram of the following transfer function and determine the frequency at the magnitude crossing with 0 dB:

$$H(s) = \frac{16000}{(s + 0.5)(s^2 - 6s + 900)}$$

Exercise 8.6.2-13:
Draw the Bode diagram of the following transfer function and determine the frequency at the magnitude crossing with 0 dB:

$$H(s) = \frac{3(1 + 2s)(1 + \frac{8}{9}s)}{(1 + 8s)(1 + \frac{2}{3}s)}$$

Exercise 8.6.2-14:
Draw the Bode diagram of the following transfer function and determine the frequency at the magnitude crossing with 0 dB:

$$H(s) = \frac{31(s + 50)(s + 200)}{(s^2 + 5s + 9)(s + 600)}$$

Exercise 8.6.2-15:
Draw the Bode diagram of the following transfer function and determine the frequency at the magnitude crossing with 0 dB:

$$H(s) = \frac{2s(0.002s + 1)}{(0.25s + 1)(0.1s + 1)}$$

8.7 Conclusion

In this chapter, we have studied analyzing LTI systems in the frequency domain, by calculating and drawing their spectra (as a Bode plot). Once again, we have seen how we can use the Laplace transform as one of the blades of our Swiss army knife to solve any frequency-domain question about LTI-systems. Though drawing Bode diagrams might seem like an ancient art, make sure you master the trade, as a clear view on how poles and zeros result in a specific spectrum will help you when studying control theory in one of your next courses.

Appendix A

The Decibel

The *decibel* (dB) is a quite common (pseudo) unit. However, it is frequently misunderstood. This is partly due to some fine nuance in the defining equations and partly due to the fact that the concept is heavily (ab)used in multiple engineering domains.

Let's attempt to set things straight.

A.1 Definition

A.1.1 Power ratio

The starting point for the decibel is the concept of *power ratio*, i.e. the ratio of a power value P_2 with respect to a value P_1.[1] Formally:

$$R_p = \frac{P_2}{P_1}$$

For reasons that will become clear below, it makes sense to take the logarithm of this ratio, leading to:

$$L_p = \log_{10} R_p = \log_{10} \frac{P_2}{P_1} \qquad \text{(B)}$$

The quantity L_p is assigned the unit Bel (B) in honour of Graham Bell. It is dimensionless and only indicates that the value expresses a logarithmic power ratio.

For some reason (probably to be able to use integer numbers in most common cases), it was then decided to use the deciBel (dB) instead of the unit Bel. In physical equations we normally don't introduce constants for unit conversions[2], but in this case, it has become quite common to write:

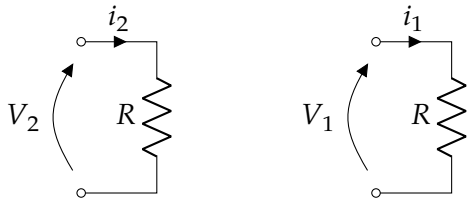
$$L_p = 10 \log_{10} \frac{P_2}{P_1} \qquad \text{(dB)}$$

Even more, from now on, we will stop writing explicitly the unit dB after the equation, but silently assume that everyone knows this.

In theory, that's all there is to it. However, in practice, people got used to calculating L_p starting from field quantities (e.g., $\mathrm{V}, \mathrm{V/m}, \mathrm{Pa}$) instead of power quantities. Instead of making the

[1] Instead of power quantities, also intensity quantities qualify, e.g. $\mathrm{W/m^2}$ or $\mathrm{lm/sr}$.
[2] We assume SI units are always used.

intermediate step, converting the field quantities to power (or intensity) quantities, a formula was developed that avoids that intermediate step.
Consider as an example a voltage V_2 applied to a resistor R, in comparison to a voltage V_1 applied to the same resistor R.



The power consumed by the resistors is V_2^2/R and V_1^2/R respectively. Therefore, the power ratio L_p can be written as:

$$\begin{aligned} L_p &= 10\log_{10}\frac{P_2}{P_1} = 10\log_{10}\frac{V_2^2/R}{V_1^2/R} = 10\log_{10}\left(\frac{V_2}{V_1}\right)^2 \\ &= 20\log_{10}\frac{V_2}{V_1} \end{aligned}$$

The last step can be taken knowing that $\log_{10} a^b = b\log_{10} a$.

