A Survey of Hyperbolic Rotations

Towards a Consistent Interpretation of Financial Instruments in Economic Engineering

E. B. Legrand

Literature Survey

Delft University of Technology

Delft Center for Systems and Control

A Survey of Hyperbolic Rotations

Towards a Consistent Interpretation of Financial Instruments in Economic Engineering

Literature Survey

E. B. Legrand

December 3, 2021

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

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

Abstract

Financial instruments are currently interpreted in economic engineering as rotational mechanical systems. This is motivated by the fact that compound interest can be viewed as a hyperbolic rotation. Hyperbolic rotations, or squeeze mappings, make various appearances in physics and mathematics; notable examples are the theory of special relativity, Möbius transformations and hyperbolic geometry. All of these subjects are related, because the group of Möbius transformations is isomorphic to the group of restricted Lorentz transformations in special relativity. The Lorentz group is the isometry group of the hyperbolic plane embedded in three-dimensional Lorentz space.

The hyperbolic rotations find their application in economic engineering because continuous reinvestment of earnings leads to a hyperbolic rotation in the capital-yield plane. Unfortunately, the financial analogy with rotational dynamics shows some severe inconsistencies when applied to real systems. It is therefore proposed to develop a new interpretation of capital and investments. This approach shall be rooted in the fundamentals of economic engineering based on principles from analytical mechanics.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

On the eighth page of the October 16, 1929 edition of the New York Times, a small article headed [6]

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

Responsible for this bold claim was Irving Fisher, a very prominent American economist who pioneered 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 unprecedented financial crisis with dramatic repercussions: in the three years that followed, unemployment rose to 20 percent and industrial production almost halved [7].

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 present state of financial markets. As it turns out, the interpretation of financial markets tends to be a problem for economic engineering as well. [introduce: economic engineering, hyperbolic rotations]

As mentioned, the main theme of this literature survey are the hyperbolic rotations or 'squeeze mappings'. The literature survey consists of two parts: the first part treats the established applications of hyperbolic rotations in mathematics and physics, while the second part goes into their relevance in economic engineering as part of the rotational analogy.

Part I discusses some notable appearances that hyperbolic rotations make in mathematics and physics, and highlights the connections between these subjects. Three subjects are discussed: special relativity, hyperbolic geometry and Möbius transformations. Chapter 2 gives a basic introduction to the theory of special relativity for complete novices in the subject, where hyperbolic rotations occur as the 'Lorentz boosts' in spacetime.

Secondly, chapter 3 first introduces some tools from differential geometry to define the (Gaussian) curvature of surfaces. Hyperbolic geometry is the geometry of negatively curved surfaces, which differs from 'traditional' Euclidean geometry in many remarkable ways. Finally, chapter 4 ties the previous two subjects together in an elegant fashion using Möbius transforms, an important subject from complex analysis.

Part II explores the foundations of economic engineering, and specifically the relevance of hyperbolic rotations to the subject. Chapter 5 essentially discusses two subjects in parallel: analytical mechanics; in particular Lagrangian and Hamiltonian mechanics, and economic engineering. The main text concerns the theory of analytical mechanics, but for every concept introduced the corresponding economic engineering interpretation is given as well. The different approach taken in chapter 5 using analytical mechanics has the benefit of producing more rigorous and universal principles than the Newtonian counterpart. Due to the hybrid nature of this chapter, it is suitable for readers familiar with classical mechanics or economic engineering, or neither of these subjects.

Secondly

The two parts of this literature study can be read more or less independently from each other; the first part is more mathematically inclined, while the second part is specifically geared towards economic engineering.

Part I

Hyperbolic Rotations in Mathematics and Physics

Chapter 2

Special relativity

This chapter gives a basic outline of the theory of special relativity. It is by no means meant to be a comprehensive overview, for which many other excellent resources exist such as Misner et al. [8], Taylor and Wheeler [9], Landau and Lifshitz [10], or Penrose [11] 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 discussed in chapter 2, and Möbius transforms which will be introduced in chapter 4.

First, section 2-1 introduces the relativity of time and the concept of spacetime. Consequently, section 2-2 proceeds with the definition of spacetime intervals, the analog of distance in spacetime. This can be generalized to the notion of a Lorentzian vector space, which is the subject of section 2-4. Finally, section 2-4 discusses the transformations of spacetime that preserve the spacetime interval: the Lorentz transformations.

2-1 Time as a relative concept

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 yield a coordinate transform between reference frames that move relative to each other at a constant linear velocity. The laws of physics should be invariant between 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 velocity 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 for time is indispensable to incorporate the relativity of time. Points in the four-dimensional *spacetime* are called *world points*, their associated trajectories are *world lines* [10].

2-2 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 traveling at the speed of light is equal to

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

Therefore, the spacetime interval s_{12} 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 traveling 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 [10]. 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 occurred *at the same location* in space, they are simply separated by the passage of time $t_{12} = \dfrac{s_{12}}{c}$.
- 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* $l_{12} = \mathrm{i}s_{12}$, where i is the imaginary unit.

- Intervals for which $s_{12}^2 = 0$ are called *lightlike*, because only light (or anything traveling at the speed of light) can travel between these events.

This begs the question what the actual time is that an observer would experience in uniform motion, i.e. what difference in time is displayed by clocks that have been traveling at different velocities. The time experienced by an observer is called *proper time*, and it can be computed by evaluating the following path integral:

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

Clearly, if the velocity v 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 traveling at the speed of light do not register the passage of time whatsoever).

2-3 Lorentzian vector spaces

The concept of the spacetime interval introduced in the previous section gives rise to a different notion of 'distance' in four-dimensional spacetime. This distance is different from the traditional Euclidean distance: it is a *Lorentzian* distance. This different notion of distance is captured by the definition of the *Lorentz(ian) product*[1].

Lorentz product

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

$$\boldsymbol{u} \bullet \boldsymbol{w} = u_1 w_1 - u_2 w_2 - \ldots - u_n w_n. \tag{2-2}$$

The Lorentz product resembles the 'normal' inner product, with the exception of the minus signs before every but the first term. These minus signs are the reason why the Lorentz product is an indefinite product; that is, it fails to be positive definite. An inner product satisfies (i) bilinearity, (ii) symmetry, and (iii) positive definiteness. As discussed in the previous section, the inner product of a spacelike vector with itself will *not* return a positive number; a lightlike vector will produce 0. As such, the Lorentz product is indefinite, which is why it is called a *pseudo-inner product* [12, 13]. The weaker condition that is imposed for pseudo-inner products asserts that it must be nondegenerate, which means that there exists no non-zero vector for which its inner product with any other vector is zero.

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

$$\|\boldsymbol{u}\|_L = (\boldsymbol{u} \bullet \boldsymbol{u})^{\frac{1}{2}}.$$

[1] 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 down to $(+,-,-,-)$, $(-,-,-,+)$, $(-,+,+,+)$ and $(+,+,+,-)$, where the first one coincides with the definition used here, in accordance with the magnificent work of Landau and Lifshitz [10]. 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.

Because the Lorentz product is not positive definite but indefinite, the result of $\boldsymbol{u} \bullet \boldsymbol{u}$ is not guaranteed to be a positive number. As such, $\|\boldsymbol{u}\|_L \in \mathbb{C}$, in stark contrast with the familiar Euclidean norm which *is* positive definite and will therefore always return a non-negative 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 $\boldsymbol{w}$ is equal to:

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

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 [14]. Elements of the Minkowski space are denoted by *four-vectors.* Often, Lorentzian spaces are denoted in short by $\mathbb{R}^{p,q}$ where the combination p, q indicates the metric signature. p and q respectively denote the number of plus and minus signs in the Lorentzian product. The total dimension of the space is then $p + q$. As such, the Minkowski space is denoted by $\mathbb{R}^{1,3}$. However, there are other Lorentzian spaces as well; in section 3-4-3 the discussion involves a three-dimensional Lorentzian space (with two spatial directions and one 'time' dimension) $\mathbb{R}^{1,2}$. Likewise, concepts in spatial relativity are usually illustrated using the Lorentz-Minkowski plane, which is basically a two-dimensional representation of spacetime referred to as $\mathbb{R}^{1,1}$. This Lorentzian space will play an important role in chapter 6.

The sign of the Lorentz norm gives rise to an equivalence relation[2] $\sim_L$ defined to be:

$$\boldsymbol{u} \sim_L \boldsymbol{w} \iff \operatorname{sgn} \|\boldsymbol{u}\|_L = \operatorname{sgn} \|\boldsymbol{w}\|_L.$$

Therefore, the quotient set of all the points in the plane $\mathbb{R} \setminus \sim_L$ contains three elements $\{-1, 0, 1\}$, these equivalence classes are given the respective names spacelike, lightlike and timelike, analogous to the terminology for the spacetime intervals from the previous section [10]. The difference between timelike, spacelike and lightlike vectors is illustrated for the Lorentz-Minkowski plane in fig. 2-1.

2-4 Lorentz transformations

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

$$x = x' + Vt, \quad y = y', \quad z = z', \quad t = t'. \tag{2-3}$$

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.

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

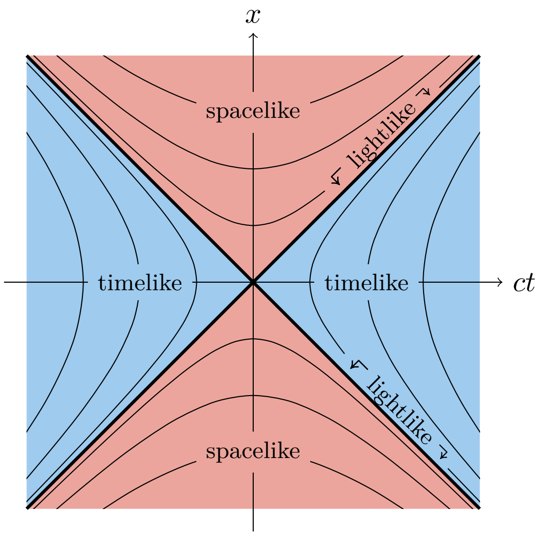


Figure 2-1: Overview of the three 'types' of vectors in the Lorentz-Minkowski plane: spacelike ($\|\boldsymbol{u}\|_L < 0$), lightlike ($\|\boldsymbol{u}\|_L = 0$) and timelike ($\|\boldsymbol{u}\|_L > 0$). The lines $ct = x$ and $ct = -x$ containing all the lightlike vectors form the so-called light cone or null cone. Points with vectors with identical length trace out hyperbolae, as they obey the equation $c^2t^2 - x^2 = K^2$.

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 [10], 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 or quaternions. 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 rotations (they are represented by hyperbolic functions rather than trigonometric functions). For example, a rotation in the tx-plane would take the following form:

$$x = ct' \sinh(\zeta) \qquad ct = ct' \cosh(\zeta); \tag{2-4}$$

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

$$x = \frac{x' + Vt'}{\sqrt{1 - \left(\frac{V}{c}\right)^2}} \qquad t = \frac{t' + \frac{V}{c^2}x'}{\sqrt{1 - \left(\frac{V}{c}\right)^2}} \qquad y = y' \qquad z = z'. \tag{2-5}$$

This indicates that the hyperbolic (boost) angle ζ can be written in terms of the velocity V of one frame with respect to the other frame:

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

This implies 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 made based on these equations:

- Clearly, V cannot be larger than c: there is no real ζ 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, the transform eq. (2-5) keeps $c^2t^2 - x^2$ unaffected (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.
- 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 in comparison to c.
- Due to the multiplication factor in the transform, two points appear closer (in the direction of x) together when traveling at speed than when they are at rest. A length measured in a rest frame is called *proper*, and contracts when measured in a moving frame: this phenomenon is called *Lorentz contraction* [10].
- 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. (2-4) and (2-5) 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: [10]

$$v_x = \frac{v'_x + V}{\sqrt{1 - \left(\frac{V}{c}\right)^2}} \qquad v_y = \frac{v'_y\sqrt{1 - \left(\frac{V}{c}\right)^2}}{1 + v'_x\left(\frac{V}{c}\right)} \qquad v_z = \frac{v'_z\sqrt{1 - \left(\frac{V}{c}\right)^2}}{1 + v'_x\left(\frac{V}{c}\right)}. \tag{2-6}$$

Much like positions, velocities have a four-dimensional counterpart in spacetime as well; these objects are called *four-velocities*, they are defined as [10]

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

with v being the three-dimensional velocity of the particle. The components of the four-velocity $\boldsymbol{u}$ are[3]

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

which are all dimensionless quantities. Clearly, the magnitude of any four-velocity amounts to one; or $u^iu_i = 1$. The indices of $\boldsymbol{u}$ are lowered by virtue of the metric tensor $\boldsymbol{g}$,

$$g_{ij} = \begin{cases} 1 & \text{if } i = j = 0, \\ -\delta_{ij} & \text{otherwise,} \end{cases} \tag{2-7}$$

[3] Here, the conventions for tensor notation are observed, where upper and lower indices refer to contravariant and covariant components respectively.

with $\boldsymbol{\delta}$ being the Kronecker delta. 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); which 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*, an important concept to which chapter 3 is entirely devoted.

The Lorentz group The space of four-vectors from the previous section allows viewing the Lorentz transformations as actions of the Lorentz group Lor. The Lorentz group is a Lie group that can be represented by 4×4 matrices. In the previous section, it was argued that the Lorentz transformations encompass $\binom{2}{4} = 6$ 'possible' rotations: three that are purely spatial and an additional three that involve the time dimension as well. To be precise, this discussion will be limited to the *restricted Lorentz group*: these are the transformations that (i) preserve the direction of time (ortochronous transformations) and (ii) preserve orientation (i.e. they only involve proper rotations) [15]. The restricted Lorentz group is often considered in lieu of the 'bigger' Lorentz group because it is most relevant for physics.

The restricted Lorentz group can be represented by matrices in the four-dimensional vector space. Because Lorentz transformations preserve the 'Lorentzian length' of vectors as dictated by the Lorentz product eq. (2-2), they constitute an *orthogonal group*. As such, the general Lorentz group is also denoted $\mathrm{O}(1,3)$, where '1,3' denotes the metric signature of the quadratic form that is preserved by the group. As a matrix group, $\mathrm{O}(1,3)$ is the following set:

$$\mathrm{O}(1,3) = \left\{A \in \mathrm{GL}(4) \;\middle|\; A^\top G A = G\right\} \qquad \text{with} \quad G = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

with $\mathrm{GL}(n)$ the general linear group of dimension n, and G the matrix representation of the metric tensor in eq. (2-7) [16]. Then, the proper Lorentz transformations (that preserve orientation) are given by the corresponding *special orthogonal group* $\mathrm{SO}(\cdot)$:

$$\mathrm{SO}(1,3) = \{A \in \mathrm{O}(1,3) \mid \det A = 1\}.$$

The transformations in the restricted Lorentz group must also be ortochronous, for which an additional restriction is required: they must map the positive sheet H_+ of the four-dimensional hyperboloid determined by the quadratic form

$$c^2t^2 - x^2 - y^2 - z^2,$$

to itself; the same holds for the negative sheet H_-. Then, one obtains the restricted Lorentz group $\mathrm{SO}^+(1,3)$:

$$\mathrm{SO}^+(1,3) = \{A \in \mathrm{SO}(1,3) \mid AH_\pm = H_\pm\},$$

which is equivalent to impose the restriction that $\operatorname{tr} A > 0$ [1, 16].

Fundamentally, the restricted Lorentz group consists of four types of 'generators' (or *conjugate classes* in the terminology of groups): elliptic, hyperbolic, loxodromic, and parabolic. The elliptic transformations are spatial rotations, while the hyperbolic transformations are

Lorentz boosts: they are the rotations that involve time. Loxodromic transformations consist of a simultaneous boost and spatial rotation. Finally, the parabolic Lorentz transformations are the ‘null rotations’, which do not have a completely intuitive meaning; but in Minkowski spacetime, they describe a parabolic trajectory that arises from the intersection of the hyperboloid with a null plane. As will be discussed in chapter 4, the restricted Lorentz group is isomorphic to the Möbius group, which shares the same distinction between hyperbolic, elliptic, parabolic, and loxodromic elements [3].

Finally, although ‘the’ Lorentz group is most often referred to as the one that acts on four-dimensional Minkowski spacetime, similar groups can be defined for other numbers of dimensions as well. For example, $\mathrm{SO}(1,1)$ is the special orthogonal group for the (1,1) Lorentz space that will play an important role in chapter 6. The one-dimensional Lorentz group is simply given by matrices of the form

$$\begin{pmatrix} \cosh(\zeta) & \sinh(\zeta) \\ \sinh(\zeta) & \cosh(\zeta) \end{pmatrix},$$

which are called *squeeze mappings* (in this case, there are no spatial rotations possible) or hyperbolic rotations of the plane. Again, the ‘1,1’ or ‘3,1’ refer to the metric signature of the space they act on, similarly to the notation of Lorentzian vector spaces.

Chapter summary

In contrast to Galilean relativity, special relativity refutes the absoluteness of time, and defines the notion of four-dimensional spacetime instead. Distances in spacetime are given by an indefinite quadratic form with metric signature (1,3), and are denoted by spacetime intervals. The distance-preserving transformations of spacetime constitute the Lorentz group. The Lorentz group contains both the spatial (‘circular’) rotations and ‘Lorentz boosts’, which are hyperbolic rotations. When generalized to other dimensions, the (1,1) Lorentz group consists simply of the hyperbolic rotations (or squeeze mappings) in the plane.

Chapter 3

Hyperbolic geometry

This chapter provides an introduction to hyperbolic geometry, which is related to both the Lorentzian vector spaces from chapter 2 and the Möbius transformations in chapter 4. In section 3-1, the reader is introduced to the two alternatives to Euclidean geometry: spherical geometry and hyperbolic geometry. Surfaces exhibit one of these geometry types depending on their Gaussian curvature; a central concept in differential (Riemannian) geometry which is discussed in section 3-2. Then, section 3-4 proceeds by outlining three representations of the hyperbolic plane.

