# Towards a consistent interpretation of financial instruments in economic engineering using Noether's theorem.

E. B. Legrand

Literature Survey

TU Delft Delft University of Technology

Delft Center for Systems and Control

# Towards a consistent interpretation of financial instruments in economic engineering using Noether's theorem.

Literature Survey

E. B. Legrand

November 10, 2021

Faculty of Mechanical, Maritime and Materials Engineering (3mE) · Delft University of Technology

Rabobank

The work in this thesis was supported by Rabobank. Their cooperation is hereby gratefully acknowledged.

TU Delft Delft University of Technology

DCSC

# Abstract

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

# Table of Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

On the eight page of the October 16, 1929 edition of the New York Times, a small article headed [5]

> *"Fisher Sees Stocks Permanently High; Yale Economist Tells Purchasing Agents Increased Earnings Justify Rise."*

Responsible for this bold statement was Irving Fisher, a very prominent American economist who pioneered on the subject of monetary economics. Unfortunately for him, on Thursday the 29th of October — merely nine days later — the Dow Jones Industrial Average dropped by 11 percent, only to lower by another 13 percent on Monday, and then by 12 percent on Tuesday. The result was the Great Depression, an unprecendented financial crisis with dramatic repercussions: in the following three years, unemployment rose to 20 percent and industrial production almost halved [6].

Of course, this anecdote is not meant to discredit Irving Fischer as an economist, for he has been a major contributor to modern economic theory, but rather to illustrate that even specialists tend to misinterpret the current state of financial markets. This thesis does not aim to solve this issue right away, but to provide a new approach to the interpretation of the financial system; this may help in the understanding of this vital component of modern day society. This new approach is founded in the recent theoretic framework of economic engineering, developed by prof. em. M. Mendel [7].

**Economic engineering** An introduction to relevant aspects of economic engineering will be provided in chapter 2, but the core idea is to include economic systems in the traditional multi-domain modeling framework that (control) engineers use to study systems of widely varying nature (mechanical, electrical, hydraulic, etc.), possibly with the intention of develop a suitable control strategy. Of course, control strategies are vital for economic systems as well — although this practice would probably be called 'policy making' — either on small scale, such as firms managing their stock levels and revenue in a changing market, or on macroeconomic scale, e.g. a central bank deciding whether to lower the interest rate or not.

Economic engineering research has been concerned with financial markets before: an economic analogy to the usage of *action-angle coordinates* was exploited by Vos [8] to improve the models used for monetary policy. Kruimer [9] developed a macroeconomic model of the U.S. economy, where he included bond and equity markets as vital components of the economic machinery by means of the so-called *rotational analogy*. Apart from the fact that both interpretations are manifestly different ways to describe the same concept, they appear to be flawed in some fundamental ways. As such, the necessity arises to reconcile these methods and address their problems in order to find a unifying economic engineering approach to incorporate financial systems; primarily concerning markets for equity and debt and perhaps their related derivatives such as futures and options.

**Current approach** As mentioned, economic engineering currently recognizes two ways to deal with 'money problems'. Firstly, money is considered analogous to *action*, as (will be explained in chapter 2); action is a quantity that represents the integral of energy over time, with units [J s]. Action-angle coordinates are a choice of coordinates in the phase space that consist of (constant) action coordinates and dimensionless 'angles', indicating a *periodic* motion. Secondly, there is the rotational analogy; which is arguably a bit more flexible than the action-angle coordinates. Again, based on dimensional analysis, an analogy can be made between 'money' and angular momentum (in physics, angular momentum and action have the same dimension). Here the role of the angle represents a return or an accumulated interest, and 'arm' of the rotation the principal of the investment/debt instrument.

**Research goal** The core idea is that, instead of *ad hoc* applying the existing theories, to return to the fundamentals of economic engineering and its ties with analytical mechanics to build a more rigorous foundation for future work. Energy and its analogy to utility in economics play a crucial role here, because (i) it lies at the foundation of analytical mechanics as well and (ii) it allows to make the connection with the existing (not strictly financial) theory of economic engineering, just like the multi-domain modeling techniques in engineering are connected through the universal concept of energy and power. Within analytical mechanics, Noether's theorem describes precisely how mathematical symmetries in the general nature of the system expressed in terms of energy (encoded in a special state function called the *Lagrangian*) dictate *conservation laws* that the system must obey. Naturally, perfect conservation of energy and momentum is equally unlikely in both physics and economics, but one strives nevertheless to construct a 'platonic ideal', a conserved and isolated financial system to form the basis of the modeling framework.

The goal of this research can therefore be stated as follows:

**Research goal**

To develop a new, consistent, and unified framework to interpret debt and equity instruments in economic engineering, using the formal methods of analytical mechanics; and to provide an economic (or financial) interpretation to Noether's theorem.

**Structure of the literature study** That being said, it is clear that the scope of this research is *theoretical*, as its purpose is to expand and refine the current economic engineering framework.

Hence, this literature study contains an overview of some related subjects that will hopefully play a role in the development of the theory. Firstly, in chapter 2 both the foundations of analytical mechanics and economic engineering are discussed in parallel; for each theoretical subject the fundamental analogies between both fields are emphasized for they form the cornerstones of this research. It is also in this chapter that Noether's theorem and the action-angle coordinates are introduced. Subsequently, chapter 3 provides a basic overview of finance and actuarial science; especially understanding the latter is vital to see why the rotational analogy works. Because the rotational analogy concerns a *hyperbolic rotation* instead of a normal one, some additional

# Chapter 2

# Economic engineering with analytical mechanics

Economic engineering is built on analogies between (macro)economic theory and common engineering disciplines such as thermodynamics, circuit theory and (classical) mechanics[1]. Escpecially for the latter, a rich variety of useful analogues can be devised. There are two common classses of interpretations mechanics: *analytical mechanics* including the formulations by Joseph-Louis Lagrange and sir William Rowan Hamilton, and *vectorial mechanics*, better known as *Newtonian mechanics*. The former variant is usually preferred in the field of engineering whereas the analytical mechanics are due to their mathematical elegance powerful interpretative value. Likewise, the theory of economic engineering can be approached in two similar ways. Usually the 'Newtonian' approach is given the most attention, but in the case of this work the energy-based approach will prove to be more useful, which is why it is the starting point of this discussion instead.

Analytical mechanics, more specifically *Lagrangian* and *Hamiltonian* mechanics, are established around the definition of special state functions, respectively called the Lagrangian $\mathcal{L}$ and the Hamiltonian $\mathcal{H}$. As per usual, Lagrangian mechanics will be introduced first, for it has the most intuitive explanation. Then, a more formal approach allows to (Legendre) transform the discussion into one of Hamiltonian mechanics.

In this chapter, frequent analogies will be made between mechanics and economics; as such, the idea is to have the 'normal text' pertain mostly to the discussion of classical mechanics, and to provide the analogies in special 'boxes' like so:

**Example**

An analogy between economics and classical mechanics.

The reason for this particular choice of layout is twofold: first, it allows to make a sharp distinction between the older, extremely rigorous theory of classical mechanics and the novel

[1] as opposed to more recent theories in relativistic mechanics and quantum mechanics

approach of economic engineering: many ideas and propositions are still tentative (especially in the realm of classical mechanics) as the field of economic engineering matures. Secondly, it allows for easier reference as to not obscure the economic analogies (which are arguably the most important aspects of this chapter) with the highly theoretical discussion of analytical mechanics.

## 2-1 Economic engineering

The discipline of 'economic engineering' is a very new one. The theoretical foundations have been developed over the past years at the Delft Center of Systems and Control primarily by prof. em. dr. ir. Mendel, combined with the contributions of several theses that have been recently written about the subject. The purpose of economic engineering is to use tools from various engineering disciplines and physics to improve the predictive power of (macro)economic models. The core idea is to extend 'domain-neutral' modeling techniques such as bond graph modeling [4] that are built on analogies between mechanical, electrical, hydraulic, ... systems to economic systems as well. Hence, one attempts to give an economic interpretation to a generalized mass (I-element), generalized spring (C-element) and generalized damper (R-element). The idea is that this consistent engineering approach leads to actual *predictive* models that provide much richer insights than the 'stylized facts' from macroeconomy or, on the other end of the spectrum, the 'black-box' econometric models that have lost all interpretative value. Indeed, applied economic engineering pursues *grey-box* modeling instead, as is most common in traditional engineering models [9]. Some analogies between mechanical, electrical and economic systems are listed in table 2-1 to merely for the sake of illustration; a thorough motivation for each of them will be given in the following section. Table 2-1 provides some examples of the analogies that are used within economic engineering. In general, their meaning is not as specific as given here and this table applies only to the most elementary cases. The main theory behind economic engineering is outlined

**Table 2-1:** Some examples of the analogies that are used in the application of economic engineering. The theory behind bond-graph modeling defines a generalization of the mechanical concepts of displacement, velocity, momentum and effort and applies these to electrical, thermodynamic, hydraulic, ... systems too [4].

| General | Mechanical | Electrical | Economic |
|---|---|---|---|
| Displacement | Displacement | Charge | Stock level |
| Flow | Velocity | Current | Flow of goods |
| Momentum | Momentum | Flux linkage | Price |
| Effort | Force | Voltage | Economic want |
| I-element | Mass | Inductor | Market |
| C-element | Spring | Capacitor | Storage of goods |
| R-element | Damper | Resistor | Depreciation / Consumption |

by Mendel [7]. Some other notable results in the field have been achieved in recent years as well. Hutters and Mendel [10] demonstrated the application of Hamiltonian mechanics to dissipative systems as to include the mechanics of consumption in port-Hamiltonian systems.

A formal approach inspired by the theory of thermodynamics was used by Manders [11] to explain economic growth and productivity. Kruimer [9] and Van Ardenne [12] used economic engineering bond graph techniques to build extensive models for the U.S. economy and a 'generalized' firm (as to improve business valuation techniques).

## 2-2 Lagrangian mechanics

### 2-2-1 The configuration manifold

Central to the concept of Lagrangian mechanics is the so-called *configuration space* $M$, which is an $n$-dimensional manifold provided with some parameterization called *generalized coordinates* assembled in the vector $\boldsymbol{q}$. The configuration manifold may just be equal to $\mathbb{R}^n$, but in more interesting and realistic cases it is often some manifold embedded in $\mathbb{R}^n$. This is often the result of holonomic constraints, which are constraints that impose a restriction only on the configuration space, but not, for example, on the allowable velocities. By using the generalized coordinates, one can parameterize all the allowable positions of a system whose motions may occur in a higher-dimensional space with a smaller coordinate set (e.g. the two-dimensional motion of a pendulum can be expressed in a single coordinate due to the constraint imposed by the rigid link). The crucial insight here is that the constraint forces will never perform any work on the system[2]; as such, they act always orthogonally to the configuration manifold. This is why, provided that one is succesful in correctly describing this manifold with a suitable coordinate set, the constraint forces need not be taken into account: a major advantage over Newtonian mechanics.

Unfortunately, the pursuit of finding a set of these all-encompassing generalized coordinates is fruitless for some systems, for they cannot possibly be represented in this simple fashion. Luckily, for a certain class of constraints they may be included to constrict the motion of the system nevertheless by means of the *Lagrange multipliers*. They are applicable to holonomic constraints for which it is impractical or nonintuitive to take them into account directly in the parameterization of the configuration manifold, and a restricted class of nonholonomic constraints that can be written in the so-called *Pfaffian* form [13].

**Configuration manifold and the economic system**

In economic engineering, $\boldsymbol{q}$ has a similar intuitive notion, namely the stock values of various products, which denote the 'position' of some economic system. The name 'position' may be misleading when intuitively ported to the familiar three-dimensional space; the configuration space usually has less structure such as the absence of a metric or inner product.

Just like in mechanics, the simplest shape the configuration manifold $M$ can take is a simple $n$-dimensional vector space, but more sophisticated cases exist as well. As mentioned, a nontrivial configuration manifold is usually the result of holonomic constraints applied to the system; these constraints have their meaning in economics too. For example, when Lagrangian analysis is applied to the analysis of electrical

[2] This is known as *D'Alembert's principle*.

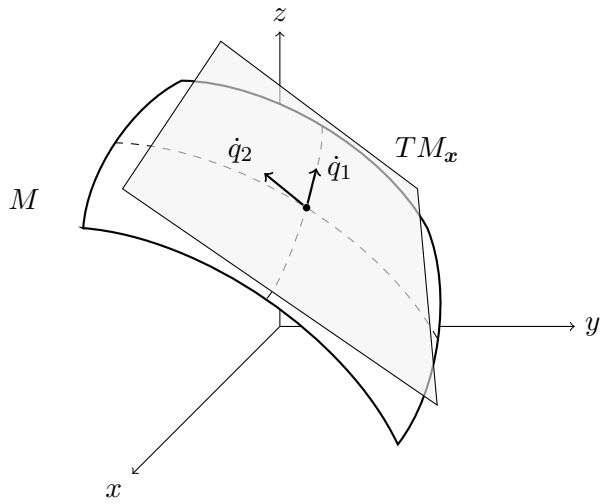



**Figure 2-1:** Schematic of a two-dimensional (configuration) manifold $M$ embedded in $\mathbb{R}^3$; the local generalized velocities associated with a point $\boldsymbol{x}$ are vectors that live in the tangent space $TM_{\boldsymbol{x}}$.

circuits, the generalized coordinates naturally reflect Kirchoff's current laws (what comes in must go out); as they provide simple constraints to the current in each part of the circuit. An intuitive extension can be made to economics, where the flow of goods is often subject to a Kirchoff-type law as well, especially when considering supply chains or transport.

### 2-2-2 Hamilton's principle of stationary action

The configuration manifold in described in the previous section is the first step in the Lagrangian approach, for it defines a single mathematical space containing all the possible motions of the system. For all intents and purposes, it usually pertains to the very nature of the system itself; e.g. the wiring of the electrical, how the mechanical parts are connected to each other or which goods are present in an economic system and whether their quantities are fundamentally related. Of course, the configuration manifold itself does not provide any information about the behavior of the system: this is where Hamilton's principle comes in — the second, crucial puzzle piece that makes Lagrangian mechanics work.

Hamilton's principle (also referred to as the principle of least or stationary action) concerns the existence of a special state function, the Lagrangian $\mathcal{L}$, determines the system's behavior [14]:

**Hamilton's principle**

Motions $\gamma : \mathbb{R} \to M$ of a mechanical system coincide with extremals of the action functional

$$S(\gamma) = \int_{t_1}^{t_2} \mathcal{L}\, \mathrm{d}t\,, \tag{2-1}$$

where $\mathcal{L}$ is the *Lagrangian function* of the system.

The Lagrangian $\mathcal{L}$ is a mapping from the *tangent bundle* $TM$ of the configuration manifold $M$, optionally paired with a time argument for time-varying problems to the reals,

$$\mathcal{L} : TM \times \mathbb{R}^+ \to \mathbb{R},$$

i.e. it takes a generalized position $\boldsymbol{q}$ and a generalized velocity $\dot{\boldsymbol{q}}$ (which live in the tangent space of $M$), and a time instance to some scalar values.

The aforementioned principle only is for now only helpful on a conceptual level, because it does not how to arrive at any solutions. For this, a branch of mathematics called the calculus of variations comes to aid, which is concerned with finding extremals of *functionals*[3], in this case $S$. A necessary condition for $S$ to attain an extremum is that

$$\delta S = 0,$$

where $\delta S$ is called the *first variation* of $S$. Just like a regular differential can be seen as an infinitesimal perturbation of a function value, the variation is a very small perturbation of a functional by means of a trajectory $h(t)$. The resulting perturbed functional can then in general be decomposed in a part that varies linearly with $h$, and a nonlinear part. The requirement for the extremal is that the *linear part vanishes for any* $h$ [14].

Landau and Lifshitz [15] use the the tools of the calculus of variations to show that the solution of eq. (2-1) is

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \mathcal{L}}{\partial \dot{\boldsymbol{q}}}\right) - \frac{\partial \mathcal{L}}{\partial \boldsymbol{q}} = 0.$$

This yields a total of $n$ second-order differential equations, or equivalently a system of $2n$ first-order equations. A special significance is assigned to the vector $\frac{\partial \mathcal{L}}{\partial \dot{\boldsymbol{q}}}$, defined as the *generalized momentum* $\boldsymbol{p}$.

**Hamilton's principle and utility maximization**

The generalization of Hamilton's principle to economics is not far-fetched; it is generally accepted that elementary economic agents act as to maximize their own utility; this is known as the *utility maximization problem*[a].

Thus, the formulation of economic motions as an extremal problem is quite straightforward. However, in order to descend from a purely philosophical debate a formalism that is actually useful, the Lagrangian must be assigned with a concrete meaning. As such, our aim is to translate the concept of energy to economics. In mechanics, the *energy* of the system is, bluntly speaking, its ability to perform *work*. For now, a purely

[3] A functional is a real-valued function on a vector space (of functions.)

mechanical interpretation is pursued, neglecting the connection between temperature and energy — the application of economic engineering and thermodynamics is desribed in the thesis of Manders [11]. The economic engineering interpretation of work is the fulfilling of wants. As such, the energy of an ecomomic system is its ability to fulfill wants. A natural dichotomy arises when viewing the economic system in terms of its configuration manifold and generalized 'velocities' (product flows), akin to the of forms energy in mechanics.

- *Kinetic energy* is related to the utility due to market (trading) activity, it therefore depends in the first place on the *flow of goods*; it is the surplus of the economic agent [16].
- *Potential energy* gives significance to utility obtained from the posession of goods; it must therefore depend on stock levels. One can see this as a sort of 'convenience yield' (a term used in futures pricing): the benefits of actually possessing the good.

A more rigorous definition of these concepts in economics will be given later in this section. The intuition behind the principle of stationary action can be found in Feynman [17]:

> *"[...] the solution is some kind of balance between trying to get more potential energy with the least amount of extra kinetic energy—trying to get the difference, kinetic minus the potential, as small as possible."*

Likewise, this can be restated as the basic least action principle in economic engineering:

> **Hamilton's principle in economic engineering**
> Economic agents try to to maximize their convenience yield (the utility of the goods they possess) by sacrificing as little economic surplus as possible.

As described by Feynman [17], this is not only a global trajectory, but also a local one at every (infinitesimal) piece of the trajectory: one can see this as a formal restatement of the rational behavior of economic agents, which is a crucial assumptions in may economic theories [16].

[a] It is important to realize that the term 'extremal' does not necessarily refer to a minimum, as is often incorrectly stated when explaining Hamilton's theorem — indeed, the least action principle is sometimes called the principle of *stationary* action, which is more in line with its mathematical definition.

### 2-2-3 Kinetic energy

In classical mechanics, the Lagrangian is defined by convention

$$\mathcal{L}(\boldsymbol{q}, \dot{\boldsymbol{q}}; t) = T^*(\boldsymbol{q}, \dot{\boldsymbol{q}}; t) - U(\boldsymbol{q}; t),$$

where $T^*$ is the kinetic co-energy[4] of the system and $U$ the potential energy. If $M$ is a Riemannian manifold and its Lagrangian has the aforementioned form, the system is called *natural* [14].