This convenience equation, though simple, has led to a lot of misunderstanding, ranging from doubt when to use the factor 10 versus the factor 20, to speaking about dB-power and dB-voltage.

The following is the only correct statement: the value of L_p represents *always* a power ratio, but it can be calculated starting from a ratio of powers (using a factor of 10), or starting from a ratio of field quantities (using a factor of 20).

A.1.2 Absolute power quantity

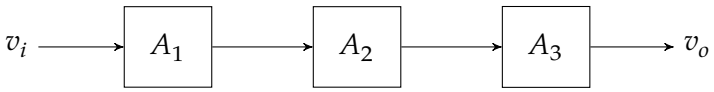
The unit decibel (dB) is also often used to express absolute quantities by comparing them to a fictive reference value.

A.2 Use

Now, what's the point of using this logarithmic value instead of just a simple ratio?

Simplification of calculations

A convincing use case is when one considers a chain of signal amplification blocks amplifying (or attenuating) an input voltage v_i to an output voltage v_o:



This could be a communications channel, with a transmit amplifier, a channel loss and a reception amplifier.
Calculating v_o as a function of v_i is simple:

$$v_o = A_1 \cdot A_2 \cdot A_3 \cdot v_i$$

Referring the voltages to a common reference voltage v_r (i.e. dividing both sides of the equation above by v_r) leads to:

$$\frac{v_o}{v_r} = A_1 \cdot A_2 \cdot A_3 \cdot \frac{v_i}{v_r}$$

Then, let's take the 20-fold logarithm of both sides and elaborate:

$$\begin{aligned} 20\log_{10}\left(\frac{v_o}{v_r}\right) &= 20\log_{10}\left(A_1 \cdot A_2 \cdot A_3 \cdot \frac{v_i}{v_r}\right) \\ &\downarrow \quad \log_{10}(a \cdot b) = \log_{10} a + \log_{10} b \\ L_{P_o} &= 20\log_{10} A_1 + 20\log_{10} A_2 + 20\log_{10} A_3 + L_{P_i} \end{aligned}$$

Or in a sloppy, but very engineeringish notation:

$$v_{i,\mathrm{dB}} = A_{1,\mathrm{dB}} + A_{2,\mathrm{dB}} + A_{3,\mathrm{dB}} + v_{o,\mathrm{dB}}$$

This means that the *complex multiplication* has been replaced by a more *simple addition*.

Logarithmic quantities

A second compelling reason to use logarithmic power ratios is that one of our human senses, hearing, is logarithmic in nature. It requires doubling the sound power to increase the perceived loudness by a fixed amount.

A.3 Examples

The decibel is used in many physics/engineering domains, each domain having its own reference power. The table below, lists some typical cases.

Unit	Reference	Use
dBV	1 V	voltages regardless of impedance.
dBmV	1 mV	voltages regardless of impedance.
dBµV	1 µV	voltages regardless of impedance.
dBm	1 mW	signal power w.r.t.
		- 600 Ω in audio electronics
		- 50 Ω in RF electronics
dBSPL[3]	20 µPa	sound pressure level in air
	1 µPa	sound pressure level in liquids
dBSIL	1×10^{-12} W/m^2	sound intensity level
dBSWL	1×10^{-12} W	sound power level
dBc	carrier power	noise or sideband power relative to the carrier

The goal of the table above was not to give an exhaustive overview, but to show you how heavily it is used in different domains. So, you'd better check the exact reference of the specific unit at hand, before you trust any readings in it.

A.4 Rules of thumb for calculating with dBs

The following estimation table shows you how to quickly determine the ratio that corresponds to a certain dB-level.

L_p	Power ratio	Field ratio
3 dB	2	$\sqrt{2}$
6 dB	4	2
10 dB	10	π
20 dB	100	10

Of course, the value of π is not correct, but it comes darned close.