3-1 Non-Euclidean geometry

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, it is either the case that

- 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*. The parallel axiom is equivalent to another fundamental principle in Euclidean geometry: the sum of the angles of a triangle is equal to π. As

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

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, i.e. the difference between the sum of the angles of a triangle and π. Three cases possible arise:

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

As a consequence of the Gauss-Bonnet theorem, the angular excess can be related to its area $\mathcal{A}$ [2]

$$\mathcal{E}(T) = k\mathcal{A}(T), \tag{3-1}$$

where k is a constant. For Euclidean geometry, $k = 0$, for hyperbolic geometry, $k < 0$, and for spherical geometry $k > 0$. This reveals 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. (3-1) is that similar triangles cannot exist in non-Euclidean geometry. 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 a surface with zero curvature the sum of the angles of a triangle will be precisely equal to π.

3-2 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 [18, 19]

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

Surface curvature can be approached more rigorously from the perspective of Riemannian geometry. The core concept behind Riemannian geometry is the eponymous manifold: a smooth

[2]Loosely translated by John M. Lee [17] as the *Totally Awesome Theorem.*

manifold equipped with a positive definite inner producton its tangent space, hence providing a *metric (tensor)* that allows determining 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*[3] $\mathbf{I}$.

First fundamental form

The first fundamental form $\mathbf{I}$ of a surface parameterized by $\boldsymbol{r}(u, v)$ is a (0, 2)-tensor defined as the inner product of two tangent vectors to the surface. It is characterized by three coefficients E, F, G:

$$E = \langle \boldsymbol{r}_u, \boldsymbol{r}_u \rangle \qquad F = \langle \boldsymbol{r}_u, \boldsymbol{r}_v \rangle \qquad G = \langle \boldsymbol{r}_v, \boldsymbol{r}_v \rangle, \tag{3-2}$$

with $\boldsymbol{r}_u = \dfrac{\partial \boldsymbol{r}}{\partial u}$ and $\boldsymbol{r}_v = \dfrac{\partial \boldsymbol{r}}{\partial v}$.

The components of the metric tensor $\boldsymbol{g}$ 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 embedded 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.

Second fundamental form

The second fundamental form $\mathbf{II}$ of a surface parameterized by $\boldsymbol{r}(u, v)$ is a (0,2)-tensor that is formally defined by

$$\mathbf{II}(v_p, w_p) = -\langle \mathrm{d}\nu(v_p), w_p \rangle,$$

where $\mathrm{d}\nu$ is called the *Weingarten map* — it is the differential of the Gauss map ν that assigns a unit vector to every point of the manifold [20]. The second fundamental form is characterized by three coefficients e, f, g:

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

These coefficients can be determined using the unit normal vector $\boldsymbol{n}$,

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

The coefficients are then given in terms of $\boldsymbol{r}$ and $\boldsymbol{n}$:

$$e = \langle \boldsymbol{r}_{uu}, \boldsymbol{n} \rangle \qquad f = \langle \boldsymbol{r}_{uv}, \boldsymbol{n} \rangle \qquad g = \langle \boldsymbol{r}_{vv}, \boldsymbol{n} \rangle, \tag{3-4}$$

with $\boldsymbol{r}_{uu} = \dfrac{\partial^2 \boldsymbol{r}}{\partial u^2}$ etc.

[3] 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 [20].

The first and second fundamental form can then be used to determine the Gaussian curvature of the parametric surface: [19, 20]

$$k = \frac{\det(\mathbf{II})}{\det(\mathbf{I})} = \frac{eg - f^2}{EG - F^2}. \tag{3-5}$$

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

1. Hyperbolic geometry is the geometry defined by 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 proportional to the area of a triangle by the negative factor 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 wishes to find a surface that exhibits constant negative Gaussian curvature. Some candidates exist that meet this criterion, the simplest of which is the so-called *pseudosphere*.

3-3 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 *pseudosphere*; which is the surface of revolution for 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 its axis has constant length ρ. Figure 3-1a shows an example of such a curve. A parametric representation of the tractrix γ is given by

$$\gamma : u \mapsto (\rho \operatorname{sech}(u),\ \rho(u - \tanh(u))) \quad u \in \mathbb{R}^+. \tag{3-6}$$

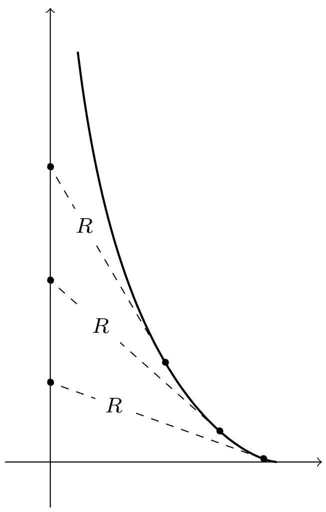
A parameterization of the pseudosphere in terms of u and v is then easily obtained by introducing a second parameter v that indicates the angle of rotation around the vertical axis:

$$\boldsymbol{r}(u, v) = \begin{pmatrix} \rho \operatorname{sech}(u)\cos(v) \\ \rho \operatorname{sech}(u)\sin(v) \\ \rho(u - \tanh(u)) \end{pmatrix} \quad u \in \mathbb{R}^+,\ v \in [0, 2\pi). \tag{3-7}$$

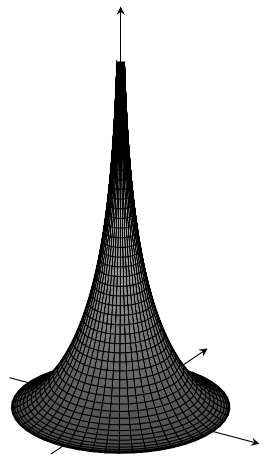
A pseudosphere is visualized by fig. 3-1b; the mesh on the surface consists of the isoparametric curves for u and v.

Curvature of the pseudosphere

As explained in section 3-2, the Gaussian curvature of a parametric surface may be calculated using the first and second fundamental form. The constant negative curvature of the pseudosphere will be formally demonstrated in this section. Let the pseudosphere be parameterized



(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 is generated by rotating the tractrix around the vertical axis.

Figure 3-1

as in eq. (3-7). Then, by virtue of eq. (3-2), the coefficients of the first fundamental form E, F and G are

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

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

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

Consequently, the coefficients of the second fundamental form are

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

using eq. (3-4). The Gaussian curvature can then be computed as follows [19]

$$k = \frac{\det(\mathbf{II})}{\det(\mathbf{I})} = \frac{eg - f^2}{EG - F^2} = \frac{-\rho^2 \operatorname{sech}^2(u)\tanh^2(u)}{\rho^4 \operatorname{sech}^2(u)\tanh^2(u)} = -\frac{1}{\rho^2}. \tag{3-8}$$

As such, the pseudosphere is shown to have a constant negative curvature everywhere but on the rim, where it ceases to be differentiable.

Because the pseudosphere exhibits constant negative curvature, it shares many properties of the hyperbolic plane. For example, the angular excess equation eq. (3-1) 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[4]). 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 [22]. This is why one must resort to models of the hyperbolic plane, they will be the subject of the next section.

3-4 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 smooth surface with the intrinsic geometry of the pseudosphere* [18]. 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 curvature-related problem that normal cartographers also face, in that the Earth can never be completely mapped on a flat surface, which is why either (i) a suitable projection method must be used or (ii) they must resort to an additional dimension and construct 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 is introduced in section 3-4-1. In section 3-4-2, the Cayley transform is used to map the Poincaré half plane to the Poincaré disk. Finally, the hyperboloid model, treated in section 3-4-3, embeds the hyperbolic plane in a three-dimensional Lorentzian space (cf. chapter 2).

3-4-1 The Poincaré half-plane

It has been pointed out that there are essentially two incompatibility problems with the pseudosphere that prevent one from devising 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

[4]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 [21]. This concept is defined more rigorously in terms of the *fundamental group* π_1, a heavily used invariant of topological spaces [13].

conformal mapping between the surface of the unit pseudosphere ($\rho = 1$) 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 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 2π traveled for v. 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 $\tilde{x}$ corresponds the angle v.

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 $\mathrm{d}x$ therefore corresponds on the pseudosphere with a traveled distance of $\mathrm{d}s = \operatorname{sech}(v)\,\mathrm{d}v$, because of the radius of the circle at that particular height. Let σ denote the distance traveled *along* any tractrix. Because the mapping is conformal, a movement $\mathrm{d}\sigma$ along a tractrix 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 using the parameterization:

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

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

$$\mathrm{d}y = \sinh(u)\,\mathrm{d}u \implies y = \cosh(u) + C. \tag{3-9}$$

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' the Euclidean metric of the pseudosphere:

$$\mathrm{d}s = \left\| \frac{\partial \boldsymbol{r}}{\partial u}\,\mathrm{d}u + \frac{\partial \boldsymbol{r}}{\partial v}\,\mathrm{d}v \right\| = \operatorname{sech}(u)\sqrt{\sinh^2(u)\,\mathrm{d}u^2 + \mathrm{d}v^2} = \frac{\sqrt{\mathrm{d}x^2 + \mathrm{d}y^2}}{y} \tag{3-10}$$

This metric is called the *Poincaré metric*. Equation (3-9) 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 traveling 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, because they would have to travel for an infinite amount of time. The line $y = 0$ 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].

A consequence of the Poincaré metric is that geodesics in the Poincaré half-plane are either

- straight vertical lines if the two points have the same x-coordinate, or
- 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 [23].

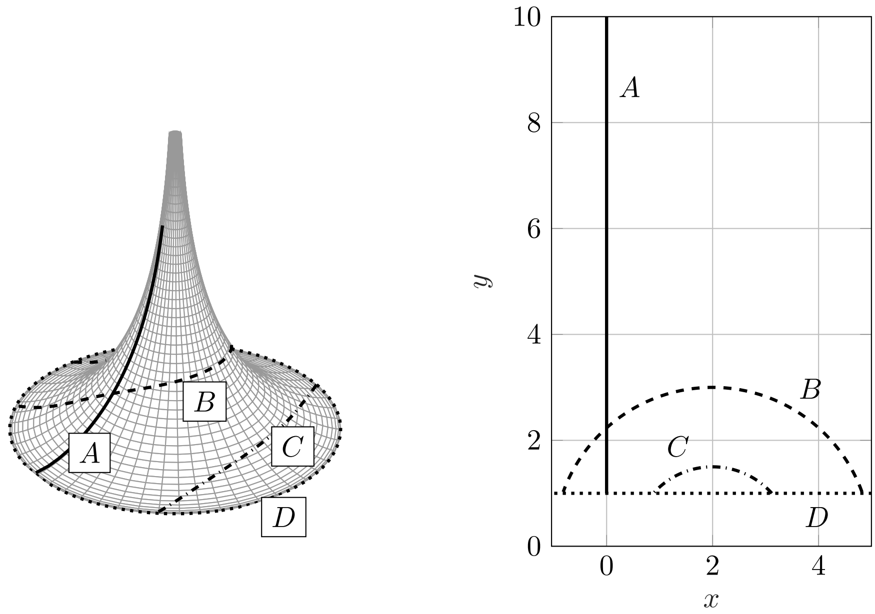


Figure 3-2: Comparison of various trajectories on the pseudosphere (left) and the Poincaré half-plane (right). Lines A, B and C are geodesics. The endpoints of B are closer together because it covers a wider range on the x-axis — if it would be larger than 2π, the curve would show an entire encirclement of the pseudosphere. Line D corresponds to the rim of the pseudosphere and the line $y = 1$ in the pseudosphere.

3-4-2 The Poincaré disk

The Poincaré disk arises naturally from the Poincaré half-plane by virtue of the Cayley transformation . When the Poincaré half-plane is viewed 'on top' of the complex plane, the Cayley transformation $\mathfrak{C}$ is defined as

$$\mathfrak{C}(z) = \frac{\mathrm{i}z + 1}{z + \mathrm{i}} \quad z \in \mathbb{C}, \tag{3-11}$$

which maps the entire half-plane to the unit disk. Furthermore, the following observations can be made:

- the horizon coincides with the rim of the disk
- the points ± 1 are invariant
- i maps to the origin
- the origin corresponds to $-\mathrm{i}$

As will become clear in chapter 4, eq. (3-11) 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. (3-10) will have the y replaced by the distance from the rim of the unit disk, which is $1 - x^2 - y^2$. Indeed, the half-plane metric can be written as in terms of the complex variable z, using the mapping eq. (3-11) 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

$$\mathrm{d}s = \frac{2\sqrt{\mathrm{d}x^2 + \mathrm{d}y^2}}{1 - x^2 - y^2}. \tag{3-12}$$

It has been pointed out that the Cayley transformation is a member of a larger class called Möbius transforms, which will be elaborated upon in chapter 4. Apart from being conformal, Möbius transforms also preserve circles (i.e. the result of a Möbius transform applied to a circle will again be a circle).

The geodesics in the half-plane have either the shape of straight lines extending from the horizontal axis, or circles that cross the horizontal axis at a right angle. Consequently, it can be shown that the geodesics in the Poincaré disk are also circles that depart orthogonally from the rim of the unit disk. Additionally, all diameters of the Poincaré disk are geodesics as well: they can be interpreted as circles with an infinite radius. Section 3-4-2 shows a comparison between some trajectories in the Poincaré half-plane and their mappings in the Poincaré disk through the Cayley transformation.

3-4-3 The hyperboloid model

Hilbert's theorem concerning the embedding of a hyperbolic surface in Euclidean space eradicates any hope to find a 'globe' for the hyperbolic plane, like a regular sphere would be for a 'spherical' plane (i.e. a plane exhibiting spherical geometry). However, as strong as it may seem, Hilbert's theorem still leaves some room for other possibilities, because it is very explicit about the nature of the surrounding space: it must be Euclidean. Chapter 2 already introduced a possible alternative: the Lorentz space. It turns out 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 space specifically? Recall that the curvature of a surface can be expressed in terms of its local radius, i.e. $1/\rho^2$. A negative curvature, therefore, implies that ρ is somehow not a real but an imaginary number. This would not make any sense in a Euclidean space, but the Lorentzian space perfectly allows this kind of odd witchcraft, as illustrated by fig. 2-1 [12]. One can think of the hyperboloid as a 'sphere' with constant imaginary radius embedded in the Lorentz space. The complete hyperboloid consists of two sheets: a positive sheet and a negative sheet, implicitly described by the following set:

$$H^2 = \left\{ \boldsymbol{x} \in \mathbb{R}^3 \;\middle|\; \|\boldsymbol{x}\|_L = \rho \right\}.$$

However, on account of the positive and the negative sheet, this set is clearly *disconnected*[5]. It is therefore common to discard the negative sheet H^2_- (cf. the sheets of the higher-dimensional

[5] 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).

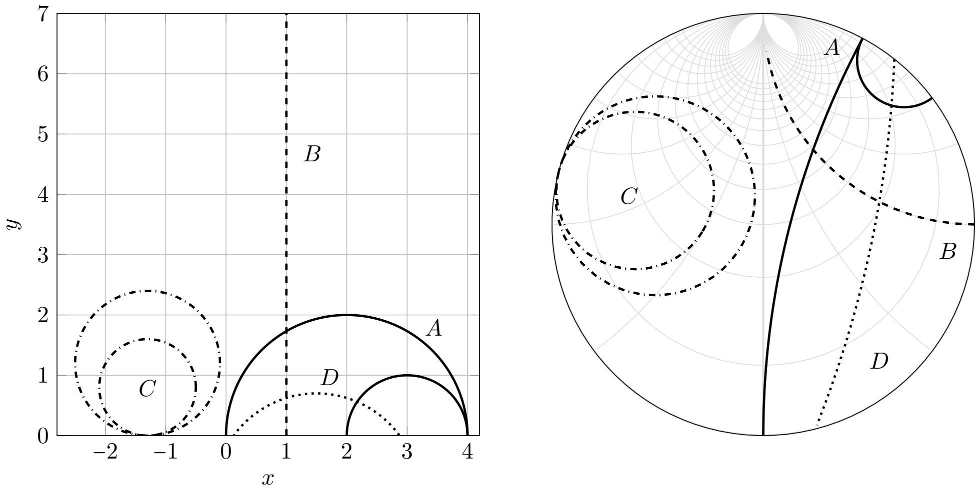


Figure 3-3: Comparison between trajectories in the Poincaré half-plane (left) and the Poincaré disk (right). Several types of trajectories are shown: the solid lines A are 'typical' geodesics, i.e. circles with finite radius in the half-plane. The dotted line B is also a geodesic but never reaches the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

hyperbola in section 2-4). The hyperboloid model of the hyperbolic 2-space is then [12]

$$H_+^2 = \left\{ \boldsymbol{r} \in \mathbb{R}^3 \;\middle|\; \|\boldsymbol{r}\|_L = \rho,\ r_0 > 0 \right\} \quad \rho \in \mathbb{R}_0^+.$$

A parametric representation of H_+^2 is the following

$$\boldsymbol{r} = \begin{pmatrix} \rho\cosh(u) \\ \rho\cos(v)\sinh(u) \\ \rho\sin(v)\sinh(u) \end{pmatrix}. \tag{3-13}$$

To show that this surface has indeed a constant negative Gaussian curvature, a slightly more general approach must be followed as for the pseudosphere, since the calculations with the second fundamental form are restricted to a Euclidean space. As described by O'Neill [24],

$$E = \partial_u \bullet \partial_v \quad F = \partial_u \bullet \partial_v \quad G = \partial_u \bullet \partial_v,$$

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. (3-13). The Gaussian curvature of the surface is then given by [24]

$$k = \frac{-1}{eg}\left(\varepsilon_1 \left(\frac{g_u}{e} \right)_u + \varepsilon_2 \left(\frac{e_v}{g} \right)_v \right) \tag{3-14}$$

with $\varepsilon_1 = \operatorname{sgn} E$, $\varepsilon_2 = \operatorname{sgn} G$, $e = \sqrt{|E|}$ and $g = \sqrt{|G|}$. Performing these calculations on eq. (3-13) yields indeed that $k = -1/\rho^2$, confirming that the one-sheet hyperboloid embedded in three-dimensional *Lorentz* space $\mathbb{R}^{1,2}$ indeed has the same curvature as the pseudosphere[6].

In the hyperboloid model, geodesics are formed by intersections between the hyperboloid and planes that intersect the origin. As mentioned in chapter 2, the restricted Lorentz group $\mathrm{SO}^{+}(1,2)$ is the isometry (distance-preserving) group for the hyperboloid in the Lorentzian space[7]. An action from the isometry group always maps geodesics to geodesics; as such, the Lorentz transformations preserve the geodesics of the hyperboloid model.