In the most general terms, the kinetic energy of the system is *defined* as a quadratic form on the tangent space of the configuration manifold. Assuming that $m$ is a Riemannian manifold (i.e. it is equipped with a Riemannian metric $\langle \boldsymbol{\xi}, \boldsymbol{\xi} \rangle$), one can define the kinetic co-energy as

$$T^* = \frac{m}{2}\langle \boldsymbol{v}, \boldsymbol{v} \rangle \qquad \boldsymbol{v} \in TM_{\boldsymbol{x}} \tag{2-2}$$

The usage of $T^*$ in the Lagrangian formulation is only useful if it is expressed in the generalized coordinates and generalized velocities; in general $T^*$ will be of the form

$$T^*(\boldsymbol{q}, \dot{\boldsymbol{q}}) = \frac{1}{2} m_{ij}(\boldsymbol{q}) \dot{q}_i \dot{q}_j,$$

observing the Einstein summation convention. This interpretation of kinetic co-energy as a Riemannian metric on the configuration manifold must not be overlooked; indeed, a free particle (i.e. in the absence of potential forces) will follow a trajectory along the *geodesic* dictated by the 'kinetic co-energy metric'; this is called the Maupertuis-Jacobi principle [14].

The distinction between energy and co-energy is not very common in literature, although thoroughly discussed by Jeltsema and Scherpen [18]. While energy is the ability to do work, and co-energy is the *complement of energy*. Since energy is defined in terms of work, kinetic energy $T$ should be defined in terms of momentum – the integral of an applied force over time, instead of a velocity. The linear relation behind the change of variables from momentum to velocity by means of the mass makes the distinction between energy and co-energy moot at first glance, but it is nevertheless important to consider. However, the kinetic energy and co-energy are not always equal, e.g. in the relativistic case where the mass will depend on the velocity as well. In the simple nonrelativistic case, for a single particle with mass $m$, the following relations hold:

$$\text{kinetic energy } T = \int \frac{p}{m}\,\mathrm{d}p = \frac{p^2}{2m} \qquad \text{kinetic co-energy } T^* = \int mv\,\mathrm{d}v = \frac{mv^2}{2}.$$

**Kinetic energy and market surplus**

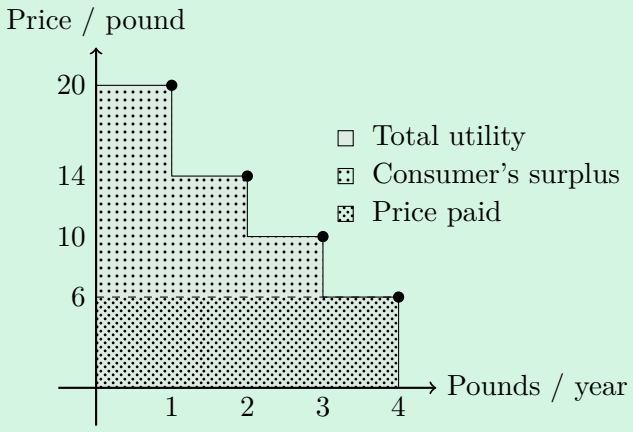
The interpretation of kinetic energy in economic engineering is a big leap forward, perhaps one of the most fundamental aspects of the entire theoretical framework. However, the formulation in eq. (2-2) obscures the intuition behind it. This is why it is more instructive to look at the simple scalar case, where kinetic energy is a notion of the amount of work it takes to accelerate a particle from rest to a certain velocity.
To explain the significance of kinetic energy as surplus, the example given by Marshall [19, chap. 6] about consumer's surplus will be recycled here in order to illustrate the point.
Imagine a woman (Jane), who likes to drink tea. When buying tea, she continuously makes the unconcious deliberations between (i) how much she likes to drink tea and

[4] As will become clear later, the term 'co-energy' is used to distinguish between the definition of kinetic energy in terms of the generalized momenta, which is the perspective of Hamiltonian mechanics.

(ii) how much she likes to pay for tea — this is a consequence of the assumption that the Jane is a rational market participant. Marshall quantifies the woman's unconcious deliberation process by means of the following table:

| Price / pound | 20 | 14 | 10 | 6 |
|---|---|---|---|---|
| Pounds bought / year | 1 | 2 | 3 | 4 |

It is rather straightforward that Jane is willing to buy more tea as it gets cheaper and vice versa. Marshall explains this by means of the concept of utility: if the price of tea, say 20$, drops to 14$, Jane obtains her additional pound of tea for 14$ for instead of the 20$ she was willing to pay for the first pound. As such, she gains a surplus satisfaction (or consumer's surplus) of 6$ — this is, as Marshall states, *precisely* the additional utlity of the second pound of tea, i.e. how much Jane values it on top of the first one. The total utility of the the two pounds of tea per year is then 20$ + 14$ = 34$. More simply stated, the consumer's surplus is the total utility of the products bought minus the price actually paid. In this example, the total utility is *at least* 34$, and the price paid is 2×14$ = 28$: therefore, Jane's the consumer's surplus is 6$.




Indeed, this example illustrates that *surplus* and *utility* are closely related; indeed, they only differ by the choice of a 'setpoint' (the price at which Jane is buying tea).
Both have units of $/yr, a consequence of the demand being in quantity/yr (of course 'yr' is an arbitrary choice for a time unit for the sake of this example) – this is an important distinction from other texts in economics, which tend to be rather vague as to whether the demand is a flow or an absolute quantity of goods. Based on this discussion, the following relations can be obtained:

$$\text{total surplus} = \sum_{i=1}^{i^m} p_i \Delta v \qquad \text{consumer's surplus} = \sum p_i \Delta v - \underbrace{\sum p_m \Delta v}_{\text{amount paid}} \tag{2-3}$$

with $p$ the price and $v$ the amount of tea sold per year. Hence, the summations happen over a set of prices between two points: a 'reference' price (in this case, $20), and the market price $p_m$; of course, the value of the consumer's surplus depends on the choice for these prices. Naturally, the market price may be quite undisputed, but

the reference price is just that, and its choice is arbitrary. In this simple example, it happens to coincide with the reservation price for the first pound of tea, but that need not be the case at all.
By virtue of the foregoing discussion, it is established that *the trade utility measured at a given price is the (consumer's) surplus.* The Lagrangian represents an exchange of 'trade utility' (dependent on the flow of goods) and the 'product utility', dependent on stock levels. To establish the analogy with mechanics, observe that the part of the mechanical Lagrangian that depends on the generalized velocities is the kinetic energy. This connection is the foundation of the following economic engineering principle:

**Kinetic (co-)energy is analogous to market surplus.**

In mechanics, the calculation of kinetic energy is *dependent on the frame of reference*, and this is analogous to the reference price used to calculate the consumer's surplus. There is one additional loose end: in the example of Jane, there was extensive mention of price, while the Lagrangian and the kinetic co-energy are only dependent on the flow of goods. One can observe from the example that the price determines the additional increase of surplus for every increase in the amount of tea bought per year, or otherwise

$$\frac{\Delta(\text{total utility})}{\Delta v} = p_i$$

To generalize, one can assume that $v$ and $p$ are continuous variables instead, related to each other by the bijective 'reservation price mapping' that is denoted by $m : v \mapsto p$. The summations in the previous examples can then all be replaced by integrals:

$$\text{total utility} = \int p \, \mathrm{d}v \qquad \text{consumer's surplus} = \int p \, \mathrm{d}v - p_1 \int \mathrm{d}v$$

i.e. the total utility and consumer's surplus only differ by a choice of reference frame. With the consumer's surplus being analogous to kinetic energy, one can say that

**Price is analogous to momentum.**

### 2-2-4 Potential energy

In mechanics, the potential energy arises due to the presence of a conservative force vector field $\boldsymbol{F}$. A vector field is conservative if the work done along any path only depends on the endpoints of the path and not on the intermediate shape. If that is the case, then it is always true that one can find a function such that[5]

$$\boldsymbol{F} = -\frac{\partial U}{\partial \boldsymbol{x}}.$$

In the Lagrangian context, the position vector $\boldsymbol{x}$ can be expressed in terms of the generalized coordinates, such that (with a slight abuse of notation)

$$\boldsymbol{F} = -\frac{\partial U}{\partial \boldsymbol{q}}.$$

[5] This is a consequence of the fundamental theorem for line integrals [20].

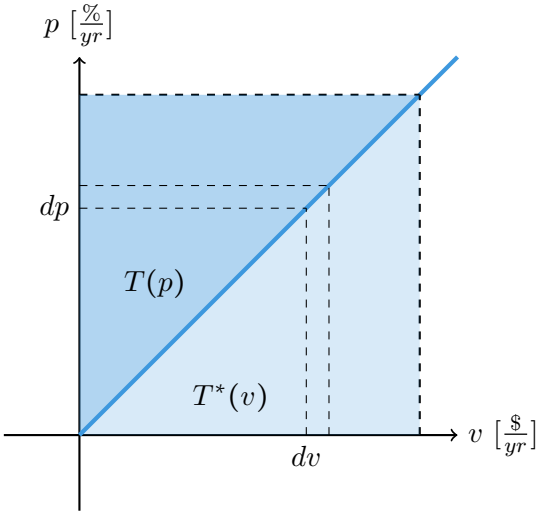



**Figure 2-2:** Blabla. Figure courtesy of B. Krabbenborg [1].

**Potential energy**

While in 'regular physics' potential energy arises due to the presence of elastic, electric, gravitational ... forces, its interpretation in economic engineering is related to the utility gained from the posession of goods; i.e. it is a function that has units of cash flow (just like surplus) that depends only on the stock levels (possibly in terms of generalized coordinates).

The simplest example is the economic analogue of a spring: just like a spring, a restoring 'force' (i.e. an economic want) occurs whenever the spring is elongated or stretched with respect to a certain reference distance. For economics agents, this means either being long or short on stock levels, again with respect to a certain reference $q_0$ (although important for practical interpretations, this reference does not really contribute on a conceptual levels, which is why it is omitted in the theoretical discusion, just like the reference momentum for the kinetic energy). Hence, there is a 'potential utility' in either being long or short, in the sense that it can be exchanged for 'market utility' or economic surplus, because it requires an exchange of goods. The continuous reciprocity between kinetic energy and potential energy is a central theme in mechanics, and it is henceforth been given a succinct economic intepretation as well.

The stereotypical example in mechanics that contains a single storage of kinetic energy (a mass) and a single storage of potential energy (a spring) is the mass-spring system, which also arises as the linearization of a multitude of undamped nonlinear systems (i.e. the pendulum); the result is a second-order scalar autonomous differential equation:

$$m\ddot{q} + \frac{\partial U}{\partial q} = 0, \quad \text{with} U = \frac{kq^2}{2} \tag{2-4}$$

with $k$ being the sprign constant; this is immediately results in Newton's second law of motion. A simple mass-spring system is shown in fig. 2-3.

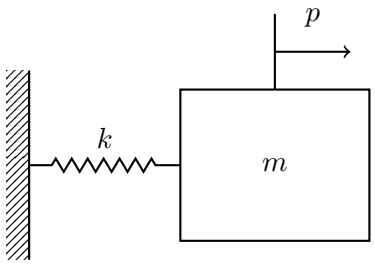



**Figure 2-3:** A simple mass-spring system. Figure courtesy of B. Krabbenborg [1].

The distiction between energy and co-energy exists for potential energy too. Again, the potential energy itself is defined in terms of work: the integral of force over distance. Comparable between the bijective simple linear relation between velocity and momentum, for simple linear springs the force $F$ and displacement $q$ are related by $F = kq$. Again, in this case the distinction may seem a bit pointless; but, when this relation is nonlinear (as it in the real world often is), the potential energy and co-energy are no longer the same.

$$\text{potential energy } U = \int F \,\mathrm{d}q \quad \text{potential co-energy} U^* = \int q \,\mathrm{d}F$$

which for in the case of the simple spring turn out to be

$$U = \frac{kq^2}{2} \qquad U^* = \frac{f^2}{2k}.$$

In the case of potential energy, the familiar equation really *is* the one that corresponds to energy (and not to co-energy, as for kinetic energy). The second equation seems really odd, intuitively turning the causal relation around; but one should bear in mind that this equation is, from a fundamental perspective, equally 'odd' as the extremely familiar $\frac{mv^2}{2}$. Figure 2-4 shows the relation between potential energy and co-energy and why they are equal for a linear relation between force and displacement. The solution of the simple mass-spring system is a

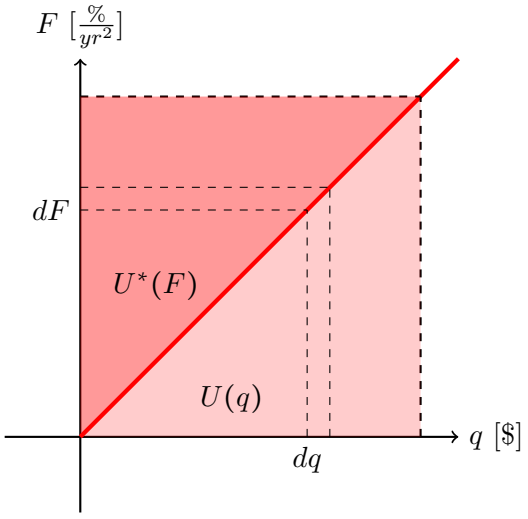



**Figure 2-4:** Blabla. Figure courtesy of B. Krabbenborg [1].

sinusoid which continuous indefinitely in time. Intuitively, this seems odd; as it does seem to apply directly to economic systems. One must, however, bear in mind that (i) the perfect mass-spring system does not exist in mechanics either, and (ii) that economic systems are much harder to 'decouple' from their surroundings due to the complex network of society. This makes the theoretical achievement that are the basic foundations economic engineering perhaps even more admirable. Because of these external factors, eq. (2-4) will in practice usually contain external inputs in the form of forcing terms in lieu of the 0 after the equality sign, and *dissipative terms* which indicate path dependent forces in the system. These forces dissipate energy and are always present in reality, both in mechanics and economics; they will be the subject of the next section.

### 2-2-5 Dissipative terms

**Dissative energy**

Foo bar

### 2-2-6 Noether's theorem

**Noether's theorem**

If the Lagrangian system $(M, \mathcal{L})$ admits the one-parameter group of transformations $h^s : M \to M, s \in \mathbb{R}$, then the Lagrangian system of equations corresponding to $\mathcal{L}$ has a first integral: $I : TM \to \mathbb{R}$. In local coordinates $q$ on $M$ the integral $I$ is written in teh form

$$I(\boldsymbol{q}, \dot{\boldsymbol{q}}) = \frac{\partial \mathcal{L}}{\partial \dot{\boldsymbol{q}}} \frac{\mathrm{d}h^s(\boldsymbol{q})}{\mathrm{d}s}\bigg|_{s=0}$$

## 2-3 Hamiltonian mechanics

### 2-3-1 The Legendre transform

Whereas the Lagrangian of a system is a function of its generalized positions and generalized velocities, the Hamiltonian is a function of the generalized position and generalized momentum. Recall from the definition that the generalized (or *conjugate*) momenta are defined in terms of the Lagrangian; $p_i = \frac{\partial \mathcal{L}}{\partial \dot{q}_i}$. Hence, the momenta and velocities are conjugate variables. To achieve this change of variables, from the

The total differential of the Lagrangian is

$$\mathrm{d}\mathcal{L} = \frac{\partial \mathcal{L}}{\partial \boldsymbol{q}} \mathrm{d}\boldsymbol{q} + \frac{\partial \mathcal{L}}{\partial \dot{\boldsymbol{q}}} \mathrm{d}\dot{\boldsymbol{q}}\,.$$

Two observations can be made: firstly, by definition, $\boldsymbol{p} = \frac{\partial \mathcal{L}}{\partial \dot{\boldsymbol{q}}}$; and secondly, the Euler-Lagrange equations dictate that $\frac{\mathrm{d}}{\mathrm{d}t} = \frac{\partial \mathcal{L}}{\partial \boldsymbol{q}}$. As such, the total differential of the Lagrangian may be

rewritten as:

$$\mathrm{d}\mathcal{L} = \dot{\boldsymbol{p}}\,\mathrm{d}\boldsymbol{q} + \boldsymbol{p}\,\mathrm{d}\dot{\boldsymbol{q}}$$

### 2-3-2 Hamilton's equations

### 2-3-3 Symplectic manifolds

### 2-3-4 The canonical formalism

One of the great advantages of Hamiltonian mechanics is that it admits a much wider range of coordinate transformations. Of course, any selection of the generalized coordinates that parameterizes the admissible motions of the system is equally valid, the generalized velocities are inherently tied to the choice of these coordinates. In Hamiltonian mechanics, the $p$ and $q$ coordinates can be chosen completely independent of each other, which is why a larger class of transformations is allowed[6] [15].

Canonical transformations are transformations such that Hamilton's equations remain valid in the new coordinate system. As shown by Landau and Lifshitz [15], each canonical transformation is characterized by a *generating function*. Poisson brackets are invariant with respect to canonical transformations.

[6] Due to the large variety of transformations, the coordinates for $q$ and $p$ may not longer just pertain to a 'spatial' component and a 'momentum' component, this distincition will then just be a matter of definition.

# Chapter 3

# Introductory actuarial concepts

(... Chapter intro ...)

## 3-1 The concept of interest

(... Intro with some history ...)

Usually, two types of interest are distinguished: simple interest and compound interest. The key difference is that in the case of compound interest, the money earned (or due) on an interest-bearing instrument is subject to interest itself. That is, if somebody takes out a loan, the corresponding interest payments are also considered as a contribution to the principal and therefore result in higher future interest payments. In contrast, the case of simple interest does not consider the compounding effect; the interest payments remain constant over time.

The compounding of interest is always associated with a certain period; this is the period over which the interest earned is calculated and added to the amount due (or 'reinvested' in case of an investment). The choice of this period is part of the agreement between the lender and the borrower, common compounding periods are daily, weekly, monthly or yearly. From a mathematical perspective, one can view the choice of compounding period as a limiting process: clearly, the choice in practice is rather arbitrary. Indeed, the 'ideal' compounding period is continuous, but before the advent of computers this was not achievable in practice; in banking applications it is still not very commonly used. [21]

### 3-1-1 Interest terminology

Many interest calculations are subject to conventions, which is why there are several terms associated 'interest' which must be clearly distinguished. Usually, an interest-bearing instrument is associated with some amount that is initially borrowed or invested, which is called the *principal*, denoted by $K$. Secondly, one can define the *accumulation function* $a$ which

yields the accumulated value of a unit investment (or loan).[1] Similarly, the *amount function* $A$ measures the accumulation of any principal value $K$: $A(k) = Ka(k)$. [22]

Concerning the period $k$, some additional remarks are in order. In most banking applications, the interest process is discrete, i.e. the compounding effect occurs over discrete intervals. Apart from the compounding process, one can distinguish a 'measurement' interval which can, but does not necessarily have to, coincide with the compounding period. For now, the interest process is assumed to be calculated on a discrete basis; the transition to continuous time will be made after that.

> The *effective rate of interest* $r$ is the ratio of the amount of interest earned during the period to the amount of principal invested at the beginning of the period. In terms of the accumulation function:
>
> $$i(k) = \frac{a(k) - a(k-1)}{a(k-1)} \tag{3-1}$$
>
> where $a$ is in this case assumed to take discrete values.

### Simple interest

For the case of simple interest, the accumulation function has the form

$$a(k) = ik + 1$$

where $i = i(0)$ denotes the constant simple interest rate, which turns out to be identical to the effective rate of interest over the first period. An important fact one has to bear in mind is that for the case of simple interest the effective rate of interest is *not* constant over time (this, in fact, motivates the existence of compound interest, as will become clear later). Using eq. (3-1):

$$i(k) = \frac{(ki+1) - [(k-1)i+1]}{(k-1)i+1} = \frac{i}{1+i(n-1)}$$

which means that simple interest results in a decreasing effective rate of interest over time. [22]

### Compound interest

As mentioned before, compound interest relies on the reinvestment of the interest already earned. At the end of the period the accumulation function is a factor $1+i$ larger than the period before. One can therefore say that

$$a(k) = (1+i)a(k-1) \quad \text{or} \quad a(k) = (1+i)^k$$

[1] In this discussion, the distinction between credit or debit (i.e. investments or loans) is quite irrelevant for the principles at hand, they just differ in sign on the balance sheet of the respective counterparties. Hence, these terms will be used interchangeably. Minus signs are there to indicate cash flows in the 'opposite' direction, regardless whether this is on the credit or debit side of the balance sheet.

assuming that the compounding period coincides with the measurement period.