As an example, let's try to estimate to what power ratio and what field ratio a decibel-level $L_p = 36\,\text{dB}$ corresponds.

$$\begin{aligned} L_p = 36\,\text{dB} &= (20 + 10 + 6)\text{dB} \\ &\Downarrow \\ \frac{P_2}{P_1} &= 100 \cdot 10 \cdot 4 = 4000 \\ \frac{V_2}{V_1} &= 10 \cdot \pi \cdot 2 = 62.8 \end{aligned}$$

If you need more accurate figures (e.g., with an accuracy of 1 dB or less), then you'd better use a calculator.

A.5 Conclusion

Don't be afraid when meeting decibels as an engineer, but make sure to exactly know what the reference value is.

Appendix B

Representing impulse trains: the Sha-function

Impulse trains are such a fundamental concept in digital signal processing that Ronald Bracewell, a former professor of electrical engineering of Stanford University, proposed a new notation for them.

The notation is very compact and makes use of a letter of the Cyrillic script: the Sha, written as ш and mostly used as a capital Ш.

The definition Bracewell proposed, combines the letter Sha (the peaks representing the individual pulses) with a subscript, indicating the spacing in between the peaks. In the continuous domain, with t the domain variable, this leads to:

$$\text{Ш}_T(t) = \sum_{i=-\infty}^{+\infty} \delta(t - iT)$$

It may also be used in the discrete domain (e.g. as a *resampling function*). With n the domain variable, this leads to:

$$\text{Ш}_N[n] = \sum_{i=-\infty}^{+\infty} \delta[n - iN]$$

The advantage of the Sha-notation is that it keeps focus on the important details (the spacing of the Dirac impulses), rather than on the cumbersome details of the summation and its indexing.

Bibliography

[BtMvdBJvdV93] R.J. Beerends, H.G. ter Morsche, van den Berg J.C., and E.M. van de Vrie. *Fourier & Laplace transformaties (in Dutch)*. Educaboek, 1993.

[GR80] Robert A. Gabel and Richard A. Roberts. *Signals and Linear Systems*. John Wiley & Sons, 2nd edition, 1980.

[Hes09] João P. Hespanha. *Linear Systems Theory*. Princeton Press, Princeton, New Jersey, 2009.

[LRW14] Paul Levrie, Penne Rudi, and Daems Walter. *4-Wiskunde - cursusnota's (in Dutch)*. Universiteit Antwerpen, 2014.

[OWN96] Alan V. Oppenheim, Alan S. Willsky, and S. Hamid Nawab. *Signals and Systems, 2nd Edition*. Prentice Hall, 1996.

[Van18] Mark Vanpaemel. *Systeemanalyse in acht domeinen (in Dutch)*. Universiteit Antwerpen, 2018.

[vdW92] P.E.M van den Wyngaert. *Analoge en digitale ketens, een signaal- en systeembenadering (in Dutch)*. Katholieke Industriële Hogeschool Antwerpen, 1992.

Index

all cosine Fourier series, 67
all sine Fourier series, 67
all-pass systems
    frequency-domain analysis, 171
asymptotic Bode plots, 177

block diagrams, 29
Bode plots, 174–197
    asymptotic, 177
    examples, 189
    in practice, 180
    in theory, 174
    procedure, 187
    standard factors, 183
Bode, Hendrik, 176

control theory, 5
convolution, 41–50
    associativity, 46
    causal simplification, 48
    commutativity, 46
    definition, 44
    distributivity, 48
    example, 48
    graphical example, 45
    impulse decomposition, 42
    natural element, 48
    properties, 46
    why do we need it?, 41

Dirac impulse, 19

eigenfunctions, 91
even symmetry, 64
even/odd, 11
examples
    signals, *see* signals, examples
exponential Fourier series, 68