Relation with the Poincaré disk

There exists a compelling relation between the Poincaré disk model of the hyperbolic plane and the hyperboloid model. As described by Balazs and Voros [1], the Poincaré disk can be obtained by projecting the hyperboloid stereographically on the unit disk with respect to the base point $(0,0,-1)$. This is demonstrated in fig. 3-4. A point on the unit hyperboloid with coordinates

$$(\cosh(u),\ \sinh(u)\cos(v),\ \sinh(u)\sin(v)),$$

will then have the following polar coordinates in the disk:

$$r = \tanh\left(\frac{u}{2}\right) \qquad \phi = v.$$

Chapter summary

Non-Euclidean geometry arises when the parallel axiom is 'removed' from Euclidean geometry; this leads to two other geometry types: spherical geometry and hyperbolic geometry. The validity of the parallel axiom and the type of geometry are related to the Gaussian curvature of the surface on which they are applied. Hyperbolic geometry is the geometry of surfaces with constant negative Gaussian curvature. The pseudosphere is an object that exhibits constant negative curvature embedded in three-dimensional Euclidean space, but it is not globally isometric to the hyperbolic plane. Alternatively, there are models of the hyperbolic plane that do manage to capture its entirety: the Poincaré half-plane, the Poincaré disk and the hyperboloid model. The latter consists of the unit hyperboloid embedded in three-dimensional Lorentz space; its isometry group is the Lorentz group acting on that space. From the hyperboloid model, the Poincaré disk can be obtained through stereographic projection on the unit disk. A particular Möbius transformation, called the Cayley transformation, provides the mapping between the Poincaré disk and the Poincaré half-plane.

[6] Some authors even refer to the hyperboloid as 'pseudosphere' as well, e.g. Balazs and Voros [1]. However, this is deemed confusing in the context of the larger discussion presented here, which is why the author refrained from using it here.

[7] Technically, an inversion of spatial orientations is allowed in this case, so the complete isometry group of the hyperboloid model comprises the ortochronous Lorentz transformations, both proper and improper, denoted by $\mathrm{O}^{+}(1,2)$.

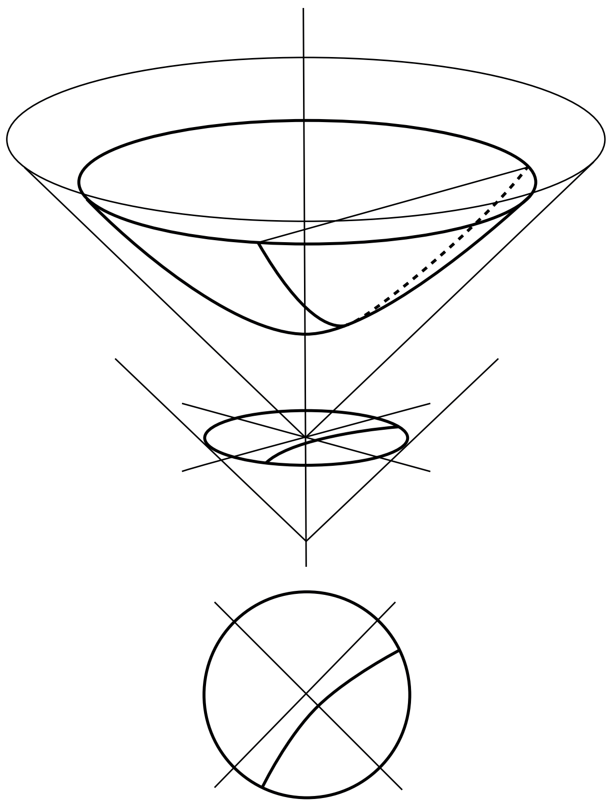


Figure 3-4: The Poincaré disk and the hyperboloid model are related through stereographic projection. A geodesic of the on the hyperboloid model is shown as well, its projected image in the Poincaré disk (also a geodesic) is part of a circle that intersects the boundary of the disk at right angles. Illustration adapted from Balazs and Voros [1].

Chapter 4

Möbius transformations

This chapter deals with a type of complex-valued functions called Möbius transformations, in particular their relation with Lorentz transformations and hyperbolic geometry. First, Möbius transformations are defined in section 4-1. Then, Möbius transformations and the corresponding Möbius group are considered from the perspective of the Riemann sphere in section 4-2. In section 4-3, the Möbius transformations are classified into hyperbolic, parabolic, elliptic and loxodromic subclasses. Finally, sections 4-4 and 4-5 discuss the subgroups of the Möbius group and their relation with other notable Lie groups.

4-1 Definition and basic properties

Möbius transformations, named after the German mathematician and astronomer August Ferdinand Möbius, constitute an important class of complex-valued functions that find many applications in mathematics and physics [2].

Möbius transformations

A Möbius transformation $\mathfrak{M}\colon \mathbb{C} \to \mathbb{C}$ is a linear fractional transformations of the form

$$\mathfrak{M}(z) = \frac{az + b}{cz + d} \tag{4-1}$$

where $a, b, c, d \in \mathbb{C}$ are constants [2].

These complex transformations carry a deep connection with hyperbolic geometry (and non-Euclidean in general) and Einstein's theory of relativity. As stated by Needham [2], the complex mappings that correspond to a Lorentz transformation are Möbius transformations and vice versa — every Möbius transformation 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:

(i) a first translation $z \mapsto z + \frac{d}{c}$;

(ii) a *complex inversion* $z \mapsto \frac{1}{z}$;

(iii) an expansion and rotation $z \mapsto -\frac{ad - bc}{c^2} z$;

(iv) a second translation $z + \frac{a}{c}$.

Each of these transformations is conformal and preserves circles, which is why general Möbius transformations inherit these properties as well.

It is quite clear from eq. (4-1) 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 Möbius transformation is a bijective mapping; its inverse is then given by [2]

$$\mathfrak{M}^{-1}(z) = \frac{dz - b}{-cz + a}. \tag{4-2}$$

4-2 The Möbius group

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 $\mathfrak{M}^{-1}$ in eq. (4-2), and the composition of two transformations $\mathfrak{M}_2 \circ \mathfrak{M}_1$ again yields a member of the class of Möbius transformations [2]. The group of nonsingular Möbius transformations is called the Möbius group, denoted by 'Möb '.

4-2-1 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 the representation of the extended complex plane[1] $\hat{\mathbb{C}}$ in three-dimensional space. That is,

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

The Riemann sphere can be visualized by placing the complex plane horizontally and considering a unit sphere centered around the origin, such that its intersection with the complex plane coincides with the unit disk. To map the plane to the sphere, a stereographic projection 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 [2].

[1] The complex plane combined with a value for infinity '∞'.

In order to see how the Riemann sphere relates to the transformations of the form eq. (4-1), 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 by $[\mathfrak{z}_1, \mathfrak{z}_2]$. This pair of complex numbers is subject to an equivalence relation that makes them projective coordinates, namely

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

with β being any nonzero complex number. The projective space is then formed by the quotient set of the complex 2-space (excluding the origin) modulo 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 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 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 $\mathbb{CP}^1$ [18]. The compactification arises from the fact that in the complex plane, one can venture in any arbitrary direction to obtain the 'same' infinity point, in contrast to the signed infinities $\pm\infty$ that are associated with the real line. This single infinity point essentially zips the edges of the plane together into a compact manifold.

4-2-2 Matrix representation

There is a compelling correspondence between Möbius transformations is and complex matrices of size 2×2 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 the multiplication of a constant, so is the 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 transformation, since multiplying the entire matrix with -1 would still correspond to the same Möbius transformation. Now, how can one actually effect a Möbius transformation and 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 transformation. This provides an alternative way of calculating the result of any Möbius transformation by means of matrix multiplication. Furthermore, the composition of two Möbius transformations, $\mathfrak{M}_1$ and $\mathfrak{M}_2$, can simply be obtained by multiplying their corresponding matrices (denoted by M_1 and M_2):

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

[2] In physics, these objects are also called *2-spinors*; they are fundamentally connected with the theory of relativity, as demonstrated by [25].

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

4-2-3 The Möbius group as PGL(2,$\mathbb{C}$)

Based on the matrix analogy, one can state that Möb is the group of linear[3] transformations on vector space $\mathbb{C}^2$ — as projective coordinates. This is equivalent to the statement 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* $\mathrm{PGL}(2,\mathbb{C})$. It was mentioned before that that the Möbius group is the automorphism group of the Riemann sphere; the fact that Möb $\cong \mathrm{PGL}(2,\mathbb{C})$ is due to the fact that the Riemann sphere 'is' the projective complex line: it consists of all directions in the complex 2-space.

It has already been established that Möbius transformations 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 itself: 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 nth root within $\mathbb{F}$. The field of complex numbers is algebraically closed, hence, the former holds for $\mathbb{F} = \mathbb{C}$. 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.

4-3 Classification of Möbius transformations

The analogy between Möbius transformations and matrix multiplication begs the question what role the eigenvalues and eigenvectors of the corresponding matrix play. Eigenvectors are vectors that remain invariant (up to scaling) under the multiplication of a particular matrix. In this case, the vector contains projective coordinates, so even when it is scaled, its coordinate representation remains identical. As such, the eigenvectors of the matrix M are the *fixed points* of the Möbius transformation $\mathfrak{M}$; any Möbius transformation has two at most. This is demonstrated by solving $z_0 = \mathfrak{M}(z_0)$, 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_0 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 transformations are called *parabolic*, more will become clear about them later [2].

Now, it remains to analyze the significance of the eigenvalues. Since eigenvalues are sensitive to scaling of the matrix, it is important to stress again that M must be normalized (have

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

a unit determinant) in order for the following to hold. A well known fact in linear algebra states that if λ_1, λ_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:

$$\lambda + \frac{1}{\lambda} = a + d = \operatorname{tr} M. \tag{4-3}$$

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

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

because then of course $\mathfrak{J}(z) = \frac{\sqrt{\kappa} z}{1/\sqrt{\kappa}} = kz$. Because conjugacy translates to a similarity transformation in the matrix analogy, it leaves the eigenvalues of the matrix unaffected. But, since J is a diagonal matrix, its eigenvalues are exactly on the main diagonal. As a result, *the multiplier of a Möbius transformation is equal to the square of its eigenvalue*, or $\kappa = \lambda^2$. Strictly speaking, every Möbius transformation has two eigenvalues and two multipliers, but since they are both each other's reciprocal, they do not have to be considered separately. With this result, eq. (4-3) can then also be restated in terms of the multiplier κ instead of the eigenvalues:

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

Solving eq. (4-3), 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 Δ one can then distinguish three possible cases:

1. $\Delta < 0$ or $(a+d)^2 < 4$: there are two complex solutions for λ. 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, a natural further categorization arises:

   (a) if $(a+d)^2 \in [0, 4)$, the argument of the square root is positive: consequently, the solutions for λ 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,

[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 transformations (and therefore the notion of similar matrices) in linear algebra.

by virtue of Euler's formula, be written as $\lambda = e^{\frac{i\theta}{2}} = \cos\left(\frac{\theta}{2}\right) + i\sin\left(\frac{\theta}{2}\right) \neq 1$, such that the multiplier $\kappa = \lambda^2 = e^{\theta i}$ — the factor of one half is only there to identify the multiplier with the actual angle θ, which is the most meaningful from a geometric standpoint. The associated matrix archetype or Jordan form for transformations of this type is:

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

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 transformations are rotation matrices.* These transformations are called *elliptic.* The edge case for which $(a + d)^2 = 0$ and $\kappa = -1$, is denoted as a *circular* transformation (which is still an elliptic transformation).

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

2. $\Delta = 0$ or $(a + d)^2 = 4$: there is one solution for λ, 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 transformation is only unique up to a sign. The multiplier for both cases is the same: $\kappa = 1$. 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 transformations 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:

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

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

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 λ, given by

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

The resulting solutions for λ are then always real and positive; they can then be expressed as the image of the exponential function: $\lambda = e^{\frac{\zeta}{2}}$ such that $\kappa = e^{\zeta}$. The usage of ζ is not at all coincidental: indeed, the argument here represents a *hyperbolic angle*, like the Lorentz boosts discussed in chapter 2. 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:

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

As such, these transformations are akin to hyperbolic rotations, which is why they are also classified as *hyperbolic*. The matrix is in this case also part of $\mathrm{SO}^+(1,1)$ (cf. section 2-4). The hyperbolic transformations are also part of the class of loxodromic transformations, together with the aforementioned class, where λ is complex. They do, however, deserve their own subclass because, apart from frequently encountered, they also represent a particular component of the special linear group over the reals $\mathrm{SL}(n, \mathbb{R})$.

To summarize, there are five different classes of Möbius transformation: circular, elliptic, hyperbolic, loxodromic, and parabolic. Circular transformations are part of the elliptic transformations and hyperbolic transformations are a subclass of loxodromic transformations. The class to which a Möbius transformation belongs is determined completely by its trace $a+d$, or equivalently, the value of the multiplier κ. For the multiplier, one can distinguish several 'regions' in the complex plane that are each associated with a class of Möbius transformations, this is visualized in fig. 4-1. The nature of the Jordan form of the matrix associated with a Möbius transformation also clearly gives away to which class a transformation belongs.

Table 4-1 provides an overview of the five Möbius classes, together with the values for the matrix trace, the multiplier, and the Jordan form. Finally, fig. 4-2 visualizes the effect of each of the transformation classes on points on the Riemann sphere. Elliptic transformations push points along with circles of constant latitude, while hyperbolic transformations move points orthogonally, along the meridians of the Riemann sphere.

4-4 Subgroups of the Möbius group

Apart from the classification considered in the preceding chapter, three subgroups of the Möbius group can be distinguished as well. A compelling observation is that each of these subgroups can each be identified with one of the 'geometry types' that were considered in

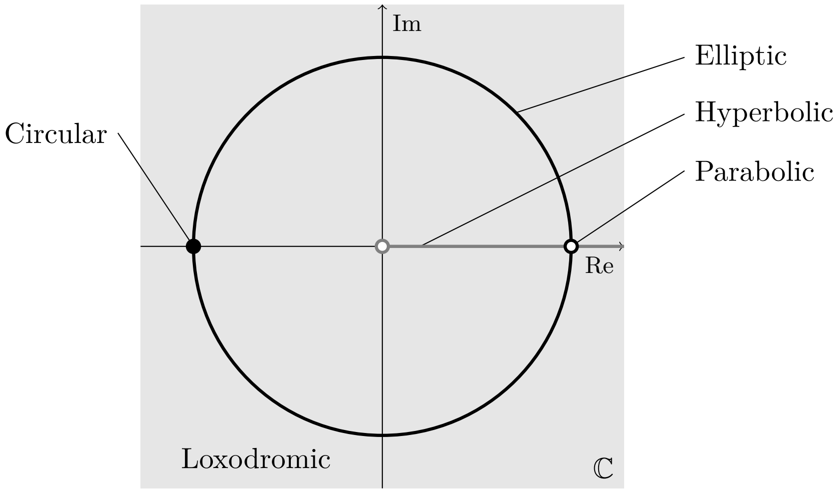


Figure 4-1: Classification of Möbius transformation in terms of the location of the multiplier κ in the complex plane. Any point that is not on the unit circle yields a loxodromic transformation; a particular subclass consists of the hyperbolic transformations, which are on the real axis except at -1 and 1 where it intersects with the unit circle, and at the origin, where the transformation becomes singular. If the multiplier lies on the unit circle (apart from $\kappa = 1$) the transformations are elliptic. A special case is the circular transformation for $\kappa = -1$. Finally, the parabolic transformations have a multiplier of 1 [2].

Table 4-1: Overview of the five classes of Möbius transformations and the corresponding values for the trace squared of the matrix ($\operatorname{tr} M = a + d$), the multiplier of the transformation, 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	$\{\kappa \in \mathbb{C} \mid \lvert\kappa\rvert = 1, \kappa \neq 1\}$	$[0,4)$	$\begin{pmatrix} \mathrm{e}^{\theta\mathrm{i}/2} & 0 \\ 0 & \mathrm{e}^{-\theta\mathrm{i}/2} \end{pmatrix}$	—
Parabolic	$\{1\}$	$\{4\}$	$\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}$	—
Hyperbolic	$\mathbb{R}_0^+ \setminus \{1\}$	$(4, +\infty)$	$\begin{pmatrix} \mathrm{e}^{\zeta/2} & 0 \\ 0 & \mathrm{e}^{-\zeta/2} \end{pmatrix}$	Loxodromic
Loxodromic	$\{\kappa \in \mathbb{C} \mid \lvert\kappa\rvert \neq 1\}$	$\mathbb{C} \setminus [0,4]$	$\begin{pmatrix} \kappa & 0 \\ 0 & \kappa^{-1} \end{pmatrix}$	—

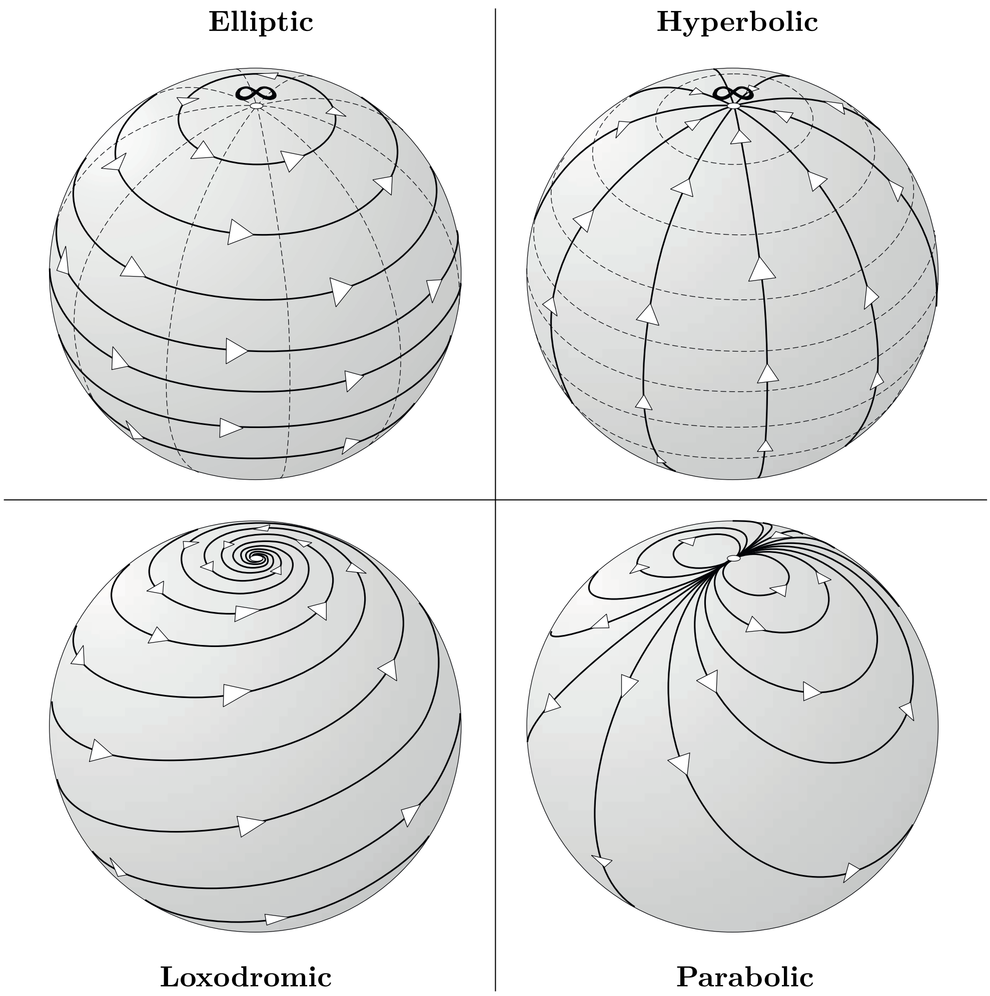


Figure 4-2: Overview of the four classes of Möbius transformation and their typical action on the Riemann sphere. The curves shown on the Riemann sphere are the *invariant curves* of the transformation, i.e. these curves as a whole remain invariant under the transformation. The elliptic, hyperbolic, and loxodromic transformations have the North and South pole, or ∞ 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 taking a loxodromic path would maintain a constant angle with respect to true North. Illustration reprinted from Needham [3, p. 78].

chapter 3: those of positive (spherical), zero (Euclidean), and negative (hyperbolic) Gaussian curvature. A specific subgroup of the Möbius group will play the role of the direct isometry group on a particular representation of the surfaces that exhibit these geometry types [3].