Similarly, eq. (3-1) can be used to compute the effective rate of interest for an arbitrary period:

$$i(k) = \frac{(1+i)^k - (1+i)^{k-1}}{(1+i)^{k-1}} = i$$

which means that for compound interest, the effective rate of interest is *constant*. This is the reason why compound interest plays an ubiquitous role in modern finance; otherwise it would become less and less profitable for investors to keep their money in a certain investment (they would rather just stay a single period and then turn to a new investment — this effectively emulates compound interest!). Only for short periods (less than one year), simple interest is sometimes used because the difference with compound interest is negligible, which can be motivated by means of the Taylor series of $a$ for compound interest:

$$(1+i)^k = 1 + i + \binom{k}{2} i^2 + \binom{k}{3} i^3 + \ldots \approx 1 + i \quad \text{for} \quad i \ll 1,\ k \text{ small}$$

In fact, from this equation it is clear that simple interest and compound interest yield the same result after the first compounding period, after which compound interest will take the upper hand.

For now, it was assumed that one compounding period is equal to one measurement period. However, this does not necessarily have to be the case. More generally, one can write the accumulation function as:

$$a(k) = \left(1 + \frac{i}{n}\right)^{nk}$$

where $n$ is the number of compounding periods per measurement periods (which is assumed to be an integer number). In actuarial sciences, the measurement period is normally equal to one year; meaning that e.g. weekly compounding amounts to $n = 52$.

Here, $i$ is called the *nominal interest rate*, which is usually denoted by $i^{(n)}$, indicating that the nominal rate $i$ is compounded $n$ times at a rate $i^{(n)}/n$. Hence, a nominal interest rate of $i^{(m)}$ per measurement period is equivalent to an effective interest rate of $i^{(m)}$ per $n$th part of that measurement period.

### Continuous time

Now, the discrete accumulation (and amount) functions will be converted to their continuous counterparts, since that will be the most convenient form from a mathematical perspective. To do so, consider the nominal interest over single measurement period, but subdivided into an ever growing number of compounding intervals. Because the number of compounding intervals grows to infinity, the choice of measurement interval becomes immaterial which is why it can simply be replaced by '$t$' denoting continuous time.

In case of simple interest, the result is trivial:

$$a(t) = \lim_{n \to +\infty} \left(\frac{i}{n}\right)(tn) + 1 = it + t$$

which basically means that the nominal interest remains the same no matter the 'compounding period'.

For compound interest, the result is more interesting:[2]

$$a(t) = \lim_{n\to\infty}\left(1+\frac{i}{n}\right)^{nt} = \mathrm{e}^{it}$$

### 3-1-2 Discounting, Net Present Value and the Internal Rate of Return

Internal Rate of Return (IRR) Net Present Value (NPV)

### 3-1-3 Lorentz structure from the compounding process

Figure 3-1 exemplifies the compounding process with 5% interest on an initial investment of $1000. For now, it is immaterial what the compounding frequency is. The total value of the investment is decomposed into two parts, which will be called the 'capital' and 'yield' fractions. Every time the investment is 'compounded', the following happens:

- The interest rate acts on the amount of capital outstanding, the result of which adds to the current total yield.
- The interest rate acts on the current total yield, the result of which is added to the capital outstanding. This is sometimes called *interest on interest* and lies at the very core of the compounding principle. If this action would not be present, the process reduces to simple interest.

Clearly, apart from the obvious symmetry, this decomposition is motivated by its intuitive interpretability. Much of what is to come in the dissertation hinges on this principle, which is why this example, obvious as it may seem, is important. Two approaches will be discussed that elegantly capture this process from a computational standpoint: discrete LTI systems and hyperbolic-complex numbers. These methods are also closely related and essentially represent two sides of the same coin. The results that follow should therefore not come as surprising, though looking at them from different perspectives is certainly instructive.

[2] There exist a few equivalent definitions of the exponential function, one of which is this limit. As such, some may argue that this statement is true *by definition*.

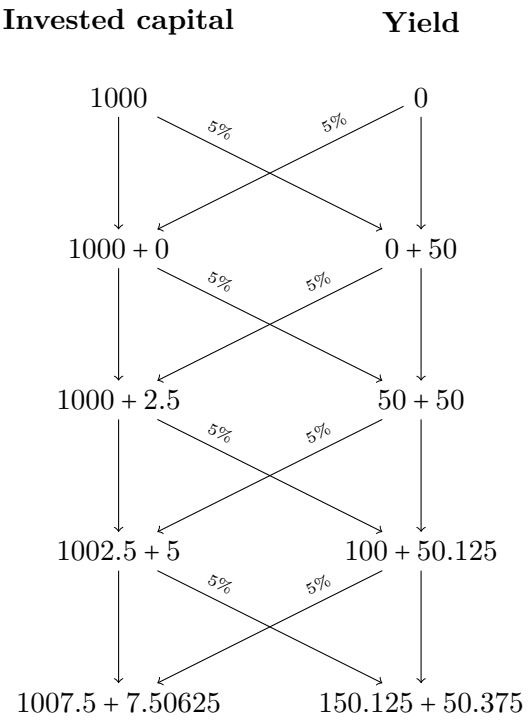



**Figure 3-1:** Caption

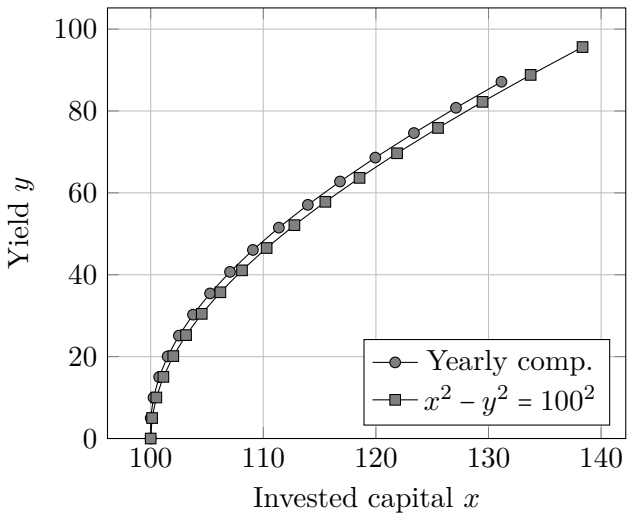



**Figure 3-2:** blabla

# Chapter 4

# The Lorentz-Minkowski Plane

In the previous chapter, it has been established that interest processes can be represented on hyperbolic curves by disregarding the time dimension and 'decomposing' the exponential in a specific way. This result is almost trivially encapsulated in the identity

$$\exp(\zeta) = \cosh(\zeta) + \sinh(\zeta) \tag{4-1}$$

The most intuitive way to interpret this identity is by inspection of the Taylor series of the functions appearing in eq. (4-1):

$$\begin{aligned} \exp(x) &= \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots \\ \cosh(x) &= \sum_{k=0}^{\infty} \frac{x^{2k}}{(2k)!} = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \dots \\ \sinh(x) &= \sum_{k=0}^{\infty} \frac{x^{2k+1}}{(2k+1)!} = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \dots \end{aligned} \tag{4-2}$$

Or, alternatively, via the definition of the hyperbolic functions

$$\begin{aligned} \cosh(x) &\triangleq \frac{\exp(x) + \exp(-x)}{2} \\ \sinh(x) &\triangleq \frac{\exp(x) - \exp(-x)}{2} \end{aligned} \tag{4-3}$$

Clearly, the cosh and sinh functions are constructed by isolating the even or the odd powers respectively from the Taylor expansion of the exponential. The astute reader may notice that eq. (3-1) bears some resemblance to Euler's formula $\exp(\mathrm{i}x) = \cos(x) + \mathrm{i}\sin(x)$. As described by Needham [2], this connection can be generalized by recognizing that

$$\cos(\mathrm{i}x) = \cosh(x) \qquad \sin(\mathrm{i}x) = \mathrm{i}\sinh(x)$$

As such, both the hyperbolic functions sinh, cosh and the trigonometric functions cos and sin can all be represented by looking at the modular surface of $|\sin(z)|$, visualized in fig. 4-1: of course, sin and cos only differ by a shift of $\pi/2$ along the real line, and cosh and sinh exist at cross-sections into the complex at integer multiples of $\pi/2$ and $\pi$ respectively.

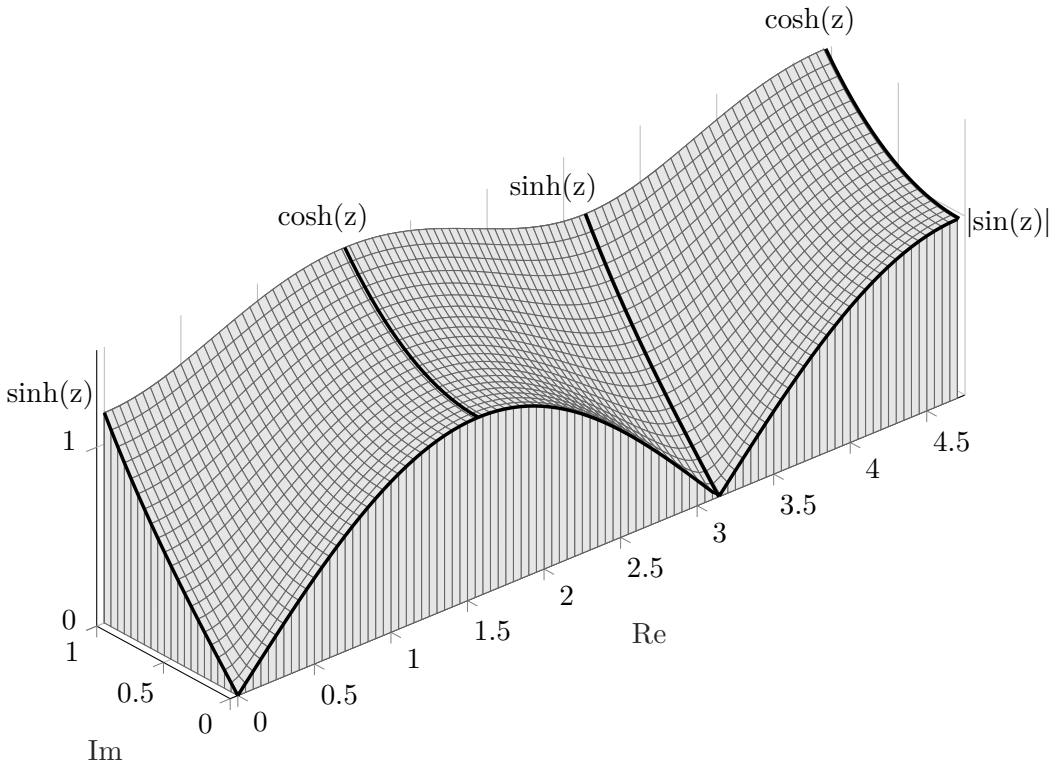



**Figure 4-1:** Modular surface of the sine over the complex plane, embedding all the trigonometric and hyperbolic functions at specific cross-sections.

## 4-1 Conic sections

There are a variety of different (representations) of hyperbolae; in this case the discussion will be limited to so-called *rectangular hyperbolae*, which are basically rotations, translations or a combination thereof of the scaled reciprocal function $y = K/y$. The term 'rectangular' refers to the fact that there is no squeeze or stretch along any particular direction. Although the term 'rectangular hyperbola' therefore corresponds to a whole range of shapes, including the unit hyperbola which is to be defined later, it will henceforth be used here as a *totum pro parte* to refer to the function $y = K/y$ in particular in order accentuate the distinction with the unit hyperbola. An implicit parameterization of this rectangular hyperbola is

$$\begin{cases} x = \pm K\mathrm{e}^{t} \\ y = \pm K\mathrm{e}^{-t} \end{cases} \quad t \in \mathbb{R},\ K \in \mathbb{R}_{*}^{+}$$

Clearly, this hyperbola has two asymptotes along the $x$ and $y$ axes. The line $x = y$ is called the *major axis* of the hyperbola. By rotating the hyperbola over a 45° angle in clockwise direction, the major axis will coincide with the $x$-axis, and a subsequent 'squeeze' by factor $\sqrt{2}/2$, one arrives at the so-called *unit hyperbola* defined by the implicit equations $x^2 - y^2 = K^2$. Applying this linear transformation (scale with $\sqrt{2}/2$ and rotate by $-\pi/4$) as a matrix operation to the

aforementioned parametric description, one arrives at

$$\frac{\sqrt{2}}{2}\begin{pmatrix}\sqrt{2}/2 & \sqrt{2}/2\\ -\sqrt{2}/2 & \sqrt{2}/2\end{pmatrix}\begin{pmatrix}\pm K\mathrm{e}^{t}\\ \pm K\mathrm{e}^{-t}\end{pmatrix} = \pm\frac{K}{2}\begin{pmatrix}\mathrm{e}^{t}+\mathrm{e}^{-t}\\ \mathrm{e}^{t}-\mathrm{e}^{-t}\end{pmatrix} = K\begin{pmatrix}\pm\cosh(t)\\ \sinh(t)\end{pmatrix}$$

which is the common parameterizaton of the unit hyperbola, this will be the standard representation throughout this text. Please note that the '±' in front of the sinh can be disregarded because the sinh is an odd function. One can now also recognize the asymptotes as an alternative axis system, which recovers the original hyperbola; this axis system will be referred to as the *idempotent axis system*. The hyperbolae and the corresponding axis system are shown in fig. 4-2.

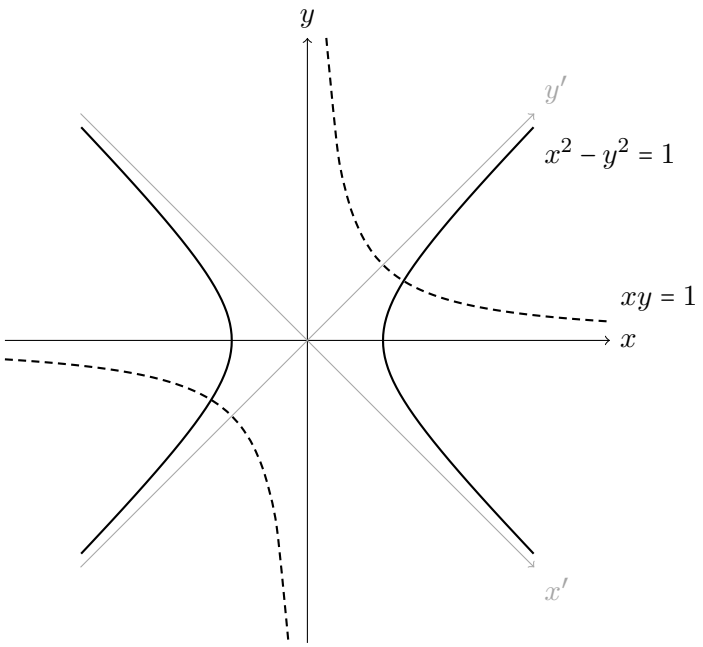



**Figure 4-2:** Comparison between the reciprocal function $(x, \frac{1}{x})$ and the unit hyperbola $(\cosh(t), \sinh(t))$ (implicit equation $x^2 - y^2 = 1$). The idempotent axis system (denoted by the gray lines) coincides with the asymptotes of the unit hyperbola; from this axis system the unit hyperbola again satisfies the equation $x'y' = 1$.

## 4-2 Hyperbolic angles

Until now, the argument of the sinh/cosh parameterization has been $t$ as is the familiar notation for a parametric curve. However, the duality between circles and hyperbolas can be exploited further by defining the arguments of the hyperbolic functions as *hyperbolic angles*, like one does for the trigonometric functions sin and cos, which could be named 'circular' angles to further highlight the distinction. In both cases, an angle refers to a certain region bounded by the curve (hyperbola/circle) at issue. The standard notation for the hyperbolic angle will be $\zeta$, in accordance to the rapidity from special relativity which can also be viewed as a hyperbolic angle as will be discussed later on.

**Hyperbolic sector**

A hyperbolic sector is the region bounded by two lines extending from the origin to each to a point on the (unit) hyperbola, and the graph of the hyperbola itself.

Clearly, hyperbolic sectors are entirely analogous to their 'traditional' circular cousins. Fixing one of the rays to the $x$-axis, one can define the corresponding hyperbolic angle:

**Hyperbolic angle**

A hyperbolic angle corresponding to a point $A$ is defined as twice the area of the hyperbolic sector based on the point $A$ and the intersection point of the unit hyperbola and the $x$-axis $(K, 0)$.

Clearly, any point on a hyperbola with radius $K$ can be parameterized using

$$(K\cosh(\zeta), K\sinh(\zeta))$$

By allowing the radius $K$ to be negative, all the points in the disconnected open set bounded by $y = x$ and $y = -x$ can be identified with a unique radius $K$ and hyperbolic angle $\zeta$. This is somewhat similar to polar coordinates, which is why these coordinates will be referred to as 'hyperbolic polar coordinates'.

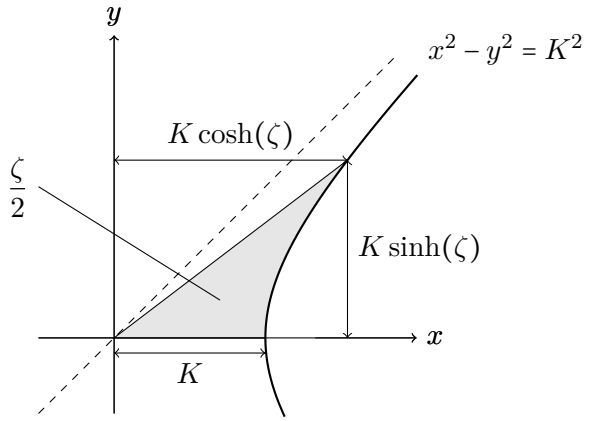



**Figure 4-3:** Illustration of a hyperbolic angle along the hyperbola with semi-major axis $K$.

It may already be clear that hyperbolic polar coordinates do *not* provide coordinates for the entire plane like regular polar coordinates do. Indeed, the mapping defined by the coordinate functions from the $K - \zeta$ space to the $x$-$y$ space is neither injective nor surjective: its image is the disconnected open set bounded by the lines $y = x$ and $y = -x$ (not surjective), and the entire line $K = 0$ in the $K - \zeta$ plane is mapped to the origin in the $x - y$ plane (not injective). As such, one can obtain a bijection by disregarding the degenerate cases for which $K = 0$ and restricting the codomain of the mapping to the set $\{(x, y) \in \mathbb{R}^2 : |x| > |y|\}$. The action of the mapping is illustrated by fig. 4-4.