feasible system, 131
first-order systems, 132
    frequency-domain analysis, 163
    with a zero, 141
Fourier family diagram, 52, 127
Fourier series, *see* Fourier transforms, Fourier series
    conversion between exponential and trigonometric, 73
    exponential, 68
    physical interpretation, 72
    trigonometric, 61
Fourier transform, *see* Fourier transforms
Fourier transforms, 51–97
    Fourier family diagram, 52
    Fourier series, 61–75
        all cosine, 67
        all sine, 67
        conversion between exponential and trigonometric, 73
        exponential, 68
        physical interpretation, 72
        symmetry, 64
        trigonometric, 61
        trigonometric (alternative), 67
    Fourier transform, 75–95
        definition, 77
        derivations, 75
        examples, 80
        properties, 85
        properties — convolution, 87
        properties — frequency differentiation, 87
        properties — frequency scaling, 86
        properties — frequency shifting, 86
        properties — frequency translation, 86
        properties — multiplication, 87
        properties — Parseval's theorem, 87
        properties — time differentiation, 87
        properties — time scaling, 86
        properties — time shifting, 86
        properties — time translation, 86
        properties — time-frequency symmetry, 86
frequency resonance, 166
frequency spectrum, 74
frequency transfer function, 89
    calculating, 93
    measuring, 93
frequency-domain analysis, 155–197
    all-pass systems, 171
    Argand plane analysis, 160
    Bode plot analysis, 174
    composing (frequency) transfer functions, 157
    first-order systems, 163
    second-order systems, 166

generalized sinors, 18
Gibbs effect, 62

half wave symmetry, 65
higher-order systems
    with a zero, 142

impulse decomposition, 42

impulse response
first-order systems, 132
second-order systems, 136
impulse response description, 33
inner product, 55

Laplace transfer function, 110
Laplace transform, 99–127
applications, 118
solving differential equations, 118
solving linear electronic networks, 120
bilateral, 99
calculating
forward, 106
inverse, 107
common transform pairs, 109
definition, 100
derivation from Fourier transform, 99
motivation, 116
one-sided, 102
properties, 103
convolution, 105
final value theorem, 105
initial value theorem, 105
linearity, 104
multiplication, 105
periodic functions, 104
time-frequency differentiation, 104
time-frequency division, 104
time-frequency scaling, 104
time-frequency shifting, 104
region of convergence, 102
relationship with Fourier transform, 101
stability of signals, 113
two-sided, 99
unilateral, 102
linear differential equations, 29
logarithmic frequency axis, 176
logarithmic magnitude axis, 176
LTI systems
eigenfunctions, 91
sine waves, 92
sinors, 91
impulse response description, 33
in the Fourier domain, 88
linear differential equations, 29
transfer function, 110
LTI systems stability, 113

MIMO, 38
minimum phase systems, 171
MISO, 38

non-minimum phase systems, 171

odd symmetry, 64
operations, *see* signals, operations

Parseval's theorem, 74
periodicity, 11
phasors, 18

region of convergence, 102
resonance, 166

Sallen Key circuit, 122
scalar product, 55
second-order systems, 134
frequency-domain analysis, 166
with a zero, 141
signals, 9–22
approximating, 57
definition, 9–10
domain, 9
examples, 12
graph, 10
inner product, 55
length, 56
nonphysical, 16
operations, 10–11
physical, 12
properties, 11–12
even/odd, 11
periodicity, 11
quantitative, *see* signals, characteristics
time limit, 11
range, 9
scalar product, 55
SIMO, 38
sine waves, 92
sinors, 16, 91
SISO, 38
spectrum, 74
stability, 113
step response
first-order systems, 133
second-order systems, 137
symmetry
even, 64
half wave, 65
odd, 64
system
feasible, 131
system analysis, 4
system identification, 5
system synthesis, 5
systems, 23–40
classification, 24
IO count, 38
linear vs nonlinear, 25
LTI systems, 27
time-variant vs. time-invariant, 24
definition, 23
properties, 27–29

differentiation is LTI, 28
LTI systems maintain differentiation , 28
systems theory
an introduction, 1–7
classification, 6
definition, 4
models, 2
variants, 4

time delay, 71
time-domain analysis, 129–154
basic technique, 130
examples, 143
RC filter, 143
RLC circuit, 148
spark plug high-voltage generator, 149
first-order systems, 132
role of the transfer function, 130
second-order systems, 134
the effect of zeros, 139
time-domain effect of zeros, 139
time-limited signal, 11
time-unlimited signal, 11
transfer function, 110
transformations
basic idea, 37
trigonometric Fourier series, 61

vectors, 52
approximating, 57
in 2D, 52
scalar product for signals, 55
signals as, 53