4-4-1 Euclidean geometry

The direct isometries in the Euclidean plane consist simply of translations and rotations. A rotation in the complex plane (around the origin) can be represented simply by $z \mapsto e^{\mathrm{i}\theta}$, 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 the particular Möbius transformations of the form

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

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

4-4-2 Spherical geometry

A well-known application of group theory in engineering is the representation of rotations in three-dimensional Euclidean space. 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 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: [26]

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

which is a general representation of a 2×2 *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 transformation (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 transformation 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 and opposite. Therefore, the general expression of a rotation of the Riemann sphere as a Möbius transformation can be written as: [3]

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

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

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 transformations: as such, the ambiguities are eliminated, and the group of transformations of type $\mathfrak{S}$ is isomorphic to $\mathrm{SO}(3,\mathbb{R})$. In the complex plane, these transformations represent the isometries of spherical geometry in the stereographic map [2].

4-4-3 Hyperbolic geometry

As described by Rovenski [27], the isometries of the Poincaré half-plane (cf. section 3-4-1) are any (composition) of the following types of transformations, i.e. they leave distances according to the Poincaré metric unaffected:

- horizontal translations: $(x,y) \mapsto (x+a,y)$ where $x,y,a \in \mathbb{R}$;
- 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 $(x,y) \mapsto \left(\dfrac{x}{x^2+y^2}, \dfrac{y}{x^2+y^2}\right)$.

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 transformations are the isometries of the Poincaré half-plane [3]. Recall from section 3-4-1 that there are two types of geodesics in the half-plane: semicircles centered at the origin and straight vertical lines. These are precisely invariant curves for the transformations listed (isometries map geodesics to geodesics) [17].

4-5 Relation with other groups

It has been established that the Möbius group as the automorphism group of the Riemann sphere is 'trivially' isomorphic to the projective linear group over the complex numbers $\mathrm{PGL}(2,\mathbb{C})$, which for this case is identical to the projective special linear group over the complex numbers $\mathrm{PSL}(2,\mathbb{C})$. This triviality is due to the fact that the Riemann sphere is equal to the complex projective line. Furthermore, it mentioned in chapter 2 that the Möbius group is isomorphic to the restricted Lorentz group $\mathrm{SO}^+(1,3)$. As described by Needham [3], the four-vectors from special relativity may be associated with the projective coordinates as follows:

$$\begin{pmatrix} t+z & x+\mathrm{i}y \\ x-\mathrm{i}y & t-z \end{pmatrix} = 2\begin{pmatrix} \mathfrak{z}_1 \\ \mathfrak{z}_2 \end{pmatrix}\begin{pmatrix} \mathfrak{z}_1 \\ \mathfrak{z}_2 \end{pmatrix}^H,$$

this is the *spinorial representation* of four-vectors. One can also see this as a parameterization of any vector in the future (positive direction of time) light cone in terms of two complex numbers. The determinant of this matrix coincides with the spacetime interval, which is the quadratic form that the Lorentz transformations must preserve. As such, a Möbius transformation effected on the projective coordinates will preserve this interval as well. It can be shown that this parameterization yields an isomorphism between the Möbius group and the restricted Lorentz group [25]. Furthermore, as discussed in chapter 2, the classification of Möbius transformations carries over to the Lorentz transformations too, where hyperbolic elements are Lorentz boosts and elliptic elements are spatial rotations.

Another classic Lie group is the special linear group $\mathrm{SL}(n, \mathbb{R})$, 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 2×2 matrices with unit determinant. It has been mentioned previously that any such matrix coincides with a Möbius transformation, albeit surjectively: for every Möbius transformation, 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 2×2 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 for the normalized Möbius transformations, 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. 4-1, that means that the matrices with real entries contain exactly the hyperbolic (both real), elliptic (unit circle), and parabolic classes. The remaining part consists of the loxodromic transformations which have a nonreal trace. $\mathrm{SL}(2, \mathbb{R})$ is in turn isomorphic to two other important groups: the two-dimensional symplectic group Sp(2,) and the special unitary group $\mathrm{SU}(1, 1)$. As discussed in the previous section, this is the same class of transformations that also constitutes the isometry group of the Poincaré half-plane.

Finally, the isomorphisms of the subgroups of the Möbius transformations have been discussed in section 4-4. Three subgroups can be identified, denoted by $\mathfrak{S}$, $\mathfrak{E}$ and $\mathfrak{H}$, which are respectively isomorphic to the two-dimensional Euclidean group, the three-dimensional rotations ($\mathrm{SO}(3, \mathbb{R})$ covered by the quaternions in $\mathrm{SU}(2)$) and the special linear group $\mathrm{SL}(2, \mathbb{R})$ which also represents the isometries of the Poincaré half-plane.

The relations between the groups are visualized in fig. 4-3.

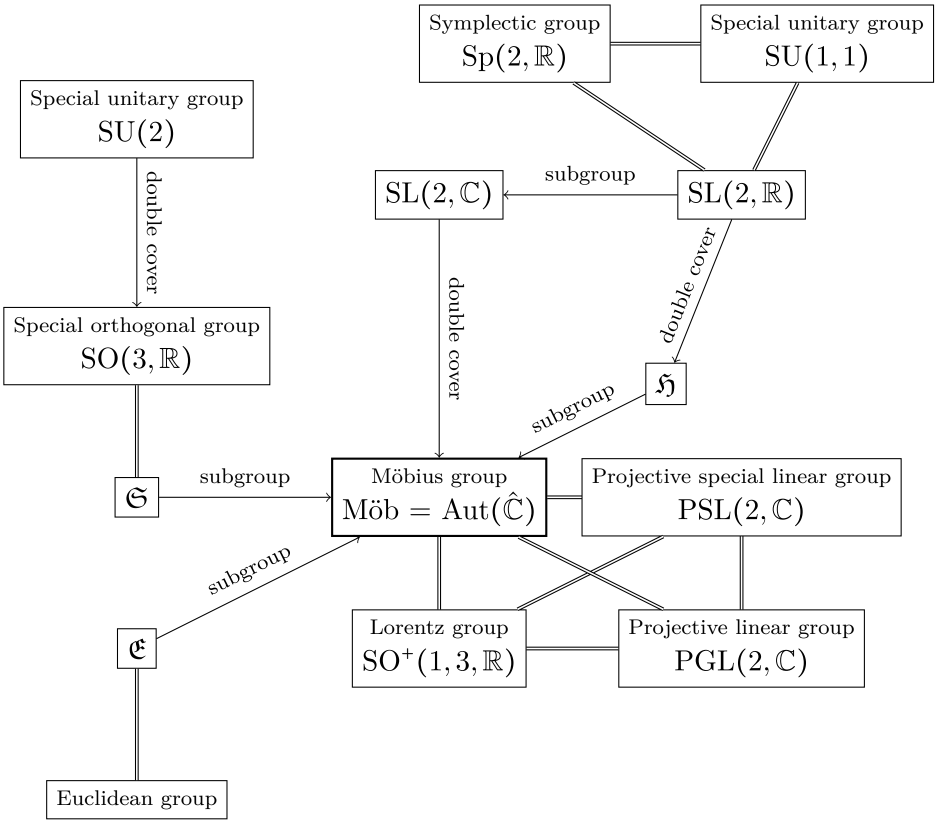


Figure 4-3: Schematic of the relation of the Möbius group with other important groups such as the special orthogonal group, special unitary group, Lorentz group, etc. Double lines represent isomorphisms, double covers and subgroups are indicated.

Chapter summary

Möbius transformations are complex-valued linear fractional transformations. They constitute a group under composition, called the Möbius group. The Möbius group is the automorphism group of the Riemann sphere. The Riemann sphere also represents the complex projective line; the Möbius group can therefore be thought of as the projective linear group over the complex numbers. The Möbius transformations are categorized into five classes: hyperbolic, elliptic, circular, loxodromic and parabolic, based on their conjugacy class in the Möbius group. There are three subgroups of the Möbius group, each of which can be identified as the isometry group (of some representation) of the three geometry types: hyperbolic geometry, spherical geometry and Euclidean geometry. Finally, an isomorphism exists between the restricted Lorentz group and the Möbius group through the spinorial representation of four-vectors.

Part II

Hyperbolic Rotations in Economic Engineering

Chapter 5

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]. Especially for the latter, a rich variety of useful analogs can be devised. There are two common approaches to mechanics: *analytical mechanics* including the formalisms developed by Joseph-Louis Lagrange and Sir William Rowan Hamilton, and *vectorial mechanics*, better known as *Newtonian mechanics*. Analytical mechanics are more common in physics due to their mathematical elegance and powerful theoretical foundation, while Newtonian mechanics prevail for most practical (engineering) applications. 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 subject of this chapter. 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 in section 5-2, for it has the most intuitive explanation. Then, in section 5-3, a more formal approach allows to (Legendre) transform the discussion into one of Hamiltonian mechanics.

This chapter is essentially a 'hybrid': in parallel with the discussion about analytical mechanics, the analogous concepts in economic engineering are introduced as well. A short introduction to economic engineering is provided in section 5-1. This discussion will mostly pertain to 'regular' economic systems, but it is the underlying aspiration of this research that these formal methods will allow establishing a sound framework for financial systems as well.

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:

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

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 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 rather theoretical discussion of analytical mechanics.

5-1 Introduction to 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. 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 pivotal idea in economic engineering is to extend 'domain neutral modeling techniques such as bond graph modeling [5] that are built on analogies between mechanical, electrical, hydraulic, ... systems to economic systems as well. More specifically, 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 macroeconomics or, on the other end of the spectrum, the 'black-box' econometric models that have lost all interpretative value. Indeed, applied economic engineering pursues *gray-box* modeling instead, as is most common in traditional engineering models [28]. Some analogies between mechanical, electrical and economic systems are listed in table 5-1 merely for the sake of illustration; a thorough motivation for each of them will be given in the following section. In general, their meaning is not as specific as given here and this table applies only to the most elementary cases.

The theoretical foundations of economic engineering have been developed by Mendel [29]. Some other notable results in the field have been achieved in recent years as well. Hutters and Mendel [30] 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 [31] to explain economic growth and productivity. Kruimer [28] and Van Ardenne [32] 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).

Table 5-1: 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 [5].

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

5-2 Lagrangian mechanics

Within analytical mechanics, Lagrangian mechanics is the most natural generalization for people familiar with Newtonian mechanics. Lagrangian mechanics defines the system as a 'configuration space' with an associated Lagrangian function that gives rise to the principle of stationary action.

5-2-1 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 coordinate tuple $\boldsymbol{q}$.

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 consequence of holonomic constraints present in the system, which are constraints that impose a restriction only on the configuration space, but not, for example, on the allowable velocities. Using the generalized coordinates, one can parameterize all the allowable configurations of a system, whose motions may occur in a higher-dimensional space, with a smaller coordinate set. For example, the two-dimensional motion of a pendulum can be expressed with 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 successful 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 constrain the motion of the system nevertheless by means of the *Lagrange multipliers*. They are applicable to holonomic constraints for which it is impractical or non-intuitive to take them into account directly in

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

the parameterization of the configuration manifold, and a restricted class of nonholonomic constraints that can be written in the so-called *Pfaffian* form [33].

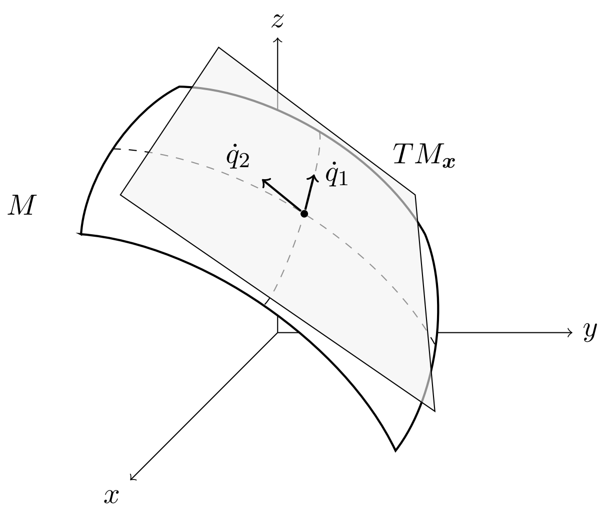


Figure 5-1: Schematic of a two-dimensional (configuration) manifold M embedded in $\mathbb{R}^3$; the local generalized velocities associated with a point $\boldsymbol{x}$ are vectors that live in the tangent space $TM_{\boldsymbol{x}}$.

The 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 an inner product.

Just like in mechanics, the simplest shape that 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 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.

A simple example in economic engineering might be a market with a supplier, a trader, and a demander, each with their own storage. Because the total amount of product is limited by the endowment of the system (the so-called Edgeworth Box), a constraint is imposed on the allowed configurations.

5-2-2 Hamilton's principle of stationary action

The configuration manifold 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 circuit, 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 $\mathcal{L}$ [34].

Hamilton's principle

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

$$\Phi(\gamma) = \int_{t_1}^{t_2} \mathcal{L} \, \mathrm{d}t \,, \tag{5-1}$$

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 is for now only helpful on a conceptual level, because it does not show 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, Φ. A necessary condition for Φ to attain an extremal is that

$$\delta\Phi = 0,$$

where $\delta\Phi$ is called the *first variation* of Φ. Just like a regular differential can be seen as a linear approximation to an infinitesimal perturbation of a function value, the variation is a linear approximation to 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* [34].

Landau and Lifshitz [35] use the the tools of the calculus of variations to show that the solution of eq. (5-1) is

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

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

which is called the *Euler-Lagrange equation*. 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 (co)vector $\frac{\partial \mathcal{L}}{\partial \dot{q}}$, 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 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 to a formalism that is actually useful, the Lagrangian must be assigned with a concrete mathematical 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 mechanical interpretation is pursued, neglecting the connection between temperature and energy — the application of economic engineering and thermodynamics is described in the thesis of Manders [31]. The economic engineering interpretation of work is the fulfilling of wants. As such, the energy of an economic 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 [36].
- *Potential energy* gives significance to the utility gained from the possession 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 [37]:

> *"[...] 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 maximize their convenience yield (the utility of the goods they possess) by sacrificing as little economic surplus as possible.

As described by Feynman [37], this is not only a global principle 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 assumption in many economic theories [36].

[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.

5-2-3 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* [34].

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

$$T^* = \frac{m}{2} \langle \boldsymbol{v}, \boldsymbol{v} \rangle \quad \text{with} \quad \boldsymbol{v} \in TM_{\boldsymbol{x}}. \tag{5-2}$$

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

$$T^*(\boldsymbol{q}, \dot{\boldsymbol{q}}) = \frac{1}{2} m_{ij}(\boldsymbol{q}) \dot{q}^i \dot{q}^j, \tag{5-3}$$

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 [34].

The distinction between energy and co-energy is not very common in literature, although thoroughly discussed by Jeltsema and Scherpen [38]. Energy is the ability to do work, while co-energy is the *complement of energy*. Because energy is defined in terms of work, kinetic energy T should be defined in terms of momentum – the integral of a force applied 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. The kinetic energy and co-energy are not always equal after all, e.g. in the relativistic case where the mass will depend on the velocity as well. In the simple non-relativistic case, for a single particle with mass m, the following relations hold:

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

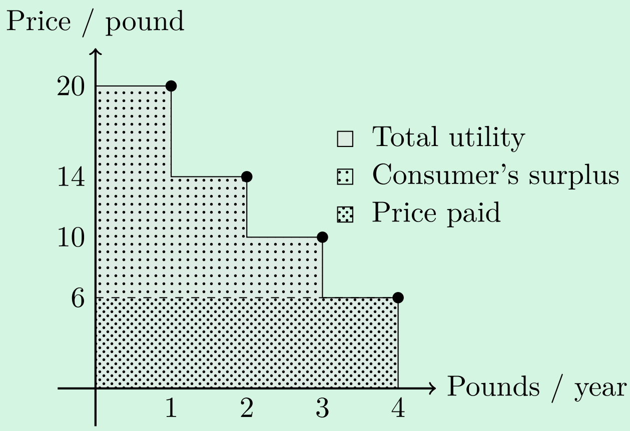
[4] As will become clear later, the term 'co-energy' is used to distinguish between the definition of kinetic energy in terms of the generalized momenta, which is the perspective of Hamiltonian mechanics.