A possible workaround for this problem can be found in the so-called *generalized trigonometry* as described by Harkin and Harkin [23]: in this case, the notion of the hyperbolic angle

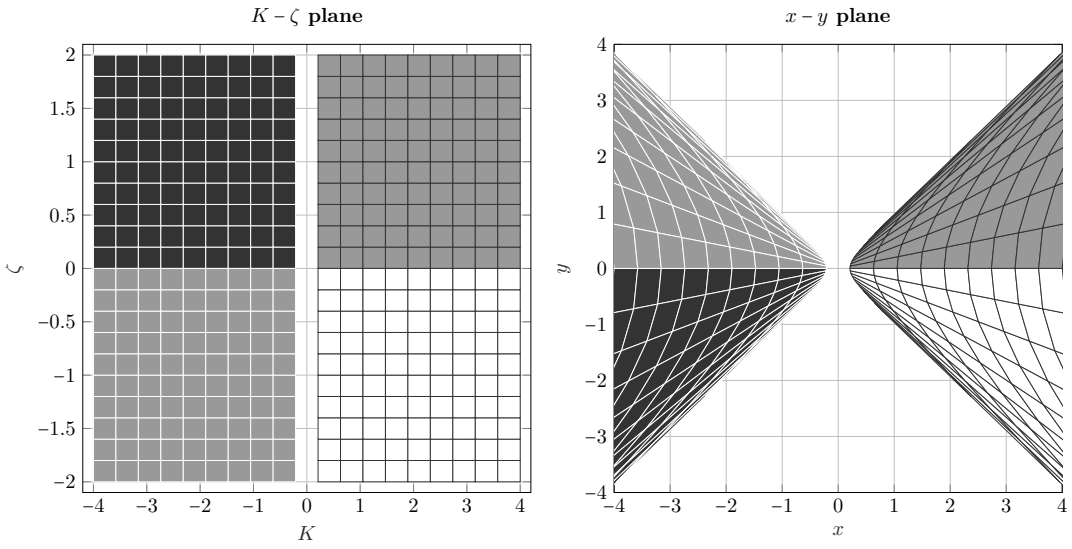



**Figure 4-4**

itself is considered ambiguous and must be accompanied the branch of the hyperbola that is associated with the angle. In this case four hyperbolic branches are considered, so both branches of $x^2 - y^2 = K^2$ combined with $y^2 - x^2 = K^2$ which were disregarded up till now.[1]

$$\zeta = \begin{cases} \tanh^{-1}(x/y) & \text{for branches I and III} \\ \tanh^{-1}(y/x) = \coth^{-1}(x/y) & \text{for branches II and IV} \end{cases} \tag{4-4}$$

Although this extended definition brings the other two quadrants of the hyperbolic plane within reach of the polar form as well, there is still a particular set that even now cannot be represented this way: $\{(x, y) \mid |x| = |y|\}$, or the 'light cone' from special relativity.

## 4-3 Introduction to the theory of special relativity

This section gives a basic outline of the theory of special relavity. It is by no means meant to be a comprehensive overview, for which many other excellent resources exist such as Misner et al. [24], Taylor and Wheeler [25], Landau and Lifshitz [26], or Penrose [27] for a shorter, less technical introduction. Instead, the goal of this section is to give a physical context for the concept of Lorentz geometry and the associated Lorentz space and metric, as well as their connection with hyperbolic geometry and Möbius transforms which will be introduced in chapter 5 and chapter 6.

Before the advent of the theory of special relativity, developed by i.a. Poincaré, Minkowski, Lorentz and Einstein, so-called 'Galilean relativity' was the norm. Galilean relativity entails the definition of Galilean transformations which links reference frames that move relative to each other at a constant linear speed. The laws of physics should be invariant between

[1] In fact, Harkin and Harkin provide an even more general treatment, covering so-called generalized complex numbers of the form $z = x + \iota y \quad x, y \in \mathbb{R}$ with $\iota^2 = \iota q + p \quad p, q \in \mathbb{R}$. However, in this case $p = 1$ and $q = 0$, as will become clear later in the discussion about hyperbolic numbers.

these inertial reference frames. However, a problem presented itself in the form of the famous Maxwell equations that pose the governing laws in electromagnetism. A consequence from Maxwell's laws is the finite propagation speed of light and therefore all possible interactions between particles in the universe. It is not hard to imagine that the Galilean invariance breaks down as a result of the introduction of a 'special' speed - indeed, the Galilean transform proclaims a complete independence between the physical laws and the constant velocity of the frame in which they are applied. But, as a direct consequence from Maxwell's equations, the laws of physics in a moving reference frame *will* depend on the speed of that particular reference frame: a major inconsistency with the traditional train of thought.

The new principle of relativity that brought reconciliation with Maxwell's ideas not only required space to be relative (i.e. dependent on a frame of reference), but also views *time as a relative concept* whereas it had always been assumed to be absolute in classical mechanics. As such, the notion of time is dependent on the choice of reference frame too. This has the immediate consequence that the traditional three-dimensional setting of classical mechanics (often with Cartesian coordinates $x, y, z$) will not suffice for the description of special relativity: a fourth coordinate four time is indespensable to incorporate the relativity of time. Points in the four-dimensional *spacetime* are called *world points*, their associated trajectories are *world lines* [26].

### 4-3-1 Spacetime intervals

Overwhelming experimental evidence has pointed out that the propagation of light is completely independent of its direction. This can be encapsulated in spacetime by means of *spacetime intervals* which provide a notion of distance between two world points. If a signal travels at the speed of light $c$, the distance between two world points along its trajectory should be zero. The spatial distance squared between two points is equal to

$$(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2$$

whereas the distance squared covered by a signal travelling at the speed of light is equal to

$$c^2(t_2 - t_1)^2.$$

Therefore, the spacetime interval between two world points is

$$s_{12} = \sqrt{c^2(t_2 - t_1)^2 - (x_2 - x_1)^2 - (y_2 - y_1)^2 - (z_2 - z_1)^2}$$

which will amount to zero for the world lines corresponding to a signal travelling at the speed of light. For an infinitesimal distance $\mathrm{d}s$, the spacetime interval can be expressed as

$$\mathrm{d}s^2 = c^2\,\mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2\,.$$

The spacetime interval is the same in any inertial reference frame. This is the mathematical translation of the invariance of the speed of light in the universe [26]. Based on the sign of the spacetime interval, three classes can be distinguished.

- If $s_{12}^2 > 0$, the interval is *timelike* and there exists a frame of reference in which both events occured *at the same location* in space, they are simply separated by the passage of time $t_{12} = \frac{s_{12}}{c}$;

- in contrast, when $s_{12}^2 < 0$, the interval is *spacelike* and the events are 'too far apart' to reach within the limits of the speed of light — the events must therefore be at different locations (absolutely remote), and there exists a reference frame in which the events occur *simultaneously* at distance $l_{12} = \mathrm{i}s_{12}$;

- intervals for which $s_{12}^2 = 0$ are called *lightlike*, because only light can travel between these events.

Now, one can ask the question what the actual time is that an observer would experience in uniform motion, i.e. what difference in time is there on clocks which have been travelling at different velocities? The time experienced by an observer is called *proper time*, and it can be computed by evaluating the following path integral:

$$t_2' - t_1' = \int_{t_1}^{t_2} \mathrm{d}t \sqrt{1 - \left(\frac{v}{c}\right)^2}. \tag{4-5}$$

Clearly, if the velocity makes up a larger fraction of the speed of light, the proper time is lower; that is: moving clocks run slower than a clock at rest (hypothetical clocks travelling at the speed of light do not register the passage of time whatsoever). (`... Twin paradox ? ...`)

### 4-3-2 Lorentz transformations

As mentioned, special relativity corrects the flaw of the Galilean transforms, which represent the classical view of inertial reference systems: for example, if one coordinate system moves at constant velocity with respect to the other in $x$-direction (the coordinate directions are assumed to coincide for simplicity), the Galilean transform takes the form:

$$x = x' + Vt, \quad y = y', \quad z = z', \quad t = t'. \tag{4-6}$$

The statement $t = t'$ encodes the traditional assumption in mechanics that time has an absolute character. Of course, it is precisely this statement that is refuted by special relativity. As such, one could devise a new type of transformation that takes this (and the invariance of spacetime intervals between events, as discussed in the previous section) into account. These transformations are called *Lorentz transformations*.

As described by Landau and Lifshitz [26], these transformations comprise the rotations in four-dimensional space: since there are six ways to pick a plane (or two coordinates) from a set of four axes, every rotation in four-space can be decomposed into six successive rotations. Of these six rotations, three are purely spatial: they are the familiar rotations that can be parameterized by e.g. Euler angles. On the other hand, the three other rotations involve time as well, and they are of a different nature. Whereas the spatial rotations are circular, the time-rotations are hyperbolic (they are represented by hyperbolic functions rather than trigonometric functions). For example, a rotation in the $tx$-plane would take the following form:

$$x = ct' \sinh(\zeta) \qquad ct = ct' \cosh(\zeta); \tag{4-7}$$

or, using the fact that $V = x/t$:

$$x = \frac{x' + Vt'}{\sqrt{1 - \left(\frac{V}{c}\right)^2}} \qquad t = \frac{t' + \frac{V}{c^2}x'}{\sqrt{1 - \left(\frac{V}{c}\right)^2}} \qquad y = y' \qquad z = z'; \tag{4-8}$$

which also indicates that the hyperbolic (boost) angle $\zeta$ can be written in terms of the velocity $V$ of one frame with respect to the other

$$\tanh(\zeta) = \frac{V}{c},$$

which means that the argument of the hyperbolic rotation purely depends on the relative velocity between the two reference frames as a fraction of the speed of light. A few observations can be be made based on these equations:

- Clearly, $V$ cannot be larger than $c$; there is no real $\zeta$ for which this could be true. This again reaffirms the statement that there can be no motions with velocities larger than the speed of light.
- Secondly, this transform keeps $c^2t^2 - x^2$ unaffected (of course, $z$ and $y$ keep their value for obvious reasons); all points in the $tx$-plane that remain invariant under this type of transformations lie on the same hyperbola. This underlines the connection with the capital-yield plane discussed in the previous sections: in that analogy, the accumulated interest corresponds to the Lorentz boost $\zeta$.
- In the limit for $c \to \infty$, the original Galilean transform is recovered; as such, the original laws still function as an approximation when $V$ is of negligible size with respect to $c$.
- Due to the multiplication factor in the transform, two points $x_1$ and $x_2$ are closer together when travelling at speed than when they are at rest. A length measured in a rest frame are called *proper* , and contracts when in a moving frame: this phenomenon is called *Lorentz contraction* [26].
- In contrast to Galilean transforms, Lorentz transforms are generally not commutative: just like regular three-dimensional rotations, they depend on the order in which they are applied.

**Velocity transform** The Lorentz transform described by eqs. (4-7) and (4-8) shows how to transform coordinates from one frame to another. However, because the transform affects both $x$ and $t$, a velocity measured in the frame (not to be confused with the relative velocity between the frames $V$) $\boldsymbol{v}$ with components will see not only its $x$-component affected, but the other two components $v_x$ and $v_z$ as well. The transformation of $\boldsymbol{v}$ to $\boldsymbol{v'}$ is then:

$$v_x = \frac{v'_x + V}{\sqrt{1 - \left(\frac{V}{c}\right)^2}} \qquad v_y = \frac{v'_y\sqrt{1 - \left(\frac{V}{c}\right)^2}}{1 + v'_x\left(\frac{V}{c}\right)} \qquad v_z = \frac{v'_z\sqrt{1 - \left(\frac{V}{c}\right)^2}}{1 + v'_x\left(\frac{V}{c}\right)}.\text{[26]} \tag{4-9}$$

### 4-3-3 Four-vectors

Instead of the usual three-vectors that are common in classical mechanics, the points in four-dimensional spacetime may be regarded as elements in a four-dimensional vector space instead:

$$A^0 = ct \qquad A^1 = x \qquad A^2 = y \qquad A^3 = z; \tag{4-10}$$

where the superscript indices indicate *contravariant* (vector) components. These can be converted to covariant indices by virtue of the metric tensor $\boldsymbol{g}$

$$g_{ij} = g^{ij} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

Using $\boldsymbol{g}$, indices can then be lowered (or raised) like so

$$A^0 = A_0 \qquad A^1 = -A_1 \qquad A^2 = -A_2 \qquad A^3 = -A_3;$$

such that the spacetime interval may be expressed in tensor notation (observing the Einstein summation convention) as

$$s^2 = A^i A_i = c^2t^2 - x^2 - y^2 - z^2.$$

Much like position, velocities have a four-dimensional spacetime counterpart as well; these objects are called *four-velocities*, they are defined as [26]

$$u^i = \frac{\mathrm{d}x^i}{\mathrm{d}s} \qquad \text{with } \mathrm{d}s = c\,\mathrm{d}t\sqrt{1 - \left(\frac{v}{c}\right)},$$

$v$ being the three-dimensional velocity of the particle. The components of the four-velocity are then

$$u^0 = \frac{1}{\sqrt{1 - \left(\frac{v}{c}\right)^2}} \qquad u^1 = \frac{v_x}{c\sqrt{1 - \left(\frac{v}{c}\right)^2}} \qquad u^2 = \frac{v_y}{c\sqrt{1 - \left(\frac{v}{c}\right)^2}} \qquad u^3 = \frac{v_z}{c\sqrt{1 - \left(\frac{v}{c}\right)^2}}$$

which are all dimensionless quantities. Clearly, any four-velocity squared amounts to one; or $u^i u_i = 1$. This is analogous to the statement that all four-velocities live on a four-dimensional *unit hyperboloid* (due to the nature of the metric tensor); this means that the four-velocities *do not* form a vector space; the sum of two four-velocities does not generally yield another four-velocity. Instead, four-velocities exhibit a special type of geometry called *hyperbolic geometry* — this is an important concept to which chapter 5 is entirely devoted.

## 4-4 The Lorentz metric

Central in the discussion of interest movements as hyperbolic motions will be the so-called Lorentzian inner (or Lorentz) product

**Lorentz product**

Let $\boldsymbol{u}, \boldsymbol{v}$ be vectors in $\mathbb{R}^n$. The *Lorentzian inner product* of $\boldsymbol{u}$ and $\boldsymbol{v}$ is defined to be the real number

$$\boldsymbol{u} \diamond \boldsymbol{v} = u_1 v_1 - u_2 v_2 - \ldots - u_n v_n$$

The Lorentz product is an inner product, which means that it takes two elements from a vector space and returns a scalar value, while satisfying two (or three) conditions: (1) bilinearity, (2) symmetry and (3) nondegeneracy [28].

In some literature, the condition of nondegeneracy is replaced by a stronger notion of positive definiteness, which means that an inner product of a vector with itself is always nonnegative and zero if and only if the vector is the zero vector. However, this condition does not apply to the Lorentz product — this fact has rather far-reaching ramifications, as will become clear with the next definition.

Based on the Lorentz inner product it is natural to define also a corresponding *Lorentzian norm* $\|\cdot\|_L$:

$$\|\boldsymbol{v}\|_L = (\boldsymbol{v} \diamond \boldsymbol{v})^{\frac{1}{2}}$$

Because the Lorentz product is not positive definite but rather indefinite, the result of $\boldsymbol{v} \diamond \boldsymbol{v}$ is not guaranteed to be a positive number. As such, $\|\boldsymbol{v}\|_L \in \mathbb{C}$, in stark contrast with the familiar Euclidean norm which *is* positive definite and will therefore always return a nonnegative real number.

The norm provides a notion of length for a vector. Based on this length, a *metric* yields a distance between two points, i.e. the distance between $\boldsymbol{u}$ and $\boldsymbol{v}$ is equal to

$$d_L(\boldsymbol{v}, \boldsymbol{u}) \triangleq \|\boldsymbol{v} - \boldsymbol{u}\|_L$$

Because the norm is not positive definite but 'only' nondegenerate, it is called a *pseudo-Euclidean metric*, and the vector space it is associated with a *pseudo-Euclidean space*. A vector space equipped with this pseudometric (in more general terms, a bilinear nondegenerate form) is called a Lorentz space. The particular case for $\mathbb{R}^4$ sets the stage for the theory of special relativity (with one 'special' dimension for time and three spatial dimensions) and is called the Minkowski space, after the German physicist Hermann Minkowski [29]. Sometimes the two-dimensional plane that is discovered here is also called the Minkowski plane, because this is what he used to explain his ideas, being unable to draw anything like four-dimensional space. However, this text will adhere to the more mathematically inclined tradition and call it 'Lorentz(ian) space' which applies for any dimension larger than one [28].

The sign of the Lorentz norm gives rise to an equivalence relation $\sim_L$ defined to be $a \sim_L b \iff \operatorname{sgn}\|a\| = \operatorname{sgn}\|b\|$.[2] Therefore, the *quotient set* of all the points in the plane $\mathbb{R}/\sim_L$ contains three elements $\{-1, 0, 1\}$, these equivalence classes are given the respective names $\{\texttt{spacelike}, \texttt{lightlike}, \texttt{timelike}\}$ based on the terminology from special relativity [26].

### About the metric signature

In literature (both physics and mathematics) many variations on the so-called 'metric signature' of the Lorentz product make their appearance. In terms of $\mathbb{R}^4$, these variations come

[2] An equivalence relation is a binary relation that is symmetric, reflexive and transitive.

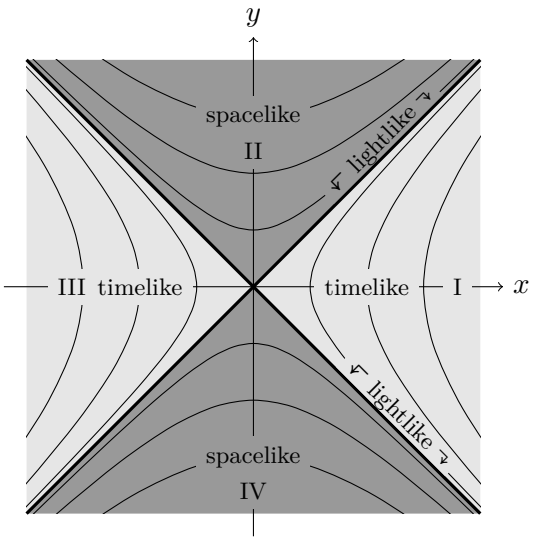



**Figure 4-5:** Overview of the three 'types' of vectors in the Lorentz-Minkowski plane: spacelike ($\|\boldsymbol{v}\|_L < 0$), lightlike ($\|\boldsymbol{v}\|_L = 0$) and timelike ($\|\boldsymbol{v}\|_L > 0$). The lines $y = x$ and $y = -x$ containing all the lightlike vectors form the so-called light cone or null cone. The hyperbola of in the spacelike region (dark gray) obey the equation $y^2 - x^2 = K^2$, they will be referred to the hyperbolic branches II ($y > 0$) and IV ($y < 0$). In contrast, the timelike hyperbolic branches with equation $x^2 - y^2 = K^2$ (light gray region) are referred to as I ($x > 0$) and III ($x < 0$).

down to $(+,-,-,-)$, $(-,-,-,+)$, $(-,+,+,+)$ and $(+,+,+,-)$, where the first one coincides with the definition used here, in accordance with the magnificent work of Landau and Lifshitz [26]. Luckily, this is just a matter of convention and its influence on the mathematical machinery at hand is limited to the switching of some signs.

## 4-5 The Lorentz group

## 4-6 Hyperbolic numbers

Perhaps the most convenient way to view the hyperbolic motions is in terms of so-called hyperbolic numbers[3]. These form an alternative number system similar to complex numbers, based on the 'hyperbolic' unit j with the defining unipotence property $\mathsf{j}^2 = 1$, where obviously $\mathsf{j} \notin \mathbb{R}$. Like complex numbers, the hyperbolic numbers can have a real and a hyperbolic part

$$z = x + y\mathsf{j} \qquad x, y \in \mathbb{R}$$

Combined with addition and multiplication defined on them, the hyperbolic numbers form a commutative ring. Each hyperbolic number $z$ is associated with its *hyperbolic conjugate*

[3]The hyperbolic number system has been assigned a myriad of names, with different terminology and mathematical notation for almost every influential paper that has been published about them. Among others, hyperbolic numbers are referred to as split-complex numbers, double numbers, perplex numbers, algebraic motors, etc. In this text, 'hyperbolic numbers' is chosen to highlight their connections with hyperbolae. For the choice of the hyperbolic unit, j will be used, though in literature also u (for unipotent) and h (for hallucinatory or hyperbolic) make their appearance [30, 31, 32, 23].

$z^* = x - y\mathrm{j}$. The product of a hyperbolic number with its own hyperbolic conjugate creates a quadratic form $zz^* = x^2 - y^2$, which always returns a real number and is equivalent Lorentzian metric. Again, three cases can be distinguished for $z$:

- *timelike* when $zz^* > 0$
- *lightlike* when $zz^* = 0$
- *spacelike* when $zz^* < 0$

However, in contrast to complex numbers (whose quadratic form would be $x^2 + y^2$), this quadratic form is *isotropic*, which means that there exists a $z \neq 0$ such that $zz^* = 0$ — this is precisely the case for all the $z$ on the light cone in the hyperbolic plane. The *hyperbolic modulus* therefore requires an absolute value in order like so

$$|z| \triangleq \sqrt{|zz^*|}$$

which considered to be the hyperbolic distance from $z$ to the origin [31].

Again, in correspondence with complex numbers, the hyperbolic numbers also have a *polar form*, i.e.

$$z = K\exp(\zeta\mathrm{j})$$

which can be evaluated using the Taylor expansion of the exponential exp:

$$z = K\exp(\zeta\mathrm{j}) = K\sum_{k=0}^{\infty}\frac{(\zeta\mathrm{j})^k}{k!} = K\sum_{k=0}^{\infty}\frac{(\zeta\mathrm{j})^{2k}}{(2k)!} + K\sum_{k=0}^{\infty}\frac{(\zeta\mathrm{j})^{2k+1}}{(2k+1)!} = K\cosh(\zeta) + K\mathrm{j}\sinh(\zeta)$$

when the hyperbolic angle $\zeta$ is associated with hyperbolic branches I and III. For branches II and IV, one essentially has to consider $z = K\mathrm{j}\exp(\zeta\mathrm{j})$ or

$$z = K\sinh(\zeta) + K\mathrm{j}\cosh(\zeta)$$

Furthermore, consider the *conjugate product* of two hyperbolic numbers $z_1 = x_1 + y_1\mathrm{j}$ and $z_2 = x_2 + y_2\mathrm{j}$:

$$z_1^* z_2 = \underbrace{(x_1x_2 - y_1y_2)}_{\text{inner product}} + \underbrace{(x_1y_2 - x_2y_1)}_{\text{outer product}}\mathrm{j}$$

of the resulting expression, the real part is called the inner product and the hyperbolic part the outer product. The inner product is equivalent to the Lorentz product while the outer product yields the directed area of the parallelogram spanned by $z_1$ and $z_2$ (or the determinant of $\begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix}$. The outer product is the same as for regular complex numbers, while the inner product recovers the Lorentz inner product instead of the Euclidean inner product. This is why one can say that the hyperbolic number plane and the complex plane have an identical notion of area. Based on the inner product, one can however recognize a different notion of orthogonality, i.e. $z_1$ and $z_2$ are *hyperbolically orthogonal* if their hyperbolic inner product equals zero — this will yield a different notion of orthogonality than their Euclidean counterpart [2, 31].

### 4-6-1 Matrix representation

For many algebraic structures (groups, rings, fields, ...) an isormorphism can be found in the rhealm of linear algebra, by identifying a specific class of matrices. For example, the complex number system $a + b\mathrm{i}$ is isomorphic to the matrices

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} \quad a, b \in \mathbb{R}.$$

Likewise, one can establish an isomorphism between the hyperbolic numbers and a certain class of matrices. The hyperbolic number system defined with addition and multiplication defined on it are ring-isomorphic to the matrix ring

$$\begin{pmatrix} x & y \\ y & x \end{pmatrix} \quad x, y \in \mathbb{R}$$

under matrix addition and matrix multiplication. The determinant of said matrix then recovers the Lorentz metric

$$\det \begin{pmatrix} x & y \\ y & x \end{pmatrix} = x^2 - y^2$$

### 4-6-2 The idempotent basis

Previously it was already mentioned that the hyperbola can be viewed conveniently in two particular axis systems, the standard axis system and the idempotent axis system, as illustrated in fig. 4-2. This basis can also be defined in terms of hyperbolic numbers; in terms of the standard basis $\{1, \mathrm{j}\}$ the idempotent basis is $\{\mathrm{j}_+, \mathrm{j}_-\}$ with

$$\mathrm{j}_+ = \frac{1}{2}(1 + \mathrm{j}) \qquad \mathrm{j}_- = \frac{1}{2}(1 - \mathrm{j})$$

The term 'idempotent' is a testament to the fact that $\mathrm{j}_+^2 = \mathrm{j}_+$ and $\mathrm{j}_+^2 = \mathrm{j}_+$. As such, a given hyperbolic number $z = x + y\mathrm{j}$ in idempotent coordinates is

$$\underbrace{(x+y)}_{z_+}\mathrm{j}_+ + \underbrace{(x-y)}_{z_-}\mathrm{j}_+$$

Furthermore, the idempotent basis is mutually annihilating, i.e. $\mathrm{j}_+\mathrm{j}_- = 0$, which is why they possess a projective property: [31]

$$z\mathrm{j}_+ = z_+\mathrm{j}_+ \quad \text{and} \quad z\mathrm{j}_- = z_-\mathrm{j}_-$$

The idempotent axis system is interesting because it returns the total amount compounded or discounted over time; in contrast to the standard yield-capital decomposition.

# Chapter 5

# Hyperbolic geometry

## 5-1 Basic facts

Hyperbolic geometry, sometimes called Lobachevskian geometry[1] is the geometry on the hyperbolic plane. This type of geometry is called non-Euclidean, because it does not satisfy all of Euclid's axioms that form the building blocks of traditional geometry. More specifically, the fifth axiom, called the parallel axiom, states that [2]

> Through any point $p$ not on the line $L$ there exists precisely one line $L'$ that does not meet $L$.

Logic dictates that if this axiom is not true, two possible alternatives arise. Given again the point $p$ and the line $L$,

- there exists *no* line through $p$ that does not meet $L$, or
- there exist *at least two* lines through $p$ that do not meet $L$.

The first statement corresponds to what is called *spherical geometry*, while the latter is the defining axiom for *hyperbolic geometry*. It does not take too much imagination to realize that the parallel axiom is equivalent to another basic fact in Euclidean geometry: the sum of the angles of a triangle is equal to $\pi$. As such, for both types of non-Euclidean geometry this will not be the case; let $\mathcal{E}(T)$ denote the *angular excess* of a triangle $T$ in a certain geometry.

- Naturally, for Euclidean geometry $\mathcal{E}(T) = 0$,
- in spherical geometry, $\mathcal{E}(T) > 0$,
- in hyperbolic geometry, $\mathcal{E}(T) < 0$.

[1] After the Russian mathematician Nikolai Lobachevsky (1832) who first published about this subject [2].

It may come as a surprise that the angular excess can be related to the size of the triangle (more specifically, its area) like so [2]

$$\mathcal{E}(T) = k\mathcal{A}(T) \tag{5-1}$$

$k$ where $k = 0$ for Euclidean geometry, $k < 0$ for hyperbolic geometry and $\mathcal{K} > 0$ for spherical geometry. This hints at the fact that spherical geometry and hyperbolic geometry are somehow 'larger' classes of geometry than Euclidean geometry, since they exist for a whole range of values for $k$, be it positive or negative. There is indeed a 'different' geometry for every value of $k$, and only one of those values corresponds to the traditional concept of Euclidean geometry that most people are familiar with. Another consequence of eq. (5-1) is that similar triangles cannot exist in non-Euclidean geometry; since apart from the trivial case where the triangles are also congruent, they must differ in area, which yields a different angular excess. The value of $k$ is, as it turns out, equal to the *Gaussian curvature* of the surface: a negatively curved surface has an angular deficiency while a positively curved surface has an angular excess; only for surface with zero curvature the sum of the angles of a triangle will be precisely equal to $\pi$.

### 5-1-1 Surface curvature

Perhaps one of the most remarkable results attributed to Carl Friedrich Gauss is his *Theorema Egregium*[2] about the (Gaussian) curvature of surfaces. The theorem states that curvature is an *intrinsic* property, which means that it remains preserved when the surface is transformed by 'bending without stretching'. Curvature can be positive, negative or zero. Positively curved surfaces 'bend away' from their tangent plane in any direction (local convexity) and negatively curved surfaces intersect their tangent plane like a saddle. For surfaces with zero curvature, there is always at least one straight line that lies in the tangent plane; examples are a cylinder (always one straight line) and a plane (straight lines in two directions). If a coordinate system is defined such that the point at issue is at the origin and the tangent plane to the surface described by $f(x, y)$ coincides with the horizontal, the Gaussian curvature can be computed by means of the determinant of the Hessian [34, 35]

$$k = \det \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix}.$$

Surface curvature can be approached more rigorously from the perspective of Riemannian geometry. The core concept behind Riemannian geometry are the eponymous manifolds: these are manifolds equipped with an positive definite inner product[3] on their tangent space, hence providing a *metric (tensor)* that allows to determine lengths, angles and curvature. The metric on a parametric surface $\boldsymbol{r}(u, v)$ can be written in terms of the so-called *first fundamental form*[4] $\mathbf{I}$, which is defined as the inner product of two tangent vectors to the

[2] Loosely translated by John M. Lee [33] as the *Totally Awesome Theorem*.

[3] In contrast to section 4-4, the positive definiteness of the inner product is essential for Riemannian geometry. When the inner product is not positive definite but nondegenerate, as the Lorentz product, the manifold is called *pseudo-Riemannian*.

[4] Unfortunately, the first (and second) fundamental forms are not forms in the modern mathematical sense of the word; this terminology is merely inherited from older works [36].

surface and is characterised by coefficients $E$, $F$ and $G$.

$$\begin{aligned} E &= \boldsymbol{r}_u \cdot \boldsymbol{r}_u \\ F &= \boldsymbol{r}_u \cdot \boldsymbol{r}_v \\ G &= \boldsymbol{r}_v \cdot \boldsymbol{r}_v \\ \text{with } \boldsymbol{r}_u &= \frac{\partial \boldsymbol{r}}{\partial u},\ \boldsymbol{r}_v = \frac{\partial \boldsymbol{r}}{\partial v}. \end{aligned} \tag{5-2}$$

$E$ $F$ $G$ $e$ $f$ $g$ e The components of the metric tensor $\boldsymbol{g}$ then coincide with the components of the first fundamental form: $g_{11} = E$, $g_{12} = g_{21} = F$ and $g_{22} = G$. Therefore, the first fundamental form determines the length of curves lying in the surface.

In contrast, the *second fundamental form* $\mathbf{II}$ provides information on the curvature or shape of the embedded surface, more specifically the rate of change of the tangent planes in any direction. Again, it is characterised by three coefficients $e$, $f$ and $g$ like so

$$e\,\mathrm{d}u^2 + 2f\,\mathrm{d}u\,\mathrm{d}v + g\,\mathrm{d}v^2\,.$$

The coefficients of the second fundamental form can be computed using the surface unit normal vector

$$\boldsymbol{n} = \frac{\boldsymbol{r}_u \times \boldsymbol{r}_v}{\|\boldsymbol{r}_u \times \boldsymbol{r}_v\|}, \tag{5-3}$$

the coefficients are then given by

$$\begin{aligned} e &= \boldsymbol{r}_{uu} \cdot \boldsymbol{n} \\ f &= \boldsymbol{r}_{uv} \cdot \boldsymbol{n} \\ g &= \boldsymbol{r}_{vv} \cdot \boldsymbol{n} \\ \text{with } \boldsymbol{r}_{uu} &= \frac{\partial^2 \boldsymbol{r}}{\partial u^2} \ \text{ etc.} \end{aligned} \tag{5-4}$$

Finally, the first and second fundamental form can then be used to determine the Gaussian curvature of the parametric surface: [35, 36]

$$k = \frac{\det(\mathrm{II})}{\det(\mathrm{I})} = \frac{eg - f^2}{EG - F^2}. \tag{5-5}$$

To summarize, the connection between hyperbolic geometry and Gaussian curvature is concisely stated again below.

1. Hyperbolic geometry is the geometry with an alternative parallel axiom, where there are always at least two parallel lines for a given line through a point.
2. The former fact is equivalent to stating that the angular excess $\mathcal{E}$ is characterised by the negative constant $k$ (constant over the entire hyperbolic plane).
3. The Gauss-Bonnet theorem states that the constant $k$ is given by the Gaussian curvature.

Therefore, in order to represent hyperbolic geometry, one whishes to find a surface that exhibits constant negative Gaussian curvature. Some candidates exist for this criterion, the simplest of which is the so-called *pseudosphere*.

### 5-1-2 The pseudosphere

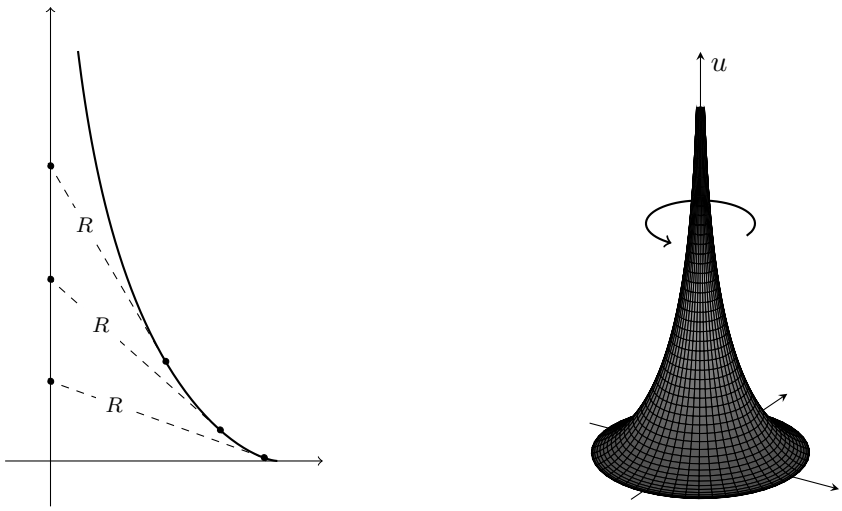
Pseudospherical surfaces are surfaces of constant negative curvature, in that sense exhibiting a certain duality to a sphere which is a surface of constant positive curvature. The most notable example is the eponymous *pseudosphere*; which is the surface of revolution of a tractrix — another name for the pseudosphere is a *tractricoid*. A tractrix is a curve defined by the property that the segment of the tangent from the point to the axis has constant length $p$. Figure 5-1a shows an example of such a curve. A parametric representation of the tractrix $\gamma$ is given by

$$\gamma : u \mapsto (p\,\mathrm{sech}(u),\ p(u - \tanh(u))) \quad u \in \mathbb{R}^+. \tag{5-6}$$

A parameterization of the pseudosphere in terms of $u$ and $v$ is then easily obtained by introducing a second parameter $v$ that indicates the angle of rotation around the vertical axis:

$$\boldsymbol{r}(u,v) = \begin{pmatrix} p\,\mathrm{sech}(u)\cos(v) \\ p\,\mathrm{sech}(u)\sin(v) \\ p(u - \tanh(u)) \end{pmatrix} \quad u \in \mathbb{R}^+,\ v \in [0, 2\pi). \tag{5-7}$$

A pseudosphere is visualized by fig. 5-1b; the mesh on the surface are isoparametric curves for $u$ and $v$.




**(a)** The tractrix - the line segments defined on the tangent to the curve between the curve and the intersection with the vertical axis all have the same length $R$.

**(b)** The pseudosphere - the horizontal lines are maintain a constant value for $v$, while the vertical lines are for a constant $u$.

**Figure 5-1**

#### Curvature of the pseudosphere

As explained in section 5-1-1, the Gaussian curvature of a parametric surface may be calculated using the first and second fundamental form. It was already alluded that the pseudosphere has a constant negative Gaussian curvature, which will be formally shown in this

section. Let the pseudosphere be parameterized as eq. (5-7). Then, by virtue of eq. (5-2), the coefficients of the first fundamental form $E$, $F$ and $G$ are

$$\begin{aligned} E &= p^4 \tanh^2(u) \\ F &= 0 \\ G &= p^4 \operatorname{sech}^2(u) \end{aligned}$$

From eq. (5-3), the unit normal vector is

$$\boldsymbol{n} = \begin{pmatrix} -\cos(v)\tanh^{-1}(u) \\ -\sin(v)\tanh(u) \\ -\cosh^{-1}(u) \end{pmatrix}$$

and, consequently, the coefficients of the second fundamental form are

$$\begin{aligned} e &= -p\operatorname{sech}(u)\tanh(u) \\ f &= 0 \\ g &= p\operatorname{sech}(u)\tanh(u) \end{aligned}$$

using eq. (5-4). The Gaussian curvature can then be computed to be [35]

$$k = \frac{\det(\mathrm{II})}{\det(\mathrm{I})} = \frac{eg - f^2}{EG - F^2} = \frac{-p^2\operatorname{sech}^2(u)\tanh^2(u)}{p^4\operatorname{sech}^2(u)\tanh^2(u)} = -\frac{1}{p^2}. \tag{5-8}$$

As such, the pseudosphere is shown to have a constant negative curvature everywhere but on the rim (where differentiability is lost).

Because the pseudosphere has constant negative curvature, it shares many properties of the hyperbolic plane. For example, the angular excess equation eq. (5-1) holds on the pseudosphere [2]. But, whereas a sphere can serve as a 'globe' for spherical geometry, the pseudosphere cannot be used as a representative for the entire hyperbolic plane. First of all, they are not homeomorphic (the pseudosphere is homeomorphic to a cylinder, which has a different fundamental group than the hyperbolic plane[5]). Secondly, the pseudosphere has a 'rim' at the bottom which prevents any line segment from extending further downwards — naturally, no such thing exists in the hyperbolic plane. As such, the pseudosphere can only serve as a model for finite regions of the hyperbolic plane, but not in its entirety. In technical terms, this means that the pseudosphere is *locally isometric* to the hyperbolic plane [39]. This is why one must resort to models of the hyperbolic plane, they will be discussed in the next section.

## 5-2 Models of the hyperbolic plane

The pseudosphere cannot serve as a model for the hyperbolic plane. In fact, a theorem attributed to David Hilbert shows that, *in Euclidean three-space, there can be no complete*

[5] Intuitively, one can see this as the number of different ways it is possible to draw loops on the surface which are distinct up to a homotopy. On the pseudosphere and the cylinder, loops can 'wrap' around the vertical axis any integer number of times, while on the (hyperbolic) plane all loops are homotopic [37]. This concept is defined more rigorously in terms of the *fundamental group* $\pi_1$, a heavily used invariant of topological spaces [38].

*smooth surface with the intrinsic geometry of the pseudosphere* [34]. This is the reason why there can only be models of the hyperbolic plane. One could say that this means 'double trouble' for make-believe cartographers of the hyperbolic plane. Of course, there is the traditional problem that normal cartographers also face, in that the Earth can never be completely mapped on a flat surface, which is why either (1) a suitable projection method must be used or (2) they must resort to an additional dimension an make a globe to obtain a completely faithful representation. Hilbert's theorem implies that even the additional dimension will be to no avail for the hyperbolic plane, as illustrated by the pseudosphere.

Several models for the hyperbolic plane have been devised in the past, and they will provide a lot more mileage than the pseudosphere alone. The most popular models will be discussed in the following sections: there is a natural way to map the pseudosphere and the Poincaré half plane, which will be discussed first. Then, via the so-called *Cayley transformation*, the half plane can be mapped to the Poincaré disk, perhaps the most illustrous model of them all. Using the disk model, one can naturally arive at the Cayley-Klein disk and the hyperboloid model — the latter will be the starting point in the financial analogy.