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. (5-2) 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 [39, 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 unconscious deliberation between (i) how much she likes to drink tea and (ii) how much she likes to pay for tea — this is a consequence of the assumption that Jane is a rational market participant. Marshall quantifies Jane's unconscious 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 utility of the second pound of tea, i.e. how much Jane values it on top of the first one. The total utility of 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 consumer's surplus is 6\$.



Indeed, this example illustrates that *surplus* and *utility* are closely related; they only differ by the choice of a 'set point' (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:

$$\text{total surplus} = \sum_{i=1}^{i_m} p_i \Delta v \qquad \text{consumer's surplus} = \sum p_i \Delta v - \underbrace{\sum p_m \Delta v}_{\text{amount paid}}, \tag{5-4}$$

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 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 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.

In this analogy, the 'mass' plays the role of the market (in)elasticity, or the slope of the supply/demand line. Some generalizations are in order with respect to the classical notion of mass in mechanics. First of all, the elasticity is allowed to depend on the stock level, which is very common in economics (scarcity affects the supply and demand lines). Secondly, the product space (configuration manifold) is generally far from isotropic, i.e. mass is direction-dependent. This seems far from realistic for mechanical systems, but when one generalizes 'mass' to the mass matrix found in the linearization of nonlinear mechanical systems, both position dependence and anisotropy are frequently encountered (e.g. double pendulum).

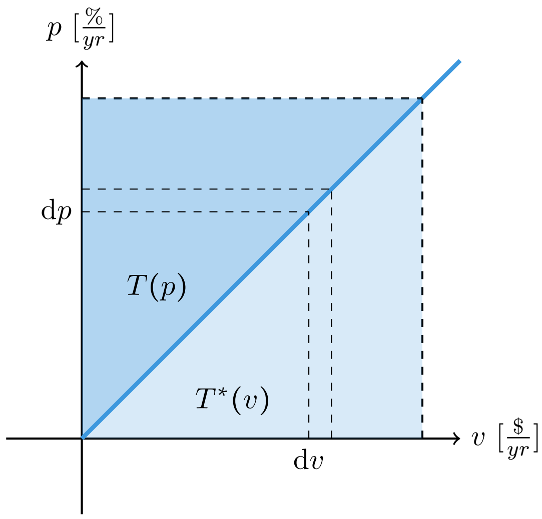


Figure 5-2: Kinetic energy and co-energy in terms of the relation between momentum and velocity. Figure courtesy of B. Krabbenborg [4].

5-2-4 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)

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

Potential energy and convenience yield

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

[5]This is a consequence of the fundamental theorem for line integrals [40].

utility gained from the possession 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 analog 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 level, which is why it is omitted in the theoretical discussion, 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 interpretation as well.

The stereotypical example in mechanics that contains a single storage element of kinetic energy (a mass) and a single storage element 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:

$$m\ddot{q} + \frac{\partial U}{\partial q} = 0 \quad \text{with} \quad U = \frac{kq^2}{2} \tag{5-5}$$

where k represents the spring constant; this immediately results in Newton's second law of motion. A simple mass-spring system is shown in fig. 5-3.

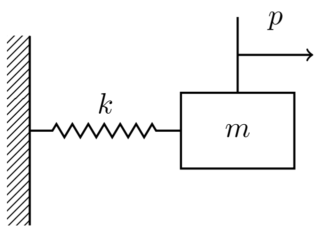


Figure 5-3: A simple mass-spring system. Figure courtesy of B. Krabbenborg [4].

The distinction 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 a 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. That is,

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

which for 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 5-4 shows the relation between potential energy and co-energy and why they are equal for a linear relation between force and displacement.

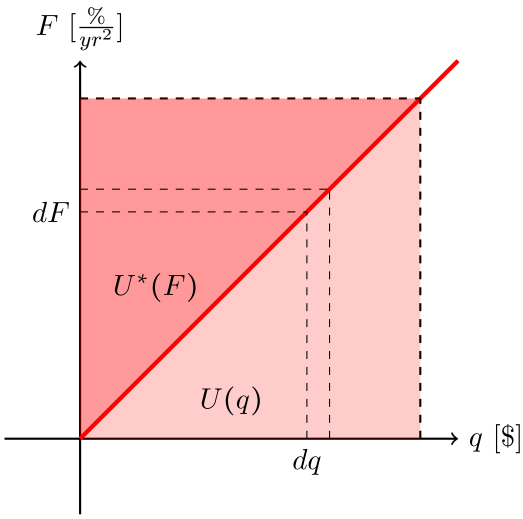


Figure 5-4: The distinction between potential energy and co-energy in terms of the relation between force and displacement. Figure courtesy of B. Krabbenborg [4].

The solution of the simple mass-spring system is a sinusoid that continues indefinitely in time. These 'perfect' sinusoids are not often encountered in real 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 interconnected nature of society. This makes the theoretical achievement that is the basic foundations of economic engineering perhaps even more admirable. Because of these external factors, eq. (5-5) 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.

5-3 Hamiltonian mechanics

From Lagrangian mechanics, one can arrive at Hamiltonian mechanics through the Legendre transformation. Hamiltonian mechanics are characterized by Hamilton's equations, which have found applications outside classical mechanics mechanics as well (e.g. optimal control theory). Hamiltonian mechanics have a strong geometric foundation; more specifically a the branch of mathematics called *symplectic geometry*.

5-3-1 The Legendre transformation

Whereas the Lagrangian of a system is a function of the generalized positions and velocities, the Hamiltonian is a function of the generalized positions and momenta. Recall from the definition that the generalized (or *conjugate*) momenta are defined in terms of the Lagrangian; $p_i = \frac{\partial \mathcal{L}}{\partial \dot{q}_i}$. This change of variables is performed in a particular fashion called the *Legendre transformation*.

The total differential of the Lagrangian is

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

Two observations can be made: first, by definition, $\boldsymbol{p} = \dfrac{\partial \mathcal{L}}{\partial \dot{\boldsymbol{q}}}$; and secondly, the Euler-Lagrange equations dictate that $\dfrac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}t} = \dfrac{\partial \mathcal{L}}{\partial \boldsymbol{q}}$. As such, the total differential of the Lagrangian may be rewritten as: [35]

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

By virtue of the product rule $\mathrm{d}(\boldsymbol{p}\dot{\boldsymbol{q}}) = \boldsymbol{p}\,\mathrm{d}\dot{\boldsymbol{q}} + \dot{\boldsymbol{q}}\,\mathrm{d}\boldsymbol{p}$, the above can be expressed as follows:

$$\mathrm{d}\,\underbrace{(\boldsymbol{p}\dot{\boldsymbol{q}} - \mathcal{L})}_{\mathcal{H}} = \dot{\boldsymbol{q}}\,\mathrm{d}\boldsymbol{p} - \dot{\boldsymbol{p}}\,\mathrm{d}\boldsymbol{q} - \frac{\partial \mathcal{L}}{\partial t}\,\mathrm{d}t\,. \tag{5-6}$$

The argument of the differential on the left-hand side is the *Hamiltonian* $\mathcal{H}$ of the system,

$$\mathcal{H}(\boldsymbol{q}, \boldsymbol{p}, t) \triangleq \boldsymbol{p}\dot{\boldsymbol{q}} - \mathcal{L}(\boldsymbol{q}, \dot{\boldsymbol{q}}, t).$$

This coincides with the Hamiltonian being the Legendre transformation of the Lagrangian, given that the Lagrangian is a convex function (in $\dot{\boldsymbol{q}}$). Roughly speaking, the Legendre transformation of a function $f(\dot{\boldsymbol{q}})$ finds $\dot{\boldsymbol{q}}(\boldsymbol{p})$ such that $g(\boldsymbol{p}) = \boldsymbol{p}\dot{\boldsymbol{q}} - f(\dot{\boldsymbol{q}})$ is maximal. Clearly, $\boldsymbol{p}$ must be chosen such that $\boldsymbol{p} = \frac{\partial f}{\partial \dot{\boldsymbol{q}}}$. This is precisely how the generalized momentum is defined in terms of the Lagrangian, which automatically implies that the Hamiltonian is indeed the Legendre transformation of the Lagrangian. The convexity of the function guarantees uniqueness and maximality.

The Legendre transformation is involutive; that is, $\mathcal{H}$ and $\mathcal{L}$ are the Legendre transformation of each other; this is why they are called *dual in the sense of Young* [34]. For the scalar case, a graphical explanation of the Legendre transformation exists is provided by fig. 5-5. Here, the Legendre transform is $g(p) = \max_x\{px - f(x)\}$; this optimality condition is met when $p = \frac{\mathrm{d}f}{\mathrm{d}x}$, i.e. where px and $f(x)$ are locally parallel.

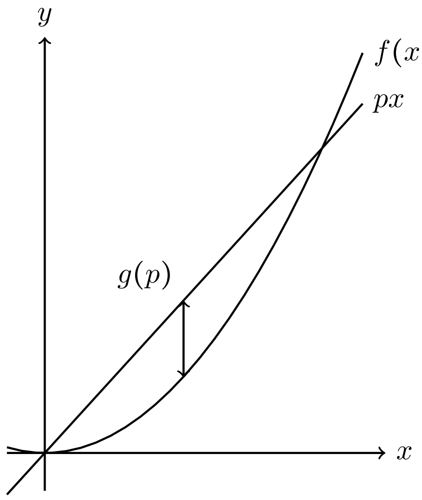


Figure 5-5: Illustration of the Legendre transformation. g is the Legendre transformation of the function f, as it maximizes the $px - f(x)$; this optimality is attained when $p = \frac{\mathrm{d}f}{\mathrm{d}x}$. In classical mechanics, this is extended to the multivariable case where $\mathcal{L}$, $\mathcal{H}$ and $\dot{\boldsymbol{q}}$ take the role of f, g and x respectively.

The Legendre transformation in economics

In economics, the Lagrangian measures the exchange between market surplus and convenience yield. This exchange has the units of $\left[\frac{\$}{\mathrm{yr}}\right]$, and is in economic engineering sometimes named 'cost' for short, although a definition in terms of utility would probably be more appropriate. The term $\boldsymbol{p}\dot{\boldsymbol{q}}$ has a very clear interpretation in economics: it is the *revenue* associated with the flow of goods and their prices. The Legendre transformation then maximizes the difference between revenue and the Lagrangian 'cost' with respect to the number of goods exchanged; $\dot{\boldsymbol{q}}$. The optimality is guaranteed by the price representing the change of utility with respect to a change in the flow of goods.
Because the Hamiltonian is equal to a difference between revenue and costs, it is said to be analogous to earnings.

5-3-2 Hamilton's equations

Equation (5-6) provides an expression for the total differential of the Hamiltonian, $\mathrm{d}\mathcal{H}$. By definition, this differential is equal to:

$$\mathrm{d}\mathcal{H}(\boldsymbol{q}, \boldsymbol{p}, t) = \frac{\partial \mathcal{H}}{\partial \boldsymbol{q}} \mathrm{d}\boldsymbol{q} + \frac{\partial \mathcal{H}}{\partial \boldsymbol{p}} \mathrm{d}\boldsymbol{p} + \frac{\partial \mathcal{H}}{\partial t} \mathrm{d}t \,.$$

Then, eq. (5-6) gives rise to the following equalities:

$$\dot{\boldsymbol{q}} = \frac{\partial \mathcal{H}}{\partial \boldsymbol{p}} \qquad \dot{\boldsymbol{p}} = -\frac{\partial \mathcal{H}}{\partial \boldsymbol{q}} \qquad \frac{\partial \mathcal{H}}{\partial t} = -\frac{\partial \mathcal{L}}{\partial t}, \tag{5-7}$$

which are known as *Hamilton's equations.* The Hamiltonian combined with Hamilton's equations yields an equivalent representation of the system dynamics as the Lagrangian with the Euler-Lagrange equations.

In mechanics, the Hamiltonian corresponding to $\mathcal{L} = T^* - U$, where T^* is the quadratic form given in eq. (5-3), the Hamiltonian is equal to:

$$\mathcal{H} = T + U, \tag{5-8}$$

which means that $\mathcal{H}$ is equal to the *total mechanical energy in the system.* A vital property of the Hamiltonian is that of its time (in)dependence, namely:

$$\frac{\mathrm{d}\mathcal{H}}{\mathrm{d}t} = \underbrace{\frac{\partial \mathcal{H}}{\partial \boldsymbol{q}}\frac{\mathrm{d}\boldsymbol{q}}{\mathrm{d}t} + \frac{\partial \mathcal{H}}{\partial \boldsymbol{p}}\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}t}}_{=0} + \frac{\partial \mathcal{H}}{\partial t} \implies \frac{\mathrm{d}\mathcal{H}}{\mathrm{d}t} = \frac{\partial \mathcal{H}}{\partial t},$$

that is, the Hamiltonian is time-invariant unless it features an explicit time-dependence. The vanishing of the first two terms is a direct consequence of Hamilton's equations. This property leads directly to the conservation of (mechanical) energy in time-invariant systems.

Whereas Lagrangian mechanics acts on the tangent bundle of the configuration manifold, Hamiltonian mechanics takes place in the *cotangent bundle* T^*M, for it is expressed in terms of the momenta rather than the generalized velocities. The generalized momenta live in the cotangent space at a certaint point of the configuration manifold because they are the differential of a scalar function (the Lagrangian) on the manifold. The cotangent bundle is a vector bundle and thus a manifold in its own right; here the base manifold of the cotangent bundle is the configuration manifold M, the fibers are the cotangent space for each point. A natural 1-form can be defined on this cotangent bundle that exploits the distinction between the base manifold and the fibers; namely a 1-form on the co-tangent bundle that is 'equivalent' to the covector in the cotangent space (with respect to their action on the tangent vector to the cotangent bundle and configuration manifold respectively). The action of this natural 1-form, referred to as the *tautological 1-form*, on a tangent vector to the cotangent bundle is equal to the action of the covector on this tangent vector *under the pushforward of the projection map.* That is, if π: $T^*M \to M$ is the projection map, $\boldsymbol{\xi} \in T(T^*M)$ is a tangent vector to the cotangent bundle, and m is an element of the cotangent bundle, the tautological 1-form ω^1 is defined by [34, 41]

$$\omega^1 \triangleq m \circ \pi_* \quad \text{or} \quad \omega^1(\boldsymbol{\xi}) = m(\pi_*(\boldsymbol{\xi})).$$

In classical mechanics, ω^1 is also named the *Poincaré 1-form* or the *canonical 1-form.* The tautological 1-form can be expressed in canonical coordinates as follows:

$$\omega^1 = \boldsymbol{p}\,\mathrm{d}\boldsymbol{q}\,.$$

The tautological 1-form is therefore a mathematical object that associates each component of the generalized momentum with the generalized position, and thus provides a fundamental connection between Hamiltonian and Lagrangian mechanics.

The exterior derivative of the Poincaré 1-form is a closed 2-form that is called the *symplectic form* ω^2:

$$\omega^2 = \mathrm{d}\omega^1\,.$$

In the given canonical coordinates $\boldsymbol{q}$ and $\boldsymbol{p}$, the symplectic form is

$$\omega^2 = \mathrm{d}\boldsymbol{p} \wedge \mathrm{d}\boldsymbol{q}\,;$$

which makes the cotangent bundle T^*M a symplectic manifold. The symplectic 2-form plays a similar role in symplectic geometry as the Riemannian metric does in Riemannian geometry; for example, by providing an isomorphism between the tangent and cotangent vectors. This isomorphism can be used to define a *Hamiltonian vector field* [34].

Hamiltonian vector field

Let $\mathcal{H}$ be a function on a $2n$-dimensional symplectic manifold (Q, ω^2). The symplectic 2-form defined on Q provides an isomorphism I between its tangent and cotangent vectors, such that $\omega^2(\boldsymbol{\eta}, \boldsymbol{\xi}) = \omega^1_{\boldsymbol{\xi}}(\boldsymbol{\eta})$. The vector field $I\,\mathrm{d}\mathcal{H}$ is called a Hamiltonian vector field, with $\mathcal{H}$ being the Hamiltonian function.

A crucial property of the Hamiltonian vector field (and the corresponding Hamiltonian flow) is that it preserves the symplectic structure: [34]

$$(g^t)^*\omega^2 = \omega^2,$$

where $(g^t)^*$ represents the pullback of the one-parameter group of diffeomorphisms of Q that are governed by the Hamiltonian flow. Hence, the symplectic form is an integral invariant of the phase flow of the Hamiltonian system.

The symplectic nature of economics

The preceding discussion presents the existence of the symplectic 2-form as a consequence of the phase space being the cotangent bundle of the configuration manifold. In economics, the presence of a symplectic structure is even more intuitive — one could even argue that economics is evidently symplectic in the first place, and that the connection with classical mechanics is a consequence of this structure. If one asserts that the economy as a system 'takes place' in the necessarily even-dimensional continuous space of products and their prices, then there is only one, natural structure in economics that ties them together: the concept of money. More specifically, multiplying a number of products with their respective price allows one to 'compare' economic goods: this is precisely the role of money as the *unit of account*, as stated by Mankiw and Taylor [36]. Said otherwise, the area of the phase plane gives the economic space its defining structure, which in turn gives rise to the canonical 2-form $\mathrm{d}\boldsymbol{p} \wedge \mathrm{d}\boldsymbol{q}$. The area is signed because money is also 'signed' from the perspective of an economic agent.

The canonical 2-form is only symplectic if it obeys two additional requirements: it must be closed and nondegenerate. Both requirements translate easily to economics:

- *Closed* means that the exterior derivative of the canonical 2-form is equal to zero. Crudely speaking, one could say that the exterior derivative is the 'divergence' of the economic phase flow. If this divergence is zero, than the economic phase flow is incompressible with money representing the volume. Intuitively, this means that the 'value' of money is consistent throughout the phase flow. This makes sense from an economic perspective, for money wouldn't be very good as a unit of account if its value were inconsistent. Of course, some objections can be made, such as the existence of different currencies and inflation, but it is argued that these are not fundamental enough to violate this elementary principle.

- *Nondegenerate* means that there is no product for which any price is necessarily 0, i.e. nothing is inherently 'free' in economics. Again, this is a reasonable (and even required) assumption for an economic system.

Hence, the economic space of prices and products naturally exhibits a symplectic structure. This can be further supported by the fact that another type of structure for this space, say a Riemannian structure (e.g. something like $\mathrm{d}p^2 + \mathrm{d}q^2$), would make much less sense as it does not capture the role of money as a means to compare and exchange goods of different nature.

The reason why one can make the additional connection with classical mechanics is that the price-product space is a cotangent bundle. This is in turn motivated by the additional observation that the prices are the partial derivative of a product's utility with respect to its transaction rate, i.e. $\boldsymbol{p} = \frac{\partial \mathcal{L}}{\partial \dot{\boldsymbol{q}}}$, i.e. the prices live in the fibers attached to the configuration manifold. Finally, there are the *additional* assumptions about the form of the Lagrangian and Hamiltonian; i.e. that they have the same form as those in classical mechanics ($\mathcal{L} = T^* - U$, $\mathcal{H} = T + U$). Although the previous sections support that this is indeed a sensible thing to do, they are essentially additional assumptions on top of the symplectic structure of the economic phase space, and the nature of prices.

The canonical formalism

One of the great advantages of Hamiltonian mechanics is that it admits a much wider range of coordinate transformations. The Lagrangian formalism already allowed point-transformations, which simply apply to the configuration manifold (i.e. a change in generalized coordinates), where the conjugate momenta transform according to the Legendre transformation. However, in the phase space the transformations may apply both to the momenta as the positions. Due to this large variety of transformations, the coordinates for $\boldsymbol{q}$ and $\boldsymbol{p}$ may no longer refer exclusively to a 'spatial' component and a 'momentum' component, and this distinction will then just be a matter of definition [35]. The 'allowed' class of coordinate transformations, called the *canonical transformations*, preserve the symplectic form ω^2 and thus the form of Hamilton's equations [34]; they are *symplectomorphisms* of the phase space. There is a rich variety in canonical transformations that can be helpful to better understand or integrate Hamiltonian systems; but this text will focus on one particular type of coordinates that has been studied in economic engineering in the past: the action-angle coordinates.

Action-angle coordinates A special choice of canonical coordinates arises for conservative systems, called the *action-angle coordinates*. These are pairs of coordinates of the form $(\boldsymbol{I}, \boldsymbol{\phi})$, where $\boldsymbol{I}$ consists of the action coordinates, and $\boldsymbol{\phi}$ are the angle coordinates. The reason why this set of coordinates exists in the first place is found in Liouville's theorem, which states that when a system with n degrees of freedom has n independent first integrals $\boldsymbol{F}$, then the level set of these first integrals is diffeomorphic to the the n-dimensional torus $\mathbb{T}^n$. This torus is completely parameterized by the n action variables, while the angle coordinates determine the position on the torus. Figure 5-6 visualizes the action-angle coordinates for the simple case of a system with one degree of freedom.

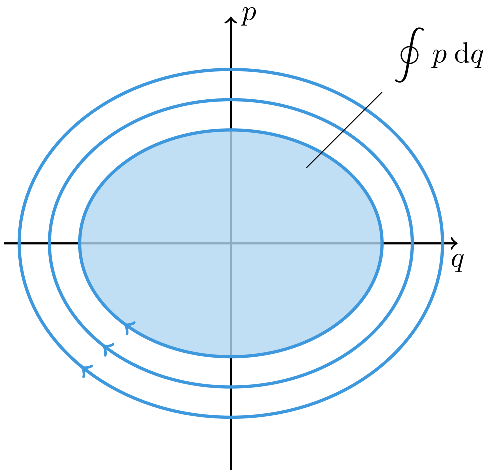


Figure 5-6: Illustration of the action-angle coordinates for a simple system where the phase space is $\mathbb{R}^2$. In that case, the trajectories are diffeomorphic to $\mathbb{T}^1$, or a circle, for a given energy level. The area enclosed by the trajectories is the area enclosed by the trajectory divided by 2π; the angle coordinate determines the position in this trajectory.

The action coordinates are functions of these first integrals so as to ensure that the action-angle coordinate set is symplectic, i.e. [34]

$$\omega^2 = \sum \mathrm{d}I_i \wedge \mathrm{d}\phi_i \,.$$

The major advantage of this particular choice of coordinates is that the corresponding differential equations are very easy to integrate. The action coordinates are constant over time, while the angle coordinates are characterized by a constant 'angular velocity' — another consequence of Liouville's theorem is that these systems are necessarily periodic. The angular velocity vector $\boldsymbol{\omega}$ depends on the value of the first integrals $\boldsymbol{f}$. That is, [34]

$$\frac{\mathrm{d}\boldsymbol{I}}{\mathrm{d}t} = 0 \qquad \frac{\mathrm{d}\boldsymbol{\phi}}{\mathrm{d}t} = \boldsymbol{\omega}(\boldsymbol{f}) \implies \boldsymbol{\phi}(t) = \boldsymbol{\phi}(0) + \boldsymbol{\omega} t.$$

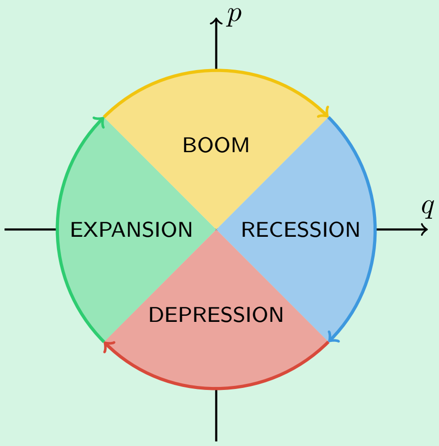
In the case where the cotangent bundle is $\mathbb{R}^{2n}$ (which is often the case for the linearized systems in economic engineering), the angle coordinates may constructed as follows:

$$I_i(\boldsymbol{f}) = \frac{1}{2\pi} \oint_{\gamma_i} \boldsymbol{p} \,\mathrm{d}\boldsymbol{q} \,, \tag{5-9}$$

where $\gamma_1, \ldots, \gamma_n$ form a *basis* for the one-dimensional cycles on the torus (that is, they each define a cycle in terms of one of the angle coordinates). This coordinate transformation can be shown to preserve the symplectic form.

Action-angle coordinates and the business cycle

Action-angle coordinates have been applied in past economic engineering research, most notably in the works of Van Ardenne [32] and Vos [42]. Based on dimensional analysis, they have been used to model financial markets; however, it is the opinion of the author that this application is a bit *ad hoc* and does not entirely respect the precise formal meaning of action-angle coordinates as outlined in this section. Rather, these coordinates can be used to express an economic system in terms of the 'magnitude' of the business cycle, where the 'angle' measures the phase of the business cycle (in Dutch, one would call this '*conjunctuur*'). Equation (5-9) essentially measures a total revenue over an entire business cycle, which has a characteristic frequency depending on the system parameters.



The above figure gives a typical business cycle with the associated terminology for the four 'phases'; it can be interpreted directly from the perspective of the phase space.

5-3-3 Noether's theorem

Conservation principles play a vital role in physics and mechanics. Notable examples are the conservation of energy, linear momentum, angular momentum and charge. A surprising result in physics is that all of the mentioned conservation laws are a consequence of the same theorem, attributed to Emmy Noether. This theorem essentially looks at the nature of the Lagrangian of the system, and sees whether it is 'ignorant' with respect to a certain group of transformations.

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: $F : TM \to \mathbb{R}$. In local coordinates q on M, the integral F is written in the form: [34]

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

The word 'admits' here means that the Lagrangian is invariant with respect to the transformation h; that is, the $\mathcal{L}$ exhibits a *symmetry* in that regard. A perspicuous consequence of this theorem is that when the Lagrangian has a cyclic coordinate (that is, it does not depend on some q_i), then the conjugate momentum of that coordinate is conserved, for the conjugate momentum of a coordinate q_i is $\dfrac{\partial \mathcal{L}}{\partial \dot{q}_i}$.

Energy conservation When the Lagrangian is time-invariant, one can define $h^s : (\boldsymbol{q}, t) \mapsto (\boldsymbol{q}, t + s)$ as translation in time. To find the conserved quantity, the Lagrangian has to be extended to consider the time t as an explicit generalized coordinate, i.e. define

$$\tilde{\mathcal{L}}\left(\boldsymbol{q}, t, \frac{\mathrm{d}\boldsymbol{q}}{\mathrm{d}\tau}, \frac{\mathrm{d}t}{\mathrm{d}\tau}\right) \triangleq \mathcal{L}\left(\boldsymbol{q}, \frac{\mathrm{d}\boldsymbol{q}/\mathrm{d}\tau}{\mathrm{d}t/\mathrm{d}\tau}, t\right)\frac{\mathrm{d}t}{\mathrm{d}\tau},$$

where τ now plays the role of the independent variable. This definition ensures that the action integrals for $\tilde{\mathcal{L}}$ and $\mathcal{L}$ are identical, which is why Noether's theorem can be applied to $\tilde{\mathcal{L}}$ instead. Of course, $\frac{\mathrm{d}h}{\mathrm{d}s}$ is equal to 1 for the time coordinate; so there is one first integral:

$$\begin{aligned} F = \frac{\partial \tilde{\mathcal{L}}}{\partial (t')} &= \mathcal{L} + \frac{\partial}{\partial t'}\left(\mathcal{L}\left(\boldsymbol{q}, \frac{\boldsymbol{q}'}{t'}, t\right)\right)t' \\ &= \mathcal{L} + \frac{\partial \mathcal{L}}{\partial \dot{\boldsymbol{q}}}\frac{\partial}{\partial t'}\left(\frac{\boldsymbol{q}'}{t'}\right)t' = \mathcal{L} - \frac{\partial \mathcal{L}}{\partial \dot{\boldsymbol{q}}}\frac{\boldsymbol{q}'}{t'} \\ &= \mathcal{L} - \frac{\partial \mathcal{L}}{\partial \dot{\boldsymbol{q}}}\dot{\boldsymbol{q}} \end{aligned}$$

with $t' = \frac{\mathrm{d}t}{\mathrm{d}\tau}$ and $\boldsymbol{q}' = \frac{\mathrm{d}\boldsymbol{q}}{\mathrm{d}\tau}$. This is the negative of the Legendre transformation of the Lagrangian, or the negative of the Hamiltonian; $F = -\mathcal{H}$. Therefore, the total energy of the system is conserved.

(Angular) momentum conservation The invariance of the Lagrangian in one or more coordinate directions implies the conservation of the associated conjugate momentum.

For example, a common assumption on Earth is that the gravitational force is constant and acts downwards. Therefore, only one coordinate appears in the potential energy expression; the other coordinate directions (that is, the two 'horizontal' directions, with proper choice of the coordinate system) are cyclic. Hence, by virtue of Noether's theorem, the linear momentum in those directions is conserved, for it is unaffected by any potential force.

Conservation of angular momentum often appears in the presence of so-called *central* force fields, which means that the potential energy is a function of the *distance* from a certain point: notable examples are Coulomb forces or gravitational forces in celestial mechanics. Because only distance matters, the orientation of the displacement vector between the origin of the force field and the body is of no importance; it will not appear in the Lagrangian either. Consequently, Noether's theorem dictates that these central force fields preserve angular momentum.

Noether's theorem and conservation in economics

Noether's theorem can be applied to economic systems as well. The most simple application is the conservation of prices in the absence of scarcity. When the utility Lagrangian does not depend on a certain stock level q_i, that essentially means that there is no utility associated with the possession (or shortage) of it; in other words, economic agents are not driven by scarcity (or abundance) of that particular product. The 'conjugate momentum' of an economic good is its price; so *in the absence of scarcity, price is conserved.*

A similar comment about cyclic coordinates can be made in the rotational analogy (which is to be discussed in chapter 6). If the utility Lagrangian is independent on the return or interest ζ, then borrowers or lenders will be indifferent about the realized returns of their investments, and never capitalize them (or in the case of borrowers, never repay their debts).

Finally, the conservation of income in the economic system is related to the time-invariance of its Lagrangian. The Lagrangian does not depend on time if (i) there are no external influences or forcing terms and (ii) all parameters in the system remain constant over time. If that is the case, the Hamiltonian (or total income) in the economic system is conserved; or volume in the phase space (i.e. total amount of money) is conserved.

Admittedly, all of these interpretations are arguably rather straightforward, since they mostly rely on either the presence of cyclic coordinates or the direct correspondence of total energy and energy of the system. However, the discussion does prove that Noether's theorem has very tangible interpretations in economics. Its real power will, though, appear with more sophisticated symmetries that are present in complex economic systems. The idea is that the Lagrangian and Noether's theorem permit the extraction of 'stylized facts' from these types of systems that would otherwise be concealed by their complexity.

Chapter summary

The basis for the Lagrangian formalism is the configuration manifold, which contains all the possible configurations of the system. The tangent bundle of the configuration manifold then comprises both the generalized positions and the associated velocities. The Lagrangian is a scalar-valued function on the tangent bundle; in classical mechanics, it amounts to the difference between the kinetic and potential energy of the system. The principle of stationary action states that motions of a Lagrangian system are trajectories for which the integral of the Lagrangian over time, the action integral, is stationary. This gives rise to an alternative formulation of Newtonian mechanics in the form of the Euler-Lagrange equations.

From the Lagrangian, the Hamiltonian is obtained through the Legendre transformation, which interchanges the generalized velocities for generalized momenta. Therefore, the Hamiltonian is defined on the cotangent bundle of the configuration manifold. The evolution of the system is given by a set of equations called Hamilton's equations. Hamiltonian systems admit a range of (canonical) coordinate transformations that preserve the symplectic structure of the phase space. For some systems, a particular coordinate transform can be used to arrive at the action-angle coordinates, which permit to integrate the system dynamics very easily.

In economic engineering, the Lagrangian and Hamiltonian are defined in terms of utility. The generalized positions are stock levels in the economic space. By virtue of the utility Lagrangian, prices are said to live in the fiber attached to every point in the economic configuration space. As such, the economic space of prices and quantities is analogous to the cotangent bundle (or phase space) of classical mechanics, where the Hamiltonian corresponds to the total earnings in the system. This is motivated by the natural symplectic structure of the economic phase space.

Chapter 6

Investment as a hyperbolic rotation

This chapter explains the connection between compound interest (and investments in general) and the hyperbolic rotations. To do so, section 6-1 first explains in depth the underlying mechanism behind compound interest: the reinvestment of earnings. Secondly, it is shown in section 6-2 that this mechanism represents a hyperbolic rotation in the plane. In section 6-3 hyperbolic angles are formally defined, analogous to 'regular' circular angles. Subsequently, the rotational analogy from economic engineering is treated in section 6-4. Finally, an alternative representation of hyperbolic rotations is found in a special number system called the hyperbolic numbers; this is the subject of section 6-5.

6-1 Interest terminology

The aim of this research is to refine the way that economic engineering currently deals with financial instruments, primarily fixed-interest assets, but also marketable equity (stocks), derivatives, etc. All of these can be regarded as 'investments', for lack of a better word, because they require a *principal*, that is the amount invested, and a certain return that is proportional to the principal. This is why the return is usually expressed as a percentage of the principal, which in the case of debt is the interest rate. The latter contains the word 'rate' because it has an inherent temporal aspect: interest rates (and returns of investments in general) are associated with a certain 'term' or duration over which the return is realized. As such, any type of investment is characterized by three components:

- a *principal*, something that is originally invested;
- a *return* proportional to the principal;
- a *period* over which the return is realized.

In the world of finance, the principal and return often refer to some amount of money, but that does not necessarily have to be the case. For example, some stocks pay *stock dividends*, which means that dividends take the form of additional company stock. Alternatively, one could look to other factors of production as well, such as land: a farmer may 'invest' a certain amount of land to gain the earnings (crops that grow on it) after the period of one year; of course, the amount of crops grown is proportional to the surface area of the land that the farmer has cultivated.

Reinvestment of earnings After the end of the investment period, the investor has two options: either he can spend his earnings, or he can reinvest them; essentially, he starts the next investment period with a higher principal than the first one. This reinvestment of capital is a crucial concept in the world of finance and is the reason for the existence of compound interest. Indeed, compound interest assumes that one's earnings after a 'compounding' period also bear interest; the so-called *interest on interest*. This can be illustrated by means of the example shown in fig. 6-1. An investor holds a fixed-interest investment with a principal of \$1000 and an interest rate of 5% annually. After one compounding period, he naturally earns \$50 directly on the principal. As such, his direct earnings, which will be called 'yield', now amount to \$50 while the invested 'capital' is still \$1000. The next compounding period, however, provided that the investor does not withdraw his earnings from the account, the \$50 generates \$2.5 of interest as well, which are *reinvested*; i.e. they take the form of capital instead of yield. Simultaneously, the investor also keeps the additional benefits from the \$1000 of capital in the form of another \$50 dollars of direct earnings.

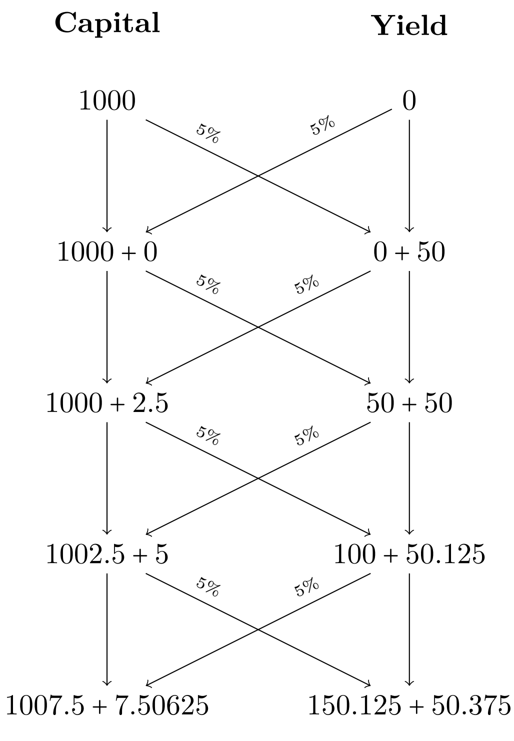


Figure 6-1: A simple example of compound interest with a principal of \$1000 and an interest of 5%. The direct earnings on the capital are called 'yield', while the (re)invested amount is called 'capital'.