### 5-2-1 Poincaré half plane

It has been pointed out that there are essentially two incompatibility problems with the pseudosphere that prevent devise a mapping between it and the complete hyperbolic plane (also, Hilbert's theorem immediately shatters any aspiration to find one): first, it is homeomorphic to the cylinder and secondly, it has an edge. This section will describe a conformal mapping between the surface of the pseudosphere and a 'half plane', which can serve as a global model for the hyperbolic plane. However, the theorem by Hilbert already hints at the fact that this plane will have some 'weird' properties, more specifically, a different notion of distance.

As described by Needham [2], one can imagine the pseudosphere to be cut open along any tractrix; the resulting surface can then be 'unfolded' in the horizontal direction as if it were a treasure map on a table. The edges that were cut can be extended towards infinity by recognizing the periodicity that exists naturally on the pseudosphere itself; a particle traveling horizontally would wrap once around the pseudosphere every distance of $2\pi$ traveled. Consequently, the circles on the pseudosphere that arise for constant values of $z$ will be mapped to horizontal lines in the half plane. Formally, this simply means that the horizontal axis of the half plane $\tilde{x}$ is equal to the angle $u$. It has already been mentioned that the mapping between the pseudosphere and the half plane is *conformal*, i.e. it locally preserves angles. Therefore, since the tractrix lines are everywhere perpendicular on the pseudosphere to the circles for constant $z$, they must map to vertical lines in order to maintain this orthogonality. A horizontal movement in the half plane $\mathrm{d}\tilde{x}$ therefore corresponds on the pseudosphere with a traveled distance of $\mathrm{d}\tilde{s} = \operatorname{sech}(v) = \operatorname{sech}(v)\,\mathrm{d}v$, because of the radius of the circle at that particular height. Because the mapping is conformal, a movement along a tractrix $\mathrm{d}\sigma$ must be scaled by the same factor:

$$\mathrm{d}\sigma = \operatorname{sech}(u)\,\mathrm{d}y\,.$$

Subsequently, the movement along the tractrix can be written in terms of $u$ and $v$ using the parameterization:

$$\mathrm{d}\sigma = \left\| \frac{\partial \boldsymbol{r}}{\partial u}\,\mathrm{d}u \right\| = \sqrt{\mathrm{d}u^2 - \operatorname{sech}^2(u)\,\mathrm{d}u^2}$$

Combining this with the previous expression found for $\mathrm{d}\sigma$, one arrives at

$$\mathrm{d}y = \sinh(u)\,\mathrm{d}u \implies y = \cosh(u) + C. \tag{5-9}$$

Thus, the conformality of the mapping imposes a restriction on the mapping for $y$ up to a constant $C$, which is usually taken to be 0 [2]. Using this information, the metric of the Poincaré half plane can also easily be deduced by *pushing forward*[6] the Euclidean metric of the pseudosphere:

$$\mathrm{d}s = \left\| \frac{\partial \boldsymbol{r}}{\partial u}\,\mathrm{d}u + \frac{\partial \boldsymbol{r}}{\partial v}\,\mathrm{d}v \right\| = \operatorname{sech}(u)\sqrt{\sinh^2(u)\,\mathrm{d}u^2 + \mathrm{d}v^2} = \frac{\sqrt{\mathrm{d}x^2 + \mathrm{d}y^2}}{y} \tag{5-10}$$

This metric is called the *Poincaré metric*. Equation (5-9) already states that for $u = 0$, i.e. the rim of the pseudosphere, maps to the horizontal line $y = 1$. This suggests that the pseudosphere is covered by the region of the half plane for which $y \geq 1$.

The Poincaré metric suggests that distances get larger and larger when travelling downwards in $y$-direction. At the line $y = 0$ they even become infinitely large! For someone living in the half plane, this line would never be reachable, as they would have to travel for an infinite amount of time. It is not part of the half plane itself, which is why it is called the *horizon*; points on the horizon are named *ideal points* [2].

Geodesics in the Poincaré half plane connecting two points are either straight vertical lines if the two points have the same $x$-coordinate. Otherwise, the geodesic is the arc defined by the portion of the circle going through each of the two points that also intersects the $x$-axis at a right angle [40]. Apart from the horizon, two other peculiar curves exist in the half plane that will also show up in later models:

- *horocycles*: circles tangent to the $x$-axis or horizontal lines,
- and *hypercycles*: circular arcs that intersect the $x$-axis at non-right angles or straight lines that intersect the $x$-axis at a non-right angle.

### 5-2-2 Poincaré disk

The Poincaré disk arises naturally from the Poincaré half plane by virtue of the Cayley transform. When the Poincaré half plane is viewed 'on top' of the complex plane, the Cayley transform is defined as [source?]

$$D(z) = \frac{\mathrm{i}z + 1}{z + \mathrm{i}} \quad z \in \mathbb{C} \tag{5-11}$$

which maps the entire half plane to the unit disk. Furthermore,

[6] It must be clearly noted that the term *pushforward metric* is not at all in place, and its usage here would in fact be quite pathological. The reason is as follows: the relation sketched here is not even strictly a mapping, as any point on the pseudosphere maps to an infinite amount of points on the half plane. Beyond this problem of the unique images there is also the obvious fact that their is no preimage for the part of the half plane where $y < 1$ (non-surjective). As such, it would definitely make more sense to define the Poincaré metric first and then define the *pullback metric* on the pseudosphere. However, since the pseudosphere exists in the familiar Euclidean space, it is more natural to introduce the pseudosphere first [38].

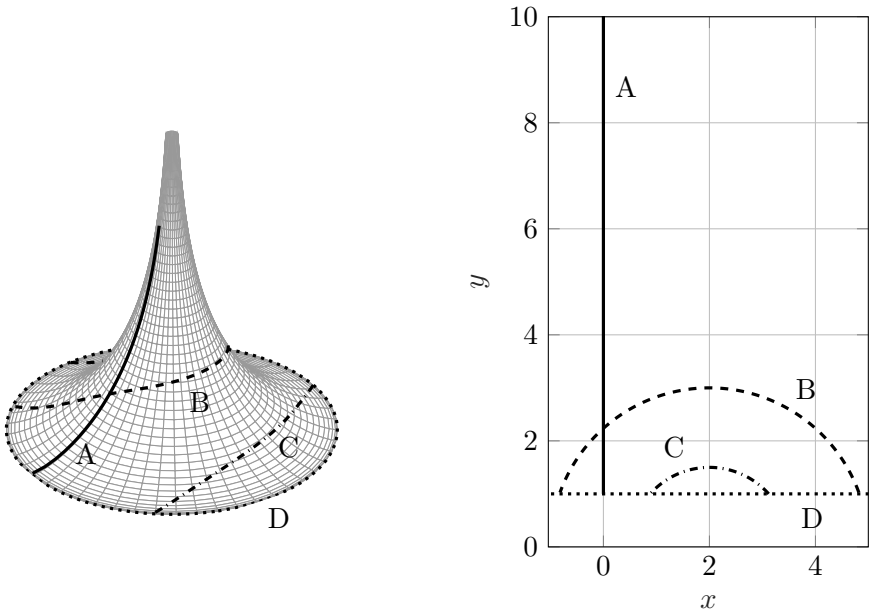



**Figure 5-2:** Comparison of various trajectories on the pseudosphere (left) and the Poincaré half plane (right). Lines $A$, $B$ and $C$ are geodesics. The endpoints of $B$ are closer together because it covers a wider range on the $x$-axis — if it would be larger than $2\pi$, the curve would show an entire encirclement of the pseudosphere. Line $D$ corresponds to the rim of the pseudosphere and the line $y = 1$ in the pseudosphere.

- the horizon coincides with the rim of the disk,
- the points ±1 are invariant,
- i maps to the origin,
- the origin corresponds to –i

As will become clear in chapter 6, eq. (5-11) belongs to the class of Möbius transforms, which means that it also must be conformal. Therefore, angles on the pseudosphere, half plane and disk will all look alike. Since the horizon containing the ideal points is now the rim of the unit disk, it makes sense that the pullback metric based on eq. (5-10) will have the $y$ replaced by the distance from the rim of the unit disk, which is $1 - x^2 - y^2$. Indeed, the halfplane metric can be written as in terms of the complex variable $z$, using the mapping eq. (5-11) the metric in the disk turns out to be [2]

$$\mathrm{d}s = \frac{|\mathrm{d}z|}{\mathfrak{Im}(z)} = \frac{2|\mathrm{d}\tilde{z}|}{1 - |z|^2} \qquad \text{with} \quad \tilde{z} = D(z).$$

In $x - y$ coordinates, this metric is

$$\mathrm{d}s = \frac{2\sqrt{\mathrm{d}x^2 + \mathrm{d}y^2}}{1 - x^2 - y^2}. \tag{5-12}$$

It has been pointed out that the Cayley transform is a member of a larger class called Möbius transforms, which will be elaborated upon in chapter 6. Apart from being conformal, Möbius

transforms also preserve circles (i.e. the result of a a Möbius transform applied to a circle will again be a circle). The geodesics in the half plane have the either the shape of straight lines extending from the horizontal axis, or circles that cross the horizontal axis at a right angle. Consequently, it is straightforward to deduce that the geodesics in the Poincaré disk are also circles that 'start' orthogonally to the rim of the unit disk. Additionally, all diameters of the Poincaré disk are geodesics as well: they can be interpreted as circles with infinite radius that are obviously also orthogonal to the rim of the disk.

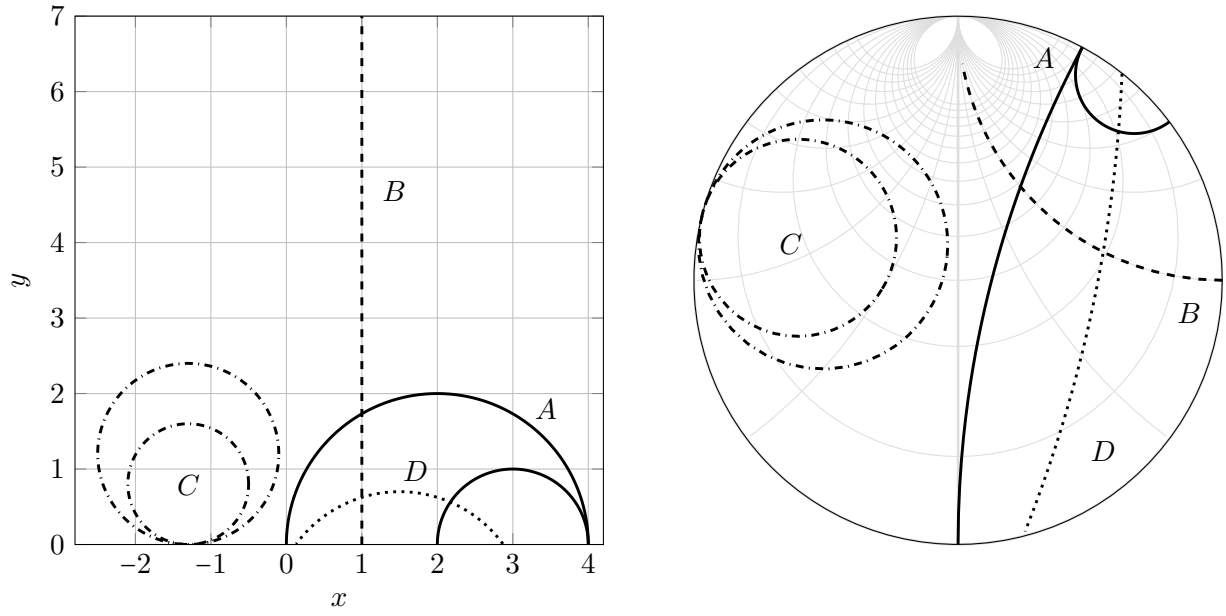



**Figure 5-3:** Comparison between trajectories in the Poincaré half plane (left) and the Poincaré disk (right). Several types of trajectories are shown: the solid lines $A$ are 'typical' geodesics, i.e. circles with finite radius in the half plane. The dotted line $B$ is also a geodesic but never reaches the horizon at its endpoint (which would take an infinite distance), this is also clearly visible in the disk. Clearly, the origin in the half plane maps to –i. The dashdotted lines $C$ are horocycles; they preserve their shape under the action of the Cayley transform. Dotted line $D$ is a hypercycle; a circle that crosses the horizon at an oblique or acute angle.

### The Smith chart

## 5-2-3 Hyperboloid model

Hilbert's theorem about the embedding of a hyperbolic surface in Euclidean space eradicates any hope to find a 'globe' for the hyperbolic space, like a regular sphere would be for an elliptic space. However, as strong as it may seem, Hilbert's theorem still leaves some room for other possiblities because it is very explicit about the nature of the space in question: it must be Euclidean. Chapter 4 already introduced a notorious alternative: the Lorentz space.

It will probably not come as a surprise that the Lorentz space does allow for the embedding of a hyperbolic surface, which is known as the *hyperboloid model*.

Now, why would one find any solace in the Lorentz plane specifically? Recall that the curvature of a surface can be expressed in terms of its local radius, i.e. $1/p^2$. A negative curvature therefore implies that $p$ is somehow not a real but imaginary number. Obviously, this can never make any sense in an Euclidean space, but the Lorentzian space perfectly allows this kind of odd witchcraft, as illustrated by fig. 4-5 [28]. The complete hyperboloid consists of two sheets, a positive sheet and a negative sheet, implicitly described by the following set:

$$H^2 = \left\{ \boldsymbol{x} \in \mathbb{R}^3 \mid \|\boldsymbol{x}\|_L = p \right\}.$$

However, on account of the positive and the negative sheet, this set is clearly *disconnected*[7]. It is therefore common to discard the negative sheet of $H^2$, the hyperboloid model of the hyperbolic 2-space is then [28]

$$H^2_+ = \left\{ \boldsymbol{x} \in \mathbb{R}^3 \mid \|\boldsymbol{x}\|_L = p,\ x_1 > 0 \right\}.$$

A parametric representation of $H^2_+$ is the following

$$\boldsymbol{r} = \begin{pmatrix} p\cosh(u) \\ p\cos(v)\sinh(u) \\ p\sin(v)\sinh(u) \end{pmatrix}. \tag{5-13}$$

To show that this surface has indeed a constant negative Gaussian curvature, a slightly more general approach must be followed than for the pseudosphere, since the calculations with the second fundamental form are restricted to a Euclidean space. As described by O'Neill [41]

$$E = \partial_u \diamond \partial_v \quad F = \partial_u \diamond \partial_v \quad G = \partial_u \diamond \partial_v$$

which correspond to the components of the first fundamental form, generalized to the Lorentz product. The coordinate system $u, v$ is called orthogonal if $F$ vanishes, which is indeed the case for the parametrization given by eq. (5-13). The Gaussian curvature of the surface is then given by [41]

$$k = \frac{-1}{eg}\left[ \varepsilon_1 \left( \frac{g_u}{e} \right)_u + \varepsilon_2 \left( \frac{e_v}{g} \right)_v \right] \tag{5-14}$$

with $\varepsilon_1 = \operatorname{sgn} E$, $\varepsilon_2 = \operatorname{sgn} G$, $e = \sqrt{|E|}$ and $G = \sqrt{|G|}$. Performing these calculations on eq. (5-13) yields indeed that $k = -1/p^2$, confirming that the one-sheet hyperboloid embedded in three-dimensional *Lorentz* space indeed has the same curvature as the pseudosphere[8].

#### Relation with Poincaré disk

#### Lorentz group

### 5-2-4 Cayley-Klein disk

$H^2$ $H^2_+$ $\mathbb{R}$ $\mathbb{C}$

[7] A space is called connected if it is impossible to find two open sets whose union is equal to that space; which in this case clearly is not the case (the open sets are the positive and negative sheet of the hyperboloid).

[8] Some authors even refer to the hyperboloid as 'pseudosphere' as well, e.g. Balazs and Voros [42]. However, this is deemed confusing in the context of the larger discussion presented here, the author refrained from using it here.

# Chapter 6

# Möbius transforms

## 6-1 Definition and basic properties

Many of the transforms and mappings discussed in the preceding chapters may be expressed as a Möbius transformation. These are mappings of the form [2]

$$\mathfrak{M}(z) = \frac{az + b}{cz + d} \tag{6-1}$$

where $a, b, c, d \in \mathbb{C}$ are constants. Möbius transformations carry a deep connection with hyperbolic geometry (and non-Euclidean in general), and Einsteins theory of relativity. As such, the connection with the financial interpretation of the Lorentz plane can be readily made. As stated by Needham [2], the complex mappings that correspond to a Lorentz transformation are Möbius transformations and vice versa — every Möbius transform corresponds to a unique Lorentz transformation.

The Möbius transformation $\mathfrak{M}$ is called *singular* if $ad - bc = 0$, which maps every point to the same image $a/c$. In general, any Möbius transformation can be decomposed into four elementary transformations:

1. A translation $z \mapsto z + \frac{d}{c}$
2. A *complex inversion* $z \mapsto \frac{1}{z}$
3. An expansion and rotation $z \mapsto -\frac{ad-bc}{c^2} z$
4. A translation $z + \frac{a}{c}$

Each of these transformations are conformal and preserve circles which is why general Möbius transformations inherit these vital properties as well.

It is quite clear from eq. (6-1) that multiplication of both the denominator and the numerator by the same constant $k$ will not affect the result of the mapping. Therefore, a Möbius transformation is uniquely determined by only three quantities $a/b, b/c$ and $c/d$. This ambiguity allows for the notion of *normalized transformations*, for which $ad - bc = 1$.

If $\mathfrak{M}$ is nonsingular, the transformation is bijective; the inverse transformation is then given by [2]

$$\mathfrak{M}^{-1}(z) = \frac{dz - b}{-cz + a}$$

## 6-2 Group structure

It is neither surprising nor hard to show that the nonsingular Möbius transformations form a group under composition; the identity mapping is a Möbius transformation, the inverse transformation is also a Möbius transformation as illustrated by the expression for $\mathfrak{M}^{-1}$ stated above and that the composition of two transformations $\mathfrak{M}_2 \circ \mathfrak{M}_1$ again yields a member of the class of Möbius transformations [2].

### 6-2-1 The Riemann sphere

The Möbius group Möb is the automorphism group of the Riemann sphere. Being the simplest compact Riemann surface, the Riemann sphere is a special representation of the extended complex plane[1] $\hat{\mathbb{C}}$ in three-dimensional space: it can be visualized by placing the complex plane horizontally and considering a unit sphere centered around the origin, such that its intersection coincides with the unit disk of the complex plane. To map the plane to the sphere, a stereographic projection is is used from the 'north pole' of the sphere. As such, anything inside the unit disk is mapped to the southern hemisphere, everything on the unit disk is mapped onto itself (since it lies at the intersection) and everything outside the unit disk lies on the northern hemisphere. The north pole of the Riemann sphere then coincides with the distinctive feature of the extended complex plane, namely the point at infinity $\infty$ [2]. Stated more formally:

$$\text{Möb} = \text{Aut}(\hat{\mathbb{C}}) \quad \hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}.$$

In order to turn towards the computational aspect of the Riemann sphere, one must consider a different coordinate system for the complex plane; namely the *homogeneous* or *projective* coordinates. These consist of an ordered pair of complex numbers[2] , i.e. an element of $\mathbb{C}^2/\{[0,0]\}$, denoted as $[\mathfrak{z}_1, \mathfrak{z}_2]$. These two complex numbers are subject to an equivalence relation that makes them projective coordinates, in that

$$[\mathfrak{z}_1, \mathfrak{z}_2] \sim [\mathfrak{w}_1, \mathfrak{w}_2] \iff \mathfrak{w}_1 = \lambda \mathfrak{z}_1 \text{ and } \mathfrak{w}_2 = \lambda \mathfrak{z}_2$$

with $\lambda$ any nonzero complex number. The projective space is then formed by the quotient set of the complex 2-space (excluding the origin) and this equivalence relation. Every equivalence class can be uniquely identified with a point $[1, \mathfrak{z}_2/\mathfrak{z}_1]$, which then corresponds to a point on the The Riemann sphere (or equivalently on the extended complex plane). The only point where this mapping breaks down is the point $[0, 1]$, which can be associated with the north pole on the Riemann sphere, or the infinity point. In technical terms, this is called a one-point

[1] The complex plane combined with a value for infinity '$\infty$'.