This reinvestment of earnings is the driving force behind compound interest: without this step, the investor would end up with *simple interest*, which sees its relative returns essentially decreasing over time. One should, however, be cautious to associate this process exclusively with financial investments. For example, a farmer could use the earnings from his land to buy more land, increase his earnings, use those again to buy even more land, and so forth — this will again give rise to exponential growth. The farming example makes it quite clear that the object that 'generates' the return (capital) is of a very different nature than the return itself (yield). They can be connected through the usage of money as both a unit of account and a medium of exchange, which is essential for the whole process to work [36].

Financial investments are arguably the culmination of this concept, since money itself earns money that can be reinvested again instantaneously. Still, the capital-yield distinction remains relevant: for example, companies pay dividends to their shareholders in proportion to the amount of stock they own; the shareholders could then choose to reinvest these earnings in more stock of that (or another) company, increasing their 'capital'.

6-2 The hyperbolic shape of compound interest

If the values from the example in fig. 6-1 are plotted against each other, with capital on the horizontal axis and yield on the vertical axis, a hyperbolic shape appears, shown in fig. 6-2. Hyperbolae are characterized by the implicit equation

$$x^2 - y^2 = K^2 \tag{6-1}$$

where K represents the 'radius' of the hyperbola, which coincides with the principal amount of the investment. The amount of capital and yield accumulated at are on the x-axis and y-axis respectively. However, as fig. 6-2 demonstrates, the shape arising from the example is not quite hyperbolic, since it essentially lags behind the actual hyperbola of that radius. The

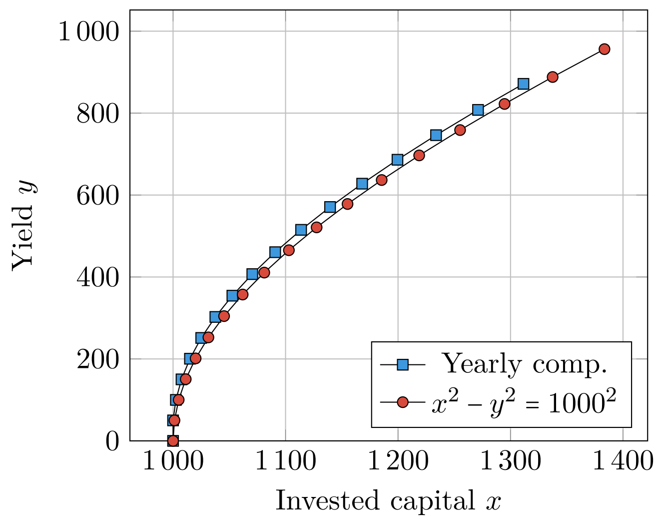


Figure 6-2: Plot of the numbers from the example in fig. 6-1. The 'ideal' hyperbola with the same radius, given by the equation $x^2 - y^2 = K^2$ is shown as well.

reason why the 'perfect' hyperbola is not recovered in the example is demonstrated as follows.

The capital-yield evolution can be modeled as a second-dimensional difference equation like so:

$$\begin{pmatrix} x(k+1) \\ y(k+1) \end{pmatrix} = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix} \begin{pmatrix} x(k) \\ y(k) \end{pmatrix} \quad \text{with} \quad x(0) = K,\, y(0) = 0,$$

where x denotes capital, y denotes yield and r is the interest rate. The solution of this autonomous system is given by:

$$\begin{pmatrix} x(k) \\ y(k) \end{pmatrix} = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}^k \begin{pmatrix} K \\ 0 \end{pmatrix}.$$

One can then write $x(k)^2 - y(k)^2$ in terms of this solution:

$$x(k)^2 - y(k)^2 = \begin{pmatrix} K & 0 \end{pmatrix} \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}^k \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}^k \begin{pmatrix} K \\ 0 \end{pmatrix}.$$

These matrix powers are easily evaluated using the eigenvalue decomposition:

$$\begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}^k = V \Lambda^k V^{-1} \quad \text{with} \quad V = \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \quad \Lambda = \begin{pmatrix} 1-r & 0 \\ 0 & 1+r \end{pmatrix}.$$

Straightforward computations then yield:

$$x(k)^2 - y(k)^2 = K^2 \left(1 - r^2\right)^k,$$

which suggests that for reasonably small values of r, the compounding process approximates the equation for a hyperbola. One way to achieve this is to compound faster, essentially decomposing the compounding period into more periods with a smaller interest, such that $r \mapsto r/n$ and $k \mapsto kn$ with $n \to \infty$. Indeed,

$$\lim_{n \to \infty} K^2 \left(1 - \left(\frac{r}{n}\right)^2\right)^{kn} = K^2.$$

That is, for infinitely fast compounding, the capital-yield decomposition of compound interest follows a hyperbolic shape.

Another way to view this is to see the difference equation as a sampled version of an underlying continuous system. One can find the continuous-time state-transition matrix by applying the matrix logarithm to its discrete-time counterpart:

$$\log \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \log(r+1) + \log(r-1) & \log(r+1) - \log(r-1) \\ \log(r+1) - \log(r-1) & \log(r+1) + \log(r-1) \end{pmatrix}.$$

This continuous-time state-transition matrix has eigenvalues $\log(1+r)$ and $\log(1-r)$, the former of which is the continuous-time equivalent interest rate, or the *force of interest*, $\dot{\zeta}$ [43]. From this discussion, one can infer that the discrete compounding process is essentially an approximation of the underlying 'ideal' continuous-time process, which is why the rotational analogy will be developed for a continuous compounding scenario; since the discrete-time equivalent may always be obtained by means of sampling.

6-3 Conic sections and hyperbolic rotations

The implicit equation for a hyperbola eq. (6-1) is remarkably similar to the equation of a circle, the only difference being the minus sign. Indeed, circles and hyperbola are both members of a larger family called conic sections, which play a prominent role in geometry. Conic sections are usually classified based on their eccentricity e, for which three general cases can be distinguished:

- $0 \leq e < 1$ are *ellipses*, were the particular case for which $e = 0$ corresponds to a circle;
- $e = 1$ are for *parabolae*;
- $e > 1$ are hyperbolae.

The elliptic-parabolic-hyperbolic distinction is a common theme within mathematics and in this literature study, as it also applies to surface curvature, which gives rise to elliptic, parabolic, or hyperbolic geometry (the subject of chapter 3), and the classification of Möbius transformations discussed in chapter 4.

Harkin and Harkin [44] use the conic sections to extend the 'classic' trigonometry based on circles (i.e. elliptic curves) to hyperbolic and parabolic trigonometry as well. Each of these types of trigonometry comes with a special number system akin to complex numbers (for the elliptic case), called double numbers (or hyperbolic numbers) and dual numbers (or parabolic numbers) — hyperbolic numbers will briefly be explored in section 6-5. The results of the previous section suggest that hyperbolic trigonometry can be used to describe investment problems: this is the basis for the *rotational analogy* in economic engineering, which is introduced in section 6-4.

6-3-1 Hyperbolic angles

For both circles and hyperbolae, an angle refers to a certain region bounded by the curve at issue. The standard notation for the hyperbolic angle will be ζ, in accordance with the rapidity from special relativity, which can also be viewed as a hyperbolic angle (cf. chapter 2).

Hyperbolic sector

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

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 $(K, 0)$.

Figure 6-3 visualizes the concept of a hyperbolic angle: the angle ζ is defined in terms of the unit hyperbola $x^2 - y^2 = 1$, or the hyperbola with 'radius' 1. By means of a special polar coordinate set, points can be expressed in terms of a radius K and an angle ζ. If the radius K is allowed to be negative, all the points in the disconnected open set bounded by $y = x$ and $y = -x$ can be uniquely identified with a radius K and hyperbolic angle ζ. This is similar to regular polar coordinates in terms of circular angles, which is why these coordinates will be referred to as 'hyperbolic polar coordinates'. These coordinates are interesting because they allow to express investments in terms of their principal and realized return in the capital-yield plane.

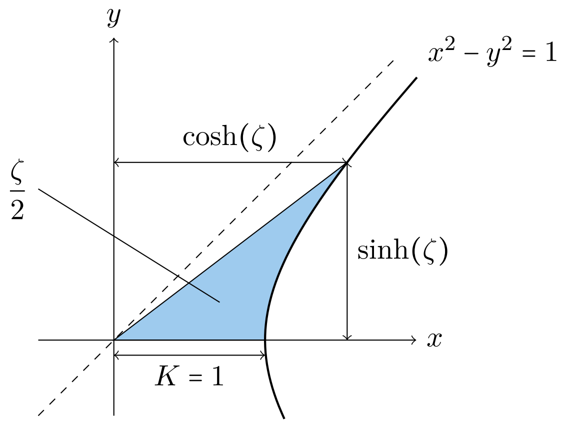


Figure 6-3: Illustration of a hyperbolic angle defined on the unit hyperbola, with projection on the axes using the hyperbolic sine and cosine.

The preceding discussion suggests 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 co-domain of the mapping to the set $\left\{(x, y) \in \mathbb{R}^2 : |x| > |y|\right\}$. The action of the mapping is illustrated by fig. 6-4. From a financial perspective, this limit signifies the maximal yield one can obtain for a given amount of capital, i.e. one can never obtain more yield than (re)invested capital at a given moment in time.

Remark A possible workaround for this problem from a mathematical perspective can be found in the so-called *generalized trigonometry* as described by Harkin and Harkin [44]: in this case, the notion of the hyperbolic angle itself is considered ambiguous and must be accompanied by 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 the other hyperbolic branches $y^2 - x^2 = K^2$, which were disregarded up till now.[1]

$$\zeta = \begin{cases} \tanh^{-1}(x/y) & \text{for branches I and III} \\ \tanh^{-1}(y/x) = \coth^{-1}(x/y) & \text{for branches II and IV} \end{cases} \tag{6-2}$$

[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.

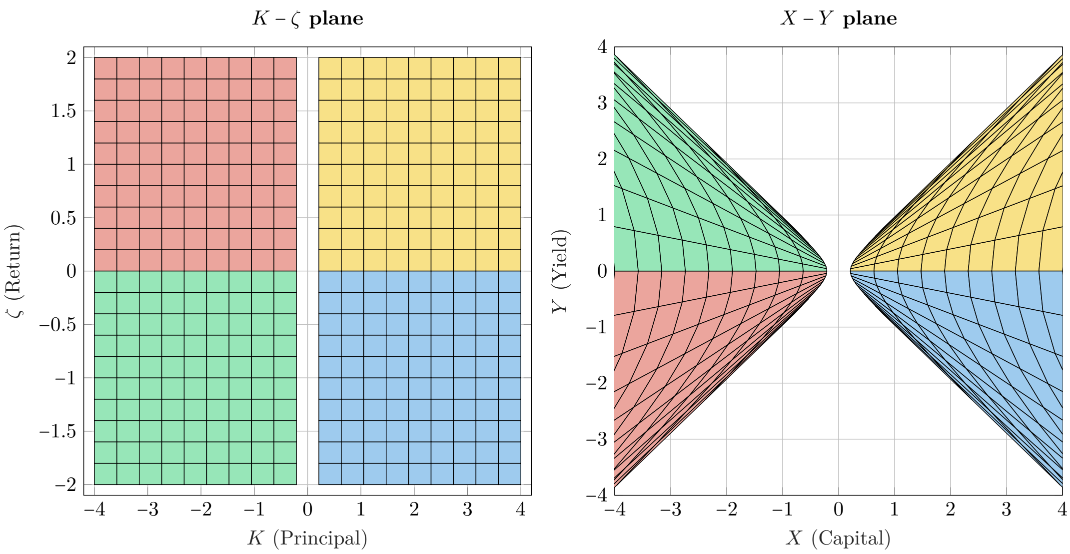


Figure 6-4: Visualization of the usage of hyperbolic polar coordinates as a mapping form the $K - \zeta$ (principal-return) space and the $X - Y$ (capital-yield) plane.

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.

6-3-2 Hyperbolic functions

Figure 6-3 also shows that any point on a hyperbola may be projected on the capital and yield axis using a special set of functions, called *hyperbolic functions*. These functions are akin to the trigonometric functions sin and cos for 'circular' angles. Capital and yield can be expressed in terms of K and ζ using the hyperbolic analogs of the trigonometric functions, the so-called *hyperbolic functions* sinh and cosh:

$$x = K\cosh(\zeta) \qquad y = K\sinh(\zeta).$$

This decomposition of an investment in capital and yield in terms of these functions is almost trivially encapsulated in the mathematical identity:

$$\exp(\zeta) = \cosh(\zeta) + \sinh(\zeta). \tag{6-3}$$

This identity is easily verified by inspection of the Taylor series of the functions appearing in eq. (6-3):

$$\begin{aligned}\exp(x) &= \sum_{k=0}^{\infty}\frac{x^k}{k!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots \\ \cosh(x) &= \sum_{k=0}^{\infty}\frac{x^{2k}}{(2k)!} = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \ldots \\ \sinh(x) &= \sum_{k=0}^{\infty}\frac{x^{2k+1}}{(2k+1)!} = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \ldots.\end{aligned} \tag{6-4}$$

Or, alternatively, via the definition of the hyperbolic functions,

$$\begin{aligned}\cosh(x) &\triangleq \frac{\exp(x)+\exp(-x)}{2}\\ \sinh(x) &\triangleq \frac{\exp(x)-\exp(-x)}{2}.\end{aligned} \tag{6-5}$$

As such, the hyperbolic functions allow finding the associated capital and yield for a given accumulated interest ζ and initial investment K.

Clearly, the cosh and sinh functions are constructed by isolating respectively the even or odd powers from the Taylor expansion of the exponential. Equation (6-3) 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

$$\cos(\mathrm{i}x) = \cosh(x) \qquad \sin(\mathrm{i}x) = \mathrm{i}\sinh(x). \tag{6-6}$$

Hence, a *hyperbolic rotation may be expressed as a 'normal' circular rotation through the usage of complex angles*, which makes it possible to connect the hyperbolic rotations for investments to normal 'circular' rotations in mechanics, as is the purpose of economic engineering.

To make eq. (6-6) more visual, the hyperbolic and trigonometric functions can all be represented by looking at the modular surface of $|\sin(z)|$, shown in fig. 6-5: 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 π respectively [2].

6-3-3 Interest in the Lorentz-Minkowski plane

The growth of compound interest in terms of capital and yield can also be viewed as a Lorentz transformation in the Lorentz-Minkowski plane. These are actions of the Lorentz group $\mathrm{SO}(1,1)$ in the plane. Observe that in this case, the transformations need not be ortochronous (indicated by the absence of the '+' in the group name). That is because the sign of K may be switched as a result of the transform. In the financial application, this is equivalent to switching from debit to credit or vice versa.

For a constant interest rate $\dot{\zeta}$, the accumulated interest itself is $\dot{\zeta}t$. The evolution of capital and yield can are then described by a continuously varying Lorentz transformation in the plane:

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} \cosh(\dot{\zeta}t) & \sinh(\dot{\zeta}t) \\ \sinh(\dot{\zeta}t) & \cosh(\dot{\zeta}t) \end{pmatrix} \begin{pmatrix} K \\ 0 \end{pmatrix}.$$

In the analogy with special relativity, the 'temporal' dimension is here reserved for capital, while the yield is the spatial dimension.

6-4 The rotational analogy for economic engineering

It has now been established that investments can be represented using (hyperbolic) angles. Since economic engineering attempts to find analogies between traditional engineering disciplines and economics, the rotational nature of investments is a strong motivation to make

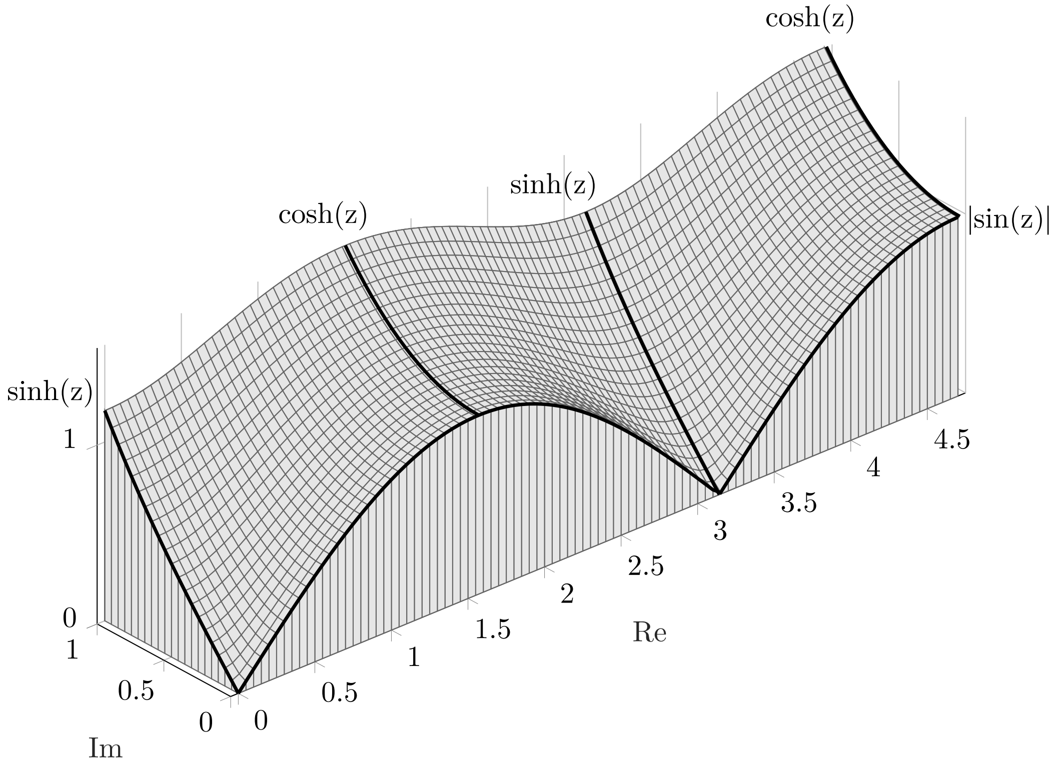


Figure 6-5: The modular surface of the sine over the complex plane, embedding all the trigonometric and hyperbolic functions at specific cross-sections.

the connection with *rotational mechanics* and the associated concepts of rotational kinetic energy, angular momentum, angular velocity, etc.