[2] In physics, these objects are also called *2-spinors* which are also fundamentally connected with the theory of relativity, as demonstrated by [43].

compactification of a plane into a sphere, with the desirable property that the infinity point *has no special meaning on the sphere*, which it inevitably has in a plane. The Riemann sphere is therefore equal to the one-dimensional complex projective space $\mathbb{CP}^1$ [34].

### 6-2-2 Matrix representation

A very powerful property of Möbius transforms is that they can each be associated to a $2 \times 2$ complex matrix like so:

$$\frac{az+b}{cz+d} \quad \leftrightsquigarrow \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Because the transformations are only defined up to multiplication of a constant, so is associated matrix. However, one can assume to have the transformation normalized, i.e. $ad - bc = 1$ which is equivalent to restricting the matrix to have a unit determinant. When normalized, there are precisely two matrices corresponding to every Möbius transform, since multiplying the entire matrix with -1 would still yield the same result. Now, how can one actually effect a Möbius transformation using this matrix representation? This is where the projective coordinates come into play. As it turns out,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} \mathfrak{z}_1 \\ \mathfrak{z}_2 \end{pmatrix} = \begin{pmatrix} a\mathfrak{z}_1 + b\mathfrak{z}_2 \\ c\mathfrak{z}_1 + d\mathfrak{z}_2 \end{pmatrix} = \begin{pmatrix} a(\mathfrak{z}_1/\mathfrak{z}_2) + b \\ c(\mathfrak{z}_1/\mathfrak{z}_2) + d \end{pmatrix}$$

which are exactly the homogeneous coordinates of the image of $z$ under this Möbius transform. This allows to calculate the result of any Möbius transformation also by means of a matrix multiplication.

The correspondence with matrices goes even deeper than this simple computational trick: the composition of two Möbius transforms can simply be obtained by multiplying their corresponding matrices (denoted by $M_1$ and $M_2$), that is

$$\mathfrak{M}_1 \circ \mathfrak{M}_2 \to M_1 M_2,$$

and furthermore, the inverse Möbius transformation $\mathfrak{M}^{-1}$ corresponds to the inverse of the matrix $M^{-1}$ [3].

### 6-2-3 The Möbius group

The link with the matrices already hints at a very important fact about Möbius transforms: they form a group under composition — the Möbius group Möb. This fact is actually immediately clear from the correspondence between matrix multiplication and composition of the transformation: a composition of two transforms is again a transform. Additionally, matrix multiplication is associative and they are assumed to be nonsingular, so inverses always exist. As such, all the group axioms are satisfied (of course, all these results could have been obtained directly from the definition of the Möbius transform, but the reasoning based on the matrices is particularly straightforward). Perhaps not very suprising is that the identity Möbius transformation $E(z)$ corresponds to the $2 \times 2$ identity matrix.

Based on the matrix analogy, the one can state that Möb is the group of linear[3] tranformations on vector space $\mathbb{C}^2$ — as projective coordinates. This is the same as saying that the Möbius group is isomorphic to the group of linear transformations *modulo* the nonzero scaling operation on $\mathbb{C}^2$: the resulting quotient group is the *projective linear group* $\mathrm{PGL}(2,\mathbb{C})$.

An interesting fact from group theory arises here as well: it has already been established that Möbius transforms are only unique up to multiplication by a scalar — as such, they can be normalized (with unit determinant) without loss of generalization. This suggests that Möb is really the action of the *special* linear group modulo scalar multiplication, resulting in the *projective special linear group*. Luckily, this fact is also reconciled within group theory: the groups $\mathrm{PGL}(n,\mathbb{F})$ and $\mathrm{PSL}(n,\mathbb{F})$ over the field $\mathbb{F}$ are isomorphic as long as every element of $\mathbb{F}$ has an $n$th root within $\mathbb{F}$. The fact that is is true for $\mathbb{F}=\mathbb{C}$ is probably the most fundamental property of the complex numbers. To summarize, the Möbius group Möb is equal to the projective linear group and the special projective group (over the field of complex numbers), which are isomorphic to each other in this particular case.

## 6-3 Classification of Möbius transforms

In the preceding discussion about the matrix representation of Möbius transforms, one important aspect has not yet been addressed: what about the eigenvalues and eigenvectors of $M$? Eigenvectors are vectors that remains invariant (up to scaling) under the multiplication of a particular matrix. Of course, one should bear in mind that the vector in this case contains projective coordinates, so that even when scaled, its coordinate representation remains identical. As such, the eigenvectors of the matrix $M$ are the *fixed points* of the Möbius transform $\mathfrak{M}$; any Möbius transform has two at most. This is even more perspicuous by solving the equation $z_0=\mathfrak{M}(z^*)$, which has solutions

$$z_0=\frac{(a-d)\pm\sqrt{(a+d)^2-4}}{2c}.$$

When converted to projective coordinates, the two solutions for $z^*$ then coincide with the eigenvectors of $M$. In the degenerate case for which $a+d=\pm 2$, the argument of the square root amounts to zero, which means that there is only one unique fixed point. These transforms are called *parabolic*, more will become clear about them later [2].

Now, it remains to analyse the significance of the eigenvalues. Of course, since eigenvalues are sensitive to scaling of the matrix, it is important to stress again that $M$ must be normalized (have a unit determinant) in order for the following to hold. A well known fact in linear algebra states that if $\lambda_1,\lambda_2$ are eigenvalues of $M$, then

$$\operatorname{tr} M=\lambda_1+\lambda_2 \quad \text{and} \quad \det M=\lambda_1\lambda_2.$$

Because the matrix is normalized, these two results can be combined into:

$$\lambda+\frac{1}{\lambda}=a+d=\operatorname{tr} M. \tag{6-2}$$

[3] It is misleading to call the Möbius transforms 'linear' in general — they are definitely nonlinear in the complex plane! However, when using the homogeneous coordinates, they become linear transforms.

It can be shown that every non-parabolic Möbius transform is conjugate[4] to a Möbius transform that has fixed points 0 and $\infty$, and is therefore of the form $\mathfrak{J}(z) = kz$, where $k$ is called the *multiplier* of this Möbius transform, and consequently all the transforms that are conjugate to it. The matrix $J$ that coincides with this transform necessarily must have the form (the letter $J$ is used to denote this transform because it is equal to the Jordan form of the transformation matrix $M$) [2]

$$J = \begin{pmatrix} \sqrt{k} & 0 \\ 0 & \frac{1}{\sqrt{k}} \end{pmatrix},$$

because then of course $\mathfrak{J}(z) = \frac{\sqrt{k}z}{1/\sqrt{k}} = kz$. Because conjugacy translates to a similarity transform in the matrix analogy, it leaves the eigenvalues of the matrix unaffected. But, for $J$ is a diagonal matrix, its eigenvalues are exactly on the main diagonal. As a result, *the multiplier of a Möbius transform is equal to the square of its eigenvalue*, or $k = \lambda^2$. Strictly speaking, every Möbius transform has of course two eigenvalues and two multipliers, but since they are both each others reciprocal, they do not have to be considered separately. With this result, eq. (6-2) can then also be restated in terms of the multiplier $k$ instead of the eigenvalues:

$$\sqrt{k} + \frac{1}{\sqrt{k}} = a + d = \operatorname{tr} M.$$

Solving eq. (6-2), one obtains

$$\lambda^2 - (a+d)\lambda + 1 = 0$$

which is a quadratic equation with discriminant $\Delta = (a+d)^2 - 4$: from the sign of $\Delta$ one can then distinguish three possible cases:

1. $\Delta < 0$ or $(a+d)^2 < 4$: there are two complex solutions for $\lambda$. It is easy to show that the solution will then be equal to

$$\frac{a+d}{2} \pm \frac{\mathrm{i}}{2}\sqrt{4 - (a+d)^2}.$$

   By inspection of this expression, it comes natural to make a further categorization:

   (a) if $(a+d)^2 \in [0, 4)$, the argument of the square root is positive: consequently, the solutions for $\lambda$ are both located on the unit circle (evidently, the unit circle as a whole is invariant under complex inversion). Any number on the unit circle can, by virtue of Euler's formula, be written as $\lambda = \mathrm{e}^{\frac{\mathrm{i}\theta}{2}} = \cos\left(\frac{\theta}{2}\right) + \mathrm{i}\sin\left(\frac{\theta}{2}\right) \neq 1$, such that the multiplier $k = \lambda^2 = \mathrm{e}^{\theta\mathrm{i}}$ — the factor of one half is only there to identify the multiplier with the actual angle $\theta$, which is the most meaningful from a geometric standpoint. The associated matrix archetype or Jordan form for transformations of this type is then

$$J = \begin{pmatrix} \mathrm{e}^{\frac{\mathrm{i}\theta}{2}} & 0 \\ 0 & \mathrm{e}^{\frac{-\mathrm{i}\theta}{2}} \end{pmatrix} = \exp\left(\begin{pmatrix} \frac{\mathrm{i}\theta}{2} & 0 \\ 0 & -\frac{\mathrm{i}\theta}{2} \end{pmatrix}\right) \quad \theta \in \mathbb{R} \setminus \{k \in \mathbb{Z} \mid 2k\pi\}.$$

[4] Two group elements $a$ and $b$ are called *conjugate* if there exists another group element $g$ such that $b = g^{-1}ag$. This is analogous to similarity transforms (and therefore the notion of similar matrices) in linear algebra.

It is more instructive to look at the *real Jordan form* of this complex diagonal matrix,

$$J = \begin{pmatrix} \cos(\theta/2) & -\sin(\theta/2) \\ \sin(\theta/2) & \cos(\theta/2) \end{pmatrix}$$

which is a rotation matrix[5]. Therefore, *matrices associated with elliptic transforms are rotation matrices.* These transforms are called *elliptic.* The edge case for which $(a+d)^2 = 0$ with a multiplier is equal to -1, is denoted as a *circular* transform (which is still an elliptic transform).

(b) Conversely, if $(a+d)^2 < 0$, the solutions will generally be complex (and not conjugate). These transforms are part of a larger class called *loxodromic.* As already stated, the loxodromic transforms also include the hyperbolic ones. Needham [2] reckons the elliptic transforms among the loxodromic transforms as well, but this is not general. In any case, the term 'loxodromic' usually refers as a pars pro toto to the non-hyperbolic transforms in particular to make the distinction.

2. $\Delta = 0$ or $(a+d)^2 = 4$: there is one solution for $\lambda$, either $1^{(2)}$ or $-1^{(2)}$, corresponding to a trace of -2 and 2 respectively (the superscript between parentheses indicates the algebraic multiplicity of the eigenvalues), because a normalized Möbius transform is only unique up to a sign. The multiplier for both cases is the same though, as $k = 1$. Möbius transformations of this kind are called *parabolic.* Because they have only one eigenvalue, there will also be one fixed point: the infinity point. The parabolic transforms give rise to translations in the complex plane of the form $z \mapsto z + b$ with matrix representation

$$M = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix},$$

which is a *unipotent matrix*[6]; these matrices form an abelian subgroup on their own (translations in the plane do indeed commute). The Jordan form of this matrix is then

$$J = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

which corresponds to the degenerate case where the geometric multiplicity of the eigenvalue is lower than its algebraic multiplicity.

3. $\Delta > 0$ or $(a+d)^2 > 4$: there are two real solutions for $\lambda$, given by

$$\lambda = \frac{a+d}{2} \pm \frac{1}{2}\sqrt{(a+d)^2 - 4}.$$

The resulting solutions for $\lambda$ are then always real and positive; they can then be expressed as the image of the exponential function: $\lambda = \mathrm{e}^{\frac{\zeta}{2}}$ such that $k = \mathrm{e}^{\zeta}$. The usage of

[5] One should be mindful that the conversion to the real Jordan form is only possible when the eigenvalues of the matrix are complex conjugate. In general, $M$ is complex, which means that this is not necessarily the case. However, on the unit circle, a complex inversion results in a reflection along the real axis, which means that the eigenvalues are in the elliptic case indeed complex conjugate.

[6] In general, a unipotent (ring) element is an element that, when the unit element is subtracted from it, yields a nilpotent element. For matrices, this means that a matrix $A$ is unipotent if $A - I$ is nilpotent, so $(A - I)^n = 0$ for some integer $n$.

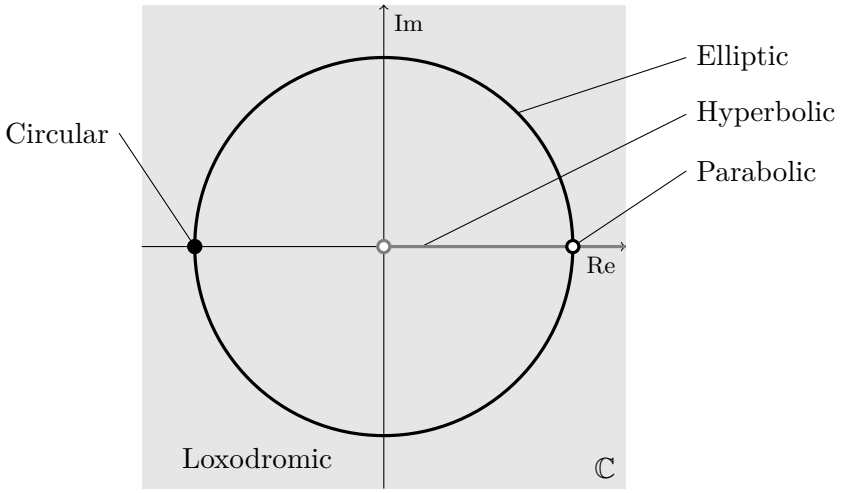



**Figure 6-1:** Classification of Möbius transform in terms of the location of the multiplier $k$ in the complex plane. Any point that is not on the unit circle yields a loxodromic transform; a particular subclass are the hyperbolic transforms, which are on the real axis except at $-1$ and $1$ where it intersects with the unit circle, and at the origin, where the transform becomes singular. If the multiplier lies on the unit circle except 1 are elliptic transforms, a special case is the circular transform for $k = -1$. Finally, the parabolic transforms have a multiplier of 1 [2].

$\zeta$ is not at all coincidental: indeed, the argument here represents a *hyperbolic angle*, as discussed in chapter 4. The corresponding Jordan form is a *squeeze mapping*:

$$J = \begin{pmatrix} e^{\frac{\zeta}{2}} & 0 \\ 0 & e^{\frac{-\zeta}{2}} \end{pmatrix} \quad \zeta \in \mathbb{R} \setminus \{0\}.$$

Similarly to the elliptic case, this matrix can also be expressed in terms of hyperbolic functions, completing the analogy that was already established in chapter 4:

$$M = \begin{pmatrix} \cosh(\zeta/2) & -\sinh(\zeta/2) \\ \sinh(\zeta/2) & \cosh(\zeta/2) \end{pmatrix}.$$

As such, these transforms are akin to hyperbolic rotations, which is why they are also classified as *hyperbolic*. The hyperbolic transforms are also part of the class of loxodromic transforms, together with the aforementioned class where $\lambda$ is complex. They do however deserve their own subclass because, apart from frequently encountered, they also represent a particular part of the special linear group over the reals $\mathrm{SL}(n, \mathbb{R})$, as will be discussed later.

To summarize, there are five different classes of Möbius transform: circular, elliptic, hyperbolic, loxodromic and parabolic. Circular transforms are part of the elliptic transforms and hyperbolic transforms are a subclass of loxodromic transforms. The class to which a Möbius transform belongs is determined completely by its trace $a + d$, or equivalently, the value of the multiplier $k$. For the multiplier, one can distinguish several 'regions' in the complex plane that are each associated with a class of Möbius transforms, this is visualized in fig. 6-1. The nature of the Jordan form of the matrix associated to a Möbius transform also clearly give away to which class a transform belongs.

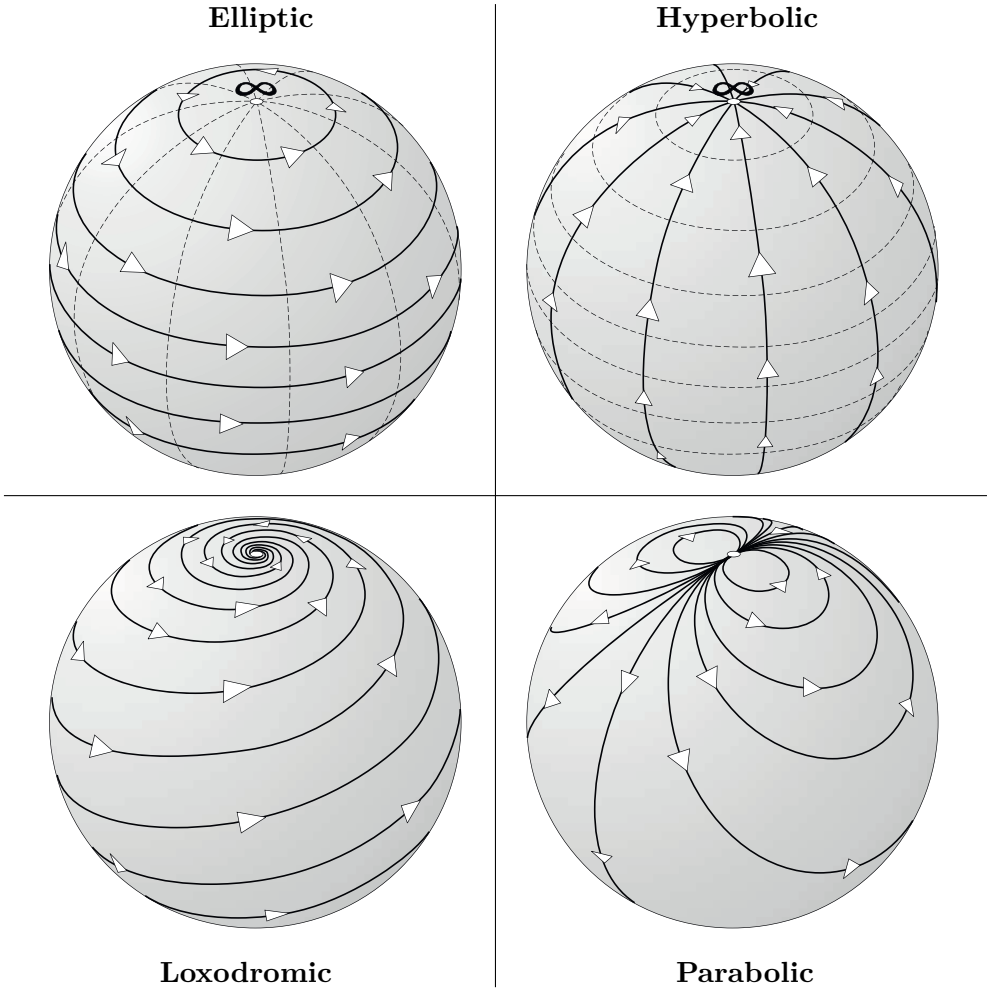



**Figure 6-2:** Overview of the four classes of Möbius transform and their typical action on the Riemann sphere. The curves shown on the Riemann sphere are the *invariant curves* of the transformation, i.e. these curves as a whole remain invariant under the transformation. Clearly, the elliptic, hyperbolic and loxodromic transformations have the North and South pole, or $\infty$ and 0 as fixed points, whereas the parabolic transformation only has a single fixed point at the North pole. The loxodromic transformations borrow their name from loxodromes, which are spiral-like trajectories on the Earth with constant bearing — a ship that taking a loxodromic path would maintain a constant angle with respect to true North. Illustration reprinted from Needham [3, p. 78].

**Table 6-1:** Overview of the five classes of Möbius transforms and the corresponding values for the trace squared of the matrix ($\operatorname{tr} M = a + d$), the multiplier of the transform and the Jordan form.

| **Class** | **Multiplier** | $(a+d)^2$ | **Jordan form** | **Parent class** |
|---|---|---|---|---|
| Circular | $\{-1\}$ | $[0, 4)$ | $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ | Elliptic |
| Elliptic | $\{k \in \mathbb{C} \mid \lvert k\rvert = 1, k \neq 1\}$ | $[0, 4)$ | $\begin{pmatrix} e^{\theta i/2} & 0 \\ 0 & e^{-\theta i/2} \end{pmatrix}$ | — |
| Parabolic | $\{1\}$ | $\{4\}$ | $\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}$ | — |
| Hyperbolic | $\mathbb{R}_0^+ \setminus \{1\}$ | $(4, +\infty)$ | $\begin{pmatrix} e^{\zeta/2} & 0 \\ 0 & e^{-\zeta/2} \end{pmatrix}$ | Loxodromic |
| Loxodromic | $\{k \in \mathbb{C} \mid \lvert k\rvert \neq 1\}$ | $\mathbb{C} \setminus [0, 4]$ | $\begin{pmatrix} k & 0 \\ 0 & k^{-1} \end{pmatrix}$ | — |

Table 6-1 provides an overview of the five Möbius classes, together with the values for the matrix trace, the multiplier and the Jordan form. Finally, fig. 6-2 visualizes the effect of each of the transform classes on points on the Riemann sphere. Elliptic transforms push points along circles of constant latitude, while hyperbolic transforms move points orthogonally, along the meridians of the Riemann sphere.

## 6-4 Subgroups

Apart from the classification considered in the preceding chapter, there are also several subgroups of the Möbius group that can be distinguished. It is compelling to observe that each of these subgroups can each be identified with one of the 'geometry types' that were considered in chapter 5: those of positive (spherical), zero (Euclidean) and negative (hyperbolic) geometry. For every geometry type, a specific subgroup of the Möbius group will play the role of the direct isometry group in a particular 'map' [3].

### 6-4-1 Euclidean geometry

The direct isometries in the Euclidean plane consist simply of translations and rotations. Clearly, a rotation in the complex plane (around the origin) can be represented simply by $z \mapsto e^{i\theta}$, while a translation is $z \mapsto z + b$ with $b \in \mathbb{C}$. Hence, the entire direct isometry group of the Euclidean plane is given by

$$\mathfrak{E}(z) = e^{i\theta} + b.$$

This group is also called the Euclidean group of dimension two.

### 6-4-2 Spherical geometry

One of the most prevalent applications of group theory is are the representation of rotations in three-dimensional Euclidean space. Of course, the group associated with these rotations is the special orthogonal group $\mathrm{SO}(3,\mathbb{R})$. This group is diffeomorphic to the three-dimensional real projective space $\mathbb{RP}^3$ (intuitively, this space consists of three 'directions'). The rotations in $\mathbb{R}^3$ can also be parameterized by *unit quaternions* (also called *versors*): these represent points on the 3-sphere $\mathbb{S}^3$. The difference between the 3-sphere and the real projective space is that the latter identifies the *antipodal* parts that are present on the sphere. As such, any rotation in $\mathbb{R}^3$ corresponds precisely to two points on the 3-sphere (or two unit quaternions) — the group of unit quaternions therefore is a double cover of $\mathrm{SO}(3,\mathbb{R})$.

Quaternions can also be represented as complex matrices: [44]

$$q = m + n\mathbf{i} + o\mathbf{j} + p\mathbf{k} \in \mathbb{H} \quad \leftrightsquigarrow \quad Q = \begin{pmatrix} m + \mathrm{i}p & -n - \mathrm{i}o \\ n - \mathrm{i}o & m - \mathrm{i}p \end{pmatrix},$$

which is a general represention of a $2 \times 2$ *unitary*[7] matrix. Likewise, the unit quaternions then translate to matrices of the above type with an additional restriction: they must have a determinant of 1: these matrices are members of the *special unitary group* $\mathrm{SU}(2)$. As such, the group of unit quaternions is isomorphic to the $\mathrm{SU}(2)$ which is therefore also a double cover of $\mathrm{SO}(3,\mathbb{R})$.

Since the members of $\mathrm{SU}(2)$ can be identified with a Möbius transform (inspection of the matrix above makes this immediately apparent), a specific subgroup of Möb can be used to represent the rotations in $\mathbb{R}^3$ — recall that every normalized Möbius transform also corresponds to two matrices, differing by a factor -1. Observing the matrix above, one can see that the entries on the main diagonal are each others' conjugate, while the entries on the antidiagonal are conjugate opposite. Therefore, the general expression of a rotation of the Riemann sphere as a Möbius transform can be written as: [3]

$$\mathfrak{S} = \frac{az + b}{-\bar{b}z + \bar{a}} \quad \text{where } |a|^2 + |b|^2 = 1;$$

the latter equivalent then enforces that $\det Q = 1$. There are always two quaternions representing the same rotation; they are antipodal and differ by a factor of $-1$. As a result, there are two possibilities for $Q$ as well, again identical but with opposite entries. Recall that the same applies to the matrix representations of Möbius transforms: as such, the ambiguities are eliminated, and the group of transforms of type $\mathfrak{S}$ is isomorphic to $\mathrm{SO}(3,\mathbb{R})$. In the complex plane, these tranformations represent the isometries of spherical geometry in the stereographic map [2].

### 6-4-3 Hyperbolic geometry

As described by Rovenski [45], the isometries of the Poincaré half plane are any (composition) of the following types of transformations, i.e. they leave distances according to the Poincaré metric unaffected:

[7] A unitary matrix is a matrix whose inverse is its conjugate transpose — it is the complex counterpart of orthogonal matrices.

- horizontal translations: $(x, y) \mapsto (x + a, y)$ where $x, y, a \in real$;
- reflection around the vertical axis: $(x, y) \mapsto (-x, y)$;
- dilations centered around the origin: $(x, y) \mapsto (ax, ay)$;
- inversions with respect to the unit circle $(x, y) \mapsto \left(\frac{x}{x^2+y^2}, \frac{y}{x^2+y^2}\right)$.

The group of all these transformations is precisely $\mathrm{PSL}(2, \mathbb{R})$, or the Möbius tranforms with real parameters:

$$\mathfrak{H}(z) = \frac{az + b}{cz + d} \qquad a, b, c, d \in \mathbb{R} \quad ad - bc = 1;$$

these are the isometries of the Poincaré half plane [3]. Recall from section 5-2-1 that there are two types of geodesics in the half plane: semicircles centered at the origin and straight (from Euclidean perspective) vertical lines. These are precisely invariant curves for the transformations listed — isometries take geodesics to geodesics [33].

## 6-5 Relation with special linear group

Another classic Lie group is the special linear group $\mathrm{SL}(n, \mathbb{F})$, which are the volume-preserving transformations on a vector space — this property turns out to be quite important. The special linear group over the complex numbers (2-dimensional) $\mathrm{SL}(2, \mathbb{C})$ can be represented as the group of all the complex $2 \times 2$ matrices with unit determinant. It has been mentioned previously that any such matrix coincides with a Möbius transform, albeit surjectively: for every Möbius transform, there are two such matrices. As such, $\mathrm{SL}(2, \mathbb{C})$ is a *double cover* of Möb.

Arguably more interesting than $\mathrm{SL}(2, \mathbb{C})$ is the special linear group over the reals $\mathrm{SL}(2, \mathbb{R})$, i.e. every invertible $2 \times 2$ matrix with real entries and a unit determinant; they form a subgroup of $\mathrm{SL}(2, \mathbb{C})$ as well. It has been mentioned before that the for the normalized Möbius transforms, the eigenvalues should be complex inverses of each other. For matrices with real entries, the eigenvalues are either both real or complex conjugates of each other. When not on the real axis, the only way eigenvalues can be complex inverses and complex conjugate is to be located on the unit circle. As can be seen on fig. 6-1, that means that the matrices with real entries contain exactly the hyperbolic (both real), elliptic (unit circle) and parabolic classes. The remaining part are the loxodromic transforms which have a nonreal trace [CHECK].

## 6-6 Relation with Lorentz group

## 6-7 Chapter summary

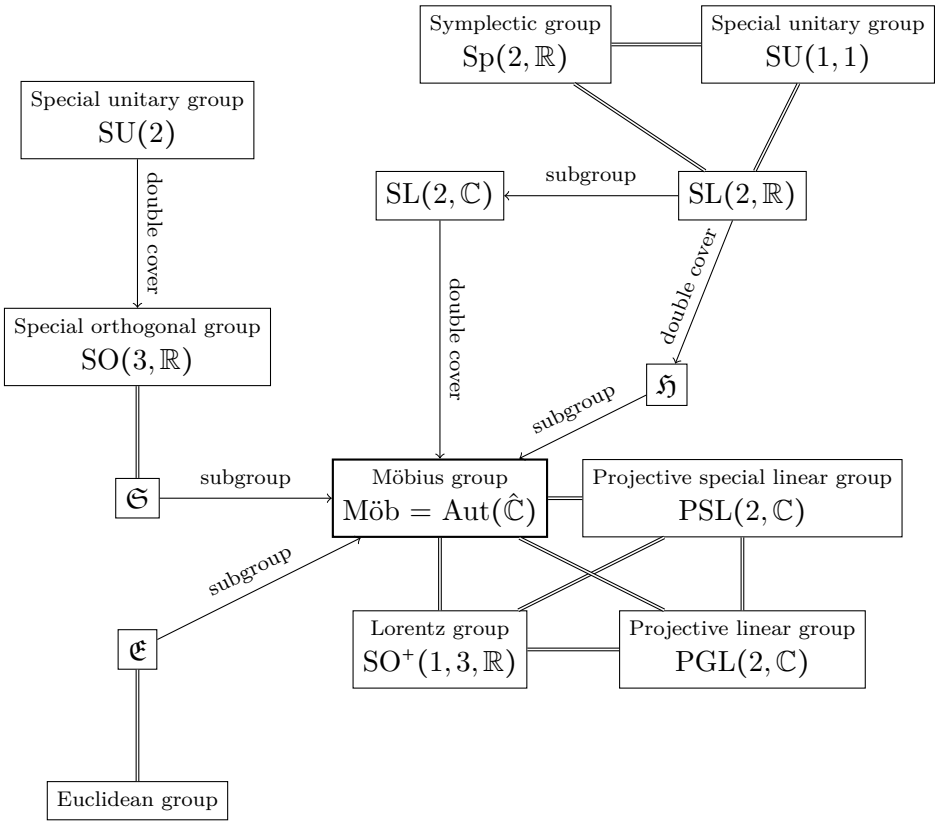



**Figure 6-3:** Bla bla

Chapter 7

# Research proposal

Chapter 8

# Summary and conclusion

# Bibliography


[1] B. N. M. Krabbenborg, "Duration Matching in the Frequency Domain - An Economic Engineering Approach to Asset and Liability Management," Literature Survey, Delft University of Technology, 2021.

[2] T. Needham, *Visual Complex Analysis*. New York: Oxford University Press, 1997.

[3] ——, *Visual Differential Geometry and Forms: A Mathematical Drama in Five Acts*, 1st ed. Princeton, NJ: Princeton University Press, 2021.

[4] D. C. Karnopp, D. L. Margolis, and R. C. Rosenberg, *Modeling, Simulation, and Control of Mechatronic Systems*, 5th ed. Hoboken, NJ: John Wiley and Sons, 2012.

[5] "FISHER SEES STOCKS PERMANENTLY HIGH; Yale Economist Tells Purchasing Agents Increased Earnings Justify Rise. SAYS TRUSTS AID SALES Finds Special Knowledge, Applied to Diversify Holdings, Shifts Risks for Clients." New York, NY, oct 1929.

[6] C. D. Romer, "Great Depression: Definition, History, Dates, Causes, Effects, and Facts," 2021. [Online]. Available: https://www.britannica.com/event/Great-Depression

[7] M. Mendel, "Economic Engineering - Lecture Notes," Delft, 2019.

[8] M. A. Vos, "The Application of System and Control Theory to Monetary Policy: Development of a First-Principles LTI Model of the Economy," Master's thesis, Delft University of Technology, 2019.

[9] G. J. L. Kruimer, "An Engineering Approach to Macroeconomic Scenario Modelling," Master's thesis, Delft University of Technology, 2021.

[10] C. Hutters and M. Mendel, "Overcoming the dissipation obstacle with Bicomplex Port-Hamiltonian Mechanics," 2020.

[11] N. Manders, "The Thermodynamics of Economic Engineering with Applications to Economic Growth," Master's thesis, Delft University of Technology, 2019.

[12] X. A. Van Ardenne, "Business Valuation in the Frequency Domain - A Dynamical Systems Approach," Master's thesis, Delft University of Technology, 2020.

[13] F. Bullo and A. D. Lewis, *Geometric Control of Mechanical Systems - Modeling, Analysis and Design for Simple Mechanical Control Systems*, 1st ed., ser. Texts in Applied Mathematics. New York, NY: Springer Science+Business Media, 2005.

[14] V. Arnol'd, *Mathematical Methods of Classical Mechanics*, 2nd ed., J. Ewing, F. Gehring, and P. Halmos, Eds. New York: Springer-Verlag, 1989.

[15] L. D. Landau and E. M. Lifshitz, *Mechanics*, 2nd ed., ser. Course of Theoretical Physics. Oxford: Pergamon Press, 1960, vol. 1.

[16] N. G. Mankiw and M. P. Taylor, *Economics*, 4th ed. Andover, Hampshire, UK: Cengage Learning, 2017.

[17] R. Feynman, *The Feynman Lectures on Physics*. New York, NY: Basic Books, 2010, vol. 1, 2, 3.

[18] D. Jeltsema and J. M. Scherpen, "Multidomain Modeling of Nonlinear Networks and Systems: Energy- and power-based perspectives," *IEEE Control Systems*, vol. 29, no. 4, pp. 28–59, 2009.

[19] A. Marshall, *Principles of Economics*, 8th ed., ser. Palgrave Classics in Economics. Macmillan Publishers, 1920.

[20] J. Stewart, *Calculus: Early Transcendentals*, 7th ed. Brooks/Cole Cengage Learning, 2012.

[21] R. Zipf, *Fixed Income Mathematics*. Academic Press - Elsevier Science, 2003.

[22] S. G. Kellison, *The Theory of Interest*, 2nd ed. Richard D. Irwin, 1991.

[23] A. A. Harkin and J. B. Harkin, "Geometry of Generalized Complex Numbers," *Mathematics Magazine*, vol. 77, no. 2, pp. 118–129, 2004.

[24] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation*. San Francisco, CA: W. H. Freeman and Company, 1970.

[25] E. F. Taylor and J. A. Wheeler, *Spacetime Physics - An Introduction to Special Relativity*, 2nd ed. New York, NY: W. H. Freeman and Company, 1992.

[26] L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, ser. Course of Theoretical Physics. Oxford: Pergamon Press, 1971, vol. 2.

[27] R. Penrose, "The Geometry of the Universe," in *Mathematics Today: Twelve Informal Essays*, L. A. Steen, Ed. Springer-Verlag, 1978, ch. 4, pp. 83–125.

[28] J. G. Ratcliffe, *Foundations of Hyperbolic Manifolds*, ser. Graduate Texts in Mathematics. Cham: Springer International Publishing, 2019, vol. 149. [Online]. Available: http://link.springer.com/10.1007/978-3-030-31597-9

[29] F. Catoni, D. Boccaletti, R. Cannata, V. Catoni, E. Nichelatti, and P. Zampetti, *The Mathematics of Minkowski Space-Time*, ser. Frontiers in Mathematics. Basel, Switzerland: Birkhaüser Verlag, 2008.

[30] P. Fjelstad, "Extending special relativity via the perplex numbers," *American Journal of Physics*, vol. 54, no. 5, pp. 416–422, 1986.

[31] G. Sobczyk, "The Hyperbolic Number Plane," *The College Mathematics Journal*, vol. 26, no. 4, pp. 268–280, 1995.

[32] A. E. Motter and M. A. F. Rosa, "Hyperbolic Calculus," *Advances in Applied Clifford Algebras*, vol. 8, no. 1, pp. 109–128, 1998.

[33] J. M. Lee, *Riemannian Manifolds: An Introduction to Curvature*, ser. Graduate Texts in Mathematics. New York, NY: Springer-Verlag, 1997.

[34] W. P. Thurston, *Three-Dimensional Geometry and Topology*, S. Levy, Ed. Princeton, NJ: Princeton University Press, 1997, vol. 1.

[35] B. O'Neill, *Elementary Differential Geometry*, 2nd ed. New York: Academic Press, 2006.

[36] M. Spivak, *A Comprehensive Introduction to Differential Geometry*, 3rd ed. Houston, TX: Publish or Perisch, Inc., 1999, vol. 2.

[37] J. M. Lee, *Introduction to Topological Manifolds*, ser. Graduate Texts in Mathematics. New York, NY: Springer-Verlag, 2000.

[38] F. P. Schuller, "The Geometric Anatomy of Theoretical Physics," Friedrich-Alexander Universität Erlangen-Nürnberg - Institut für Theoretische Physik II, Tech. Rep., 2014.

[39] E. Ghys, "Poincaré and His Disk," in *The Scientific Legacy of Poincaré*, ser. History of Mathematics. The American Mathematical Society, 2010, vol. 36, ch. 1, pp. 17–45.

[40] A. Ramsay and R. D. Richtmyer, *Introduction to Hyperbolic Geometry.* New York: Springer Science+Business Media, 1995.

[41] B. O'Neill, *Semi-Riemannian Manifolds with Applications to Relativity.* San Diego, CA: Academic Press, 1983.

[42] N. L. Balazs and A. Voros, "Chaos on the Pseudosphere," *Physics Reports*, vol. 143, no. 3, pp. 109–240, 1986.

[43] R. Penrose and W. Rindler, *Spinors and Space-Time - Two-spinor Calculus and Relativistic Fields.* Cambridge: Cambridge University Press, 1984, vol. 1.

[44] J. Stillwell, *Naive Lie Theory*, ser. Undergratuate Texts in Mathematics. New York: Springer Science+Business Media, 2008.

[45] V. Rovenski, *Modeling of Curves and Surfaces with MATLAB.* New York: Springer Science+Business Media, 2010.

# Glossary

## List of Acronyms

| | |
|---|---|
| **IRR** | Internal Rate of Return |
| **NPV** | Net Present Value |

## List of Symbols

| | |
|---|---|
| $\mathbb{C}$ | Complex numbers |
| e | Euler's constant |
| $\mathbb{R}$ | Real numbers |
| $E$ | First component of the first fundamental form |
| $e$ | First component of the second fundamental form |
| $F$ | Second component of the first fundamental form |
| $f$ | Second component of the second fundamental form |
| $G$ | Third component of the first fundamental form |
| $g$ | Third component of the second fundamental form |
| $H^2$ | Two-sheeted hyperboloid embedded in three-dimensional Lorentz space. |
| $H^2_+$ | Positive hyperboloid sheet embedded in three-dimensional Lorentz space. |
| $k$ | Gaussian curvature |

# Index