Some aspects of the analogy are already apparent from the foregoing discussion: the *angle* corresponds to the accumulated interest or return gained on the investment, while the *arm* or radius is analogous to the principal of the investment. Likewise, the time derivative of the angle $\dot{\zeta}$ is the angular velocity, i.e. the interest rate, which may itself be time-varying.

The current interpretation in economic engineering proclaims that credit corresponds to angular momentum. One reason for this is that the units of angular momentum are [N m s], in economic engineering, this would be units of currency [$]. Furthermore, the analogy can be supported by means of the Lagrangian as well, similar to the reasoning in chapter 5 connecting prices and momentum. If one assumes that the rotational kinetic energy is the 'market surplus' gained from borrowing or lending money (or investing in general) based on a certain return rate $\dot{\zeta}$, the conjugate momentum of this form of kinetic energy is necessarily the angular momentum. In the case of investments, the market surplus is associated with a new type of elasticity that relates the amount invested L with the return rate $\dot{\zeta}$, similar to the supply/demand curves for 'regular' goods and services. In economics, this is called the *curve of loanable funds.* Therefore, the 'borrower' or 'investor surplus' is well-defined: it is related to how much more the current return (or how much less the current interest) is than the minimal interest for which one is willing to invest or borrow. From this definition, the angular momentum must be analogous to the amount invested (or borrowed), being the partial derivative of this 'market surplus' with respect to the return.

Table 6-1: Overview of the analogies between rotational mechanics and economic engineering.

Rotational mechanics	**Economic engineering**
Arm	Principal
Angle	Return / accumulated interest
Angular velocity	Return rate / interest rate
Rotational kinetic energy	Borrower / lender surplus
Angular momentum	Credit

Issues with the rotational analogy Although the rotational analogy has been applied successfully in the past (see Kruimer [28] and Van Ardenne [32]), there appear to be some inconsistencies in its present interpretation.

- Currently, the difference between the arm (initial investment) and the angular momentum (credit) is not entirely clear.
- The sign of the interest rate does not determine the sign (i.e. direction) of the angular momentum; the former has to do with either discounting/appreciation while the latter denotes the direction of the contract (e.g. debit vs. credit).
- The interpretation of the elasticity as the mass moment of inertia (because it determines the relation between the amount of credit and the interest rate) has not been consistently applied. In the past, the mass moment of inertia has been put on equal footing with the

duration of a financial contract, but more recent works have pointed out that duration is a part of the valuation and is therefore associated with frequency domain analysis [4].

Besides, the rational analogy has been used in the context of bond graph modeling, where one simply needs two conjugate variables (interest and credit), and the corresponding C, R and I-elements. The underlying reasoning supported by the hyperbolic rotation that is compound interest was irrelevant for this application. In this context the rotational analogy is essentially, apart from terminology, completely equivalent to the action-angle approach used in other research.

6-5 Hyperbolic numbers

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

$$z = x + y\mathrm{j} \quad x, y \in \mathbb{R}.$$

Combined with addition and multiplication, the hyperbolic numbers form a commutative ring. Each hyperbolic number z is associated with its *hyperbolic conjugate* $\bar{z} = x - y\mathrm{j}$. The product of a hyperbolic number with its own hyperbolic conjugate produces a quadratic form $z\bar{z} = x^2 - y^2$, which always returns a real number; this quadratic form is equivalent to the Lorentzian metric. Again, three cases can be distinguished for z:

- z is *timelike* when $z\bar{z} > 0$;
- z is *lightlike* when $z\bar{z} = 0$;
- z is *spacelike* when $z\bar{z} < 0$.

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 $z\bar{z} = 0$ — this is precisely the case for all the hyperbolic numbers on the light cone in the hyperbolic plane. The *hyperbolic modulus* therefore requires an absolute value:

$$|z| \triangleq \sqrt{|z\bar{z}|},$$

which is considered to be the hyperbolic distance from z to the origin [46].

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

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

[2]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, the name '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 [45, 46, 47, 44].

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 ζ is associated with hyperbolic branches I and III. The polar form of the hyperbolic numbers reflects precisely the two different representations for investments: using principal and return (polar form), or capital and yield ('Cartesian' form). For the spacelike branches II and IV, the components are interchanged, such that $z = K\mathrm{j}\exp(\zeta\mathrm{j})$ 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 $z_2 = x_2 + y_2\mathrm{j}$:

$$\bar{z}_1 z_2 = \underbrace{(x_1x_2 - y_1y_2)}_{\text{inner product}} + \underbrace{(x_1y_2 - x_2y_1)}_{\text{outer product}}\mathrm{j}.$$

In 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. Therefore, the hyperbolic number plane and the complex plane have an identical notion of area. Based on the inner product, one can recognize a 'hyperbolic' notion of orthogonality, i.e. z_1 and z_2 are *hyperbolically orthogonal* if their hyperbolic inner product equals zero [2, 46].

The inner and outer product both have their financial interpretations: the inner product is related to the hyperbolic distance between two hyperbolic vectors, just like the Lorentz product in the previous section. The outer product is akin to the cross product, and is therefore used to rigorously define angular momentum in the Lorentzian space, which is an important component of the economic engineering analogy.

6-5-1 Matrix representation

For many algebraic structures (groups, rings, fields, ...) an isomorphism can be found in the realm 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

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

Likewise, one can establish an isomorphism between the hyperbolic numbers and a certain class of matrices. The hyperbolic number system is ring-isomorphic to the matrix ring

$$x + y\mathrm{j} \quad \leftrightsquigarrow \quad \begin{pmatrix} x & y \\ y & x \end{pmatrix} \quad x, y \in \mathbb{R},$$

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

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

6-5-2 The idempotent basis

The hyperbola can be viewed conveniently in two particular axis systems, the standard axis system (i.e. the capital-yield axes), and the *idempotent axis system*. The idempotent axis system is directed along the asympotes of the hyperbola, as illustrated in fig. 6-6.

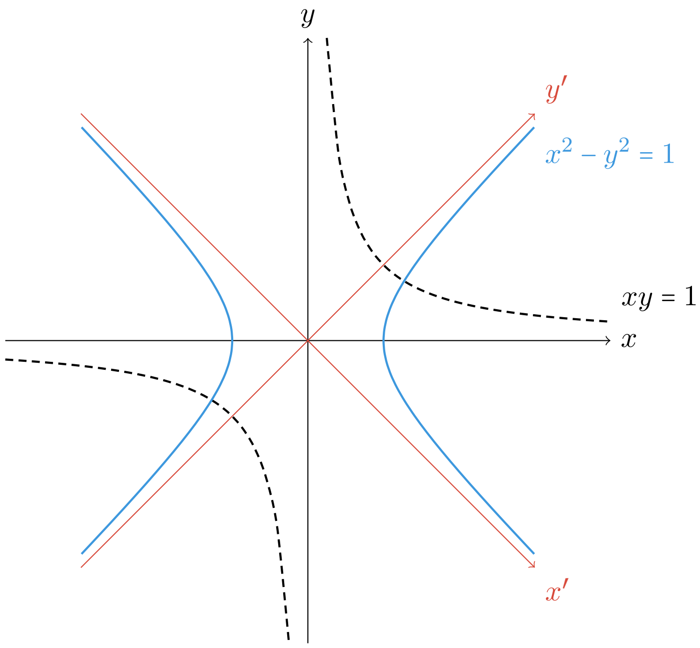


Figure 6-6: Comparison between the standard and idempotent axis system. The idempotent asix system is indicated in red; it is directed along the asymptotes of the standard hyperbola with equation $x^2 - y^2 = 1$.

This basis can also be defined in terms of hyperbolic numbers. Instead 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 it has a projective property: [46]

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

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

Chapter summary

An investment is characterized by a principal and a proportional return. The periodic reinvestment of earnings causes the total value of the investment to grow exponentially. This value can be decomposed in capital (what is invested) and yield (what is earned directly on the investement). For continuous compounding, the capital-yield decomposition follows a hyperbolic trajectory; it can be expressed in terms of a hyperbolic angle that is the accumulated return of the investment. The radius of the hyperbola represents the principal of the investment. The correspondence between rotations and compound interest is applied in economic engineering in the form of the rotational analogy: an analogy between rotational mechanical systems and financial instruments. Finally, the hyperbolic rotations can also conveniently be described in terms of the algebra of the hyperbolic number system.

Chapter 7

Research proposal

Mankiw and Taylor [36] define three roles of money in society:

(i) a *medium of exchange*; i.e. a universal means of exchanging goods and services without the necessity for bartering,

(ii) a *unit of account*; a measuring stick to compare different goods by means of their prices; and

(iii) a *store of value*, which allows translating present earnings to the future.

A sound framework in economic engineering should be able to assign a precise interpretation to each of these roles. The medium of exchange is encapsulated in the 'price' component of the price-quantity phase space that underlies the economic engineering theory. Likewise, the role of money as a unit of account is reflected in the symplectic structure of the economic phase space: as shown in section 5-3-2, the symplectic form (which represents a differential amount of revenue 'traced out' in the phase plane) provides an isomorphism between the covariant and contravariant tensor components. However, this does not cover other notions to which the unit of account may apply, such as the substitution effect; so this interpretation may have to be refined. Finally, the role as a store of value has been the most problematic in the past because of (i) its inherent temporal aspect and (ii) the idea that earnings (which are analogous to energy) are converted to either a generalized momentum or a generalized position. This 'store of value' problem is most apparent when one tries to model investments of any kind; like stocks, bonds, derivatives, debt, and equity, etc.

It has been proven in the past that financial systems (that is, economic systems that include the market dynamics of investments) are an essential component of the larger economic processes, as demonstrated by Kruimer [28], Vos [42] and Van Ardenne [32]. In past economic engineering research, two prevalent interpretations have been assigned to financial instruments: action-angle coordinates (cf. chapter 5) and the rotational analogy (cf. chapter 6). However, both problems present some crucial issues:

- The **action-angle approach** is based on a special canonical transformation of the phase space coordinates that are used in physics to facilitate the integration of complex systems or make statements about their periodicity. Their application on financial systems is based on the fact that they consist of dimensionless angle coordinates (‘return’) and action coordinates with units of money (‘investment’). However, this seems to be an ad hoc solution, since the formal applicability of action-angle coordinates is very precise; they can only be used for conservative and conditionally periodic systems [34]. For this reason, the angle coordinates have a constant ‘angular velocity’: clearly, this is not at all representative of the return rate in economics. Likewise, the action coordinates are supposed to be functions of first integrals of the system (and therefore time-invariant); again, this does seem unlikely for the investments in dynamic economic systems.

This approach may have worked in an applied modeling context by producing the correct dynamics, the direct correspondence with investments and return evidently does not hold up to more rigorous standards. This does not mean that there is no place for action-angle coordinates in economic engineering whatsoever, take for example the business cycle analogy that is proposed in section 5-3-2.

- The issues with the **rotational analogy** are more subtle, but they have far-reaching ramifications nevertheless. As section 6-4 already presented, there is definitely something to say for the ‘rotational’ part, i.e. viewing interest as a hyperbolic rotation — this is the entire point of chapter 6. Basically, one can argue that the kinematics hold, but the application of the dynamics shows some shortcomings: for example, the difference between the arm and the angular momentum, which both have identical units, the interpretation of rotational kinetic energy. Furthermore, duration has been proposed as being equivalent to the mass moment of inertia, but does this does not cover the full character of financial instruments; beside the fact that duration is now considered to be a valuation concept that belongs in frequency-domain analysis [4, 28]. In the past, the rotational analogy has been used for modeling in a similar fashion as the action-angle coordinates: simply by making the amount invested and the return conjugate variables in the same way as ‘traditional’ prices and quantities are related in economic engineering.

Given the importance of financial systems, a solution to this problem will benefit economic engineering research and its applicability to real situations. It can be observed that the quest for suitable analogies to money and financial instruments is hampered by the ‘unit problem’; that is, many quantities are expressed in terms of money while they play a very different role: for example, the distinction between money as energy (income) and as a quantity or momentum. Whereas dimensional analysis has been an instructive practice in other applications of economic engineering, employing it carelessly in this situation quickly leads to inconsistencies. For example, the rotational analogy requires the continuous transformation between direct earnings on interest to capital; from a fundamental standpoint, these may correspond to quantities of a different nature; other types of financial instruments (e.g. stocks) suggest that this process should be ‘pulled apart’, for it conceals the underlying mechanism of reinvestment.

As such, the aim of this research is to return to find the solution of this problem in the fundamentals of economic engineering as they are presented in chapter 5. These fundamentals consist of the concept of utility as energy and the symplectic structure of the economic phase plane. Because of the unit problem, it is often unclear in finance what has the role of products and prices (or even something else); this will naturally appeal to the usage of canonical coordinate transforms of the regular phase space or a Routhian approach where the Hamiltonian and Lagrangian approach are essentially merged. The idea is to find a mapping between these central concepts and the fundamental theory of interest and utility, in particular the work of Irving Fischer [48, 49].

Other questions that are to be answered pertain to financial equilibria and conservation principles. For example, the classic mechanical equilibrium (Galilean frame of reference) consists of a mass moving at a constant velocity; in economic engineering, this would be a demander with a constant consumer surplus or a producer with a constant producer surplus; consequently, market prices remain constant over time. This and other conservation laws are governed by Noether's theorem (cf. section 5-3-3). The crucial link between Noether's theorem and the structure behind the Lagrangian and Hamiltonian utility functions will provide insights that would otherwise be shrouded by the unit problem or the potentially complex differential equation that describes the evolution. Usually, the latter is used to understand the nature of the economic system, but the case is made here that analysis of the Lagrangian and Hamiltonian functions is usually more efficient and instructive.

Although the main goal of this research is to improve the understanding of financial systems, the rigorous methods that will be applied may contribute to economic engineering in general as well. For example, the application of manifold theory to economic systems, both from the perspective of the configuration manifold and the cotangent bundle or phase space. Until now, the natural structure of these spaces has not been researched in depth. Additionally, Hamiltonian and Lagrangian analysis are usually not deemed suitable for dissipative systems; but Hutters and Mendel [50] demonstrated how Hamiltonian systems can be applied to systems with linear damping. Because real economic systems are always dissipative (just like their mechanical counterparts), this method may be applied to financial systems as well.

Finally, the incorporation of the nature of capital in the economic engineering framework can be used to explain heavily debated concepts such as inflation, monetary policy, and banking regulation. For example, there have been disputes in macroeconomic theory about whether inflation is endogenous (that is, it arises naturally) or exogenous to economic systems. This is one example where conservation principles prove their value in economic systems.

Chapter 8

Summary and conclusion

Bibliography

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

[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] 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.

[5] 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.

[6] "Fisher sees stocks permanently high; Yale economist tells purchasing agents increased earnings justify rise. Says trust and aid sales finds special knowledge, applied to diversify holdings, shifts risks for clients." *The New York Times*, oct 1929.

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

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

[9] 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.

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

[11] 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.

[12] 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

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

[14] 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.

[15] W.-K. Tung, *Group Theory in Physics*. Singapore: World Scientific Publishing, 1985.

[16] A. Baker, *Matrix Groups - An Introduction to Lie Group Theory*, ser. Springer Undergraduate Mathematics Series. London: Springer-Verlag, 2002.

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

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

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

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

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

[22] 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.

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

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

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

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

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

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

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

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

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

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

[33] 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.

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

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

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

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

[38] 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.

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

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

[41] R. Abraham and J. E. Marsden, *Foundations of Mechanics.* Addison-Wesley Publishing Company, 1978.

[42] 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.

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

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

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

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

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

[48] I. Fisher, *The Nature of Capital and Income.* New York: The Macmillan Company, 1906.

[49] ——, *The Theory of Interest.* New York: The Macmillan Company, 1930.

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

Glossary

List of Symbols

$\mathrm{d}\nu$	Weingarten map
ν	Gauss map
ω^1	Tautological 1-form
ω^2	Symplectic 2-form
Φ	Action functional
π	Projection map of a fiber bundle
$\boldsymbol{\delta}$	Kronecker delta
Δ	Discriminant
κ	Multiplier of a Möbius transformation
ρ	Radius of the pseudosphere
ζ	Hyperbolic angle, accumulated interest
$\boldsymbol{p}$	Generalized momentum / price levels
$\hat{\mathbb{C}}$	Extended complex plane
i	Imaginary unit
$\mathbb{T}^n$	Torus of dimension n
$\mathfrak{C}$	Cayley transformation
$\mathfrak{J}$	'Archetype' of a conjugacy class of Möbius transforms
$\mathfrak{M}$	Generic Möbius transformation
$\mathcal{A}$	Area (of a triangle)
$\mathbb{R}^{p,q}$	Lorentzian vector space
$\boldsymbol{g}$	Metric tensor
$\boldsymbol{u}$	Four-velocity
$\boldsymbol{F}$	First integrals of a Hamiltonian system
$\boldsymbol{n}$	Normal vector
$\boldsymbol{r}$	Position vector in three-dimensional space
c	Speed of light

e, f, g	Coefficients of the second fundamental form
H_+	Positive hyperboloid sheet
H_-	Negative hyperboloid sheet
I	Isomorphism defined by the symplectic 2-form
K	Original investment
r	Discrete-time interest rate
s_{12}	Spacetime interval
u, v	Surface parameters
V	Velocity (between reference frames)
v	Velocity
x	Capital
x, y, z	Spatial coordinates of spacetime
y	Yield
$\mathbb{C}$	Complex numbers
$\boldsymbol{q}$	Generalized position / stock levels
$\dot{\boldsymbol{q}}$	Generalized velocity / flow of goods
$\mathcal{H}$	Hamiltonian
j	Hyperbolic unit
$\mathcal{L}$	Lagrangian
$\mathcal{E}(T)$	Angular excess of a triangle
$[\mathfrak{z}_1, \mathfrak{z}_2]$	Projective coordinate set
$\mathbb{R}^n$	Real n-dimensional vector space
$\mathbb{S}^3$	3-sphere
a, b, c, d	Parameters of a Möbius transform
E, F, G	Components of the first fundamental form
J	Matrix representation of an archetypical Möbius transformation
k	Gaussian curvature
M	Configuration manifold
z_0	Fixed point of a Möbius transformation
T^*	Kinetic co-energy
U	Potential energy
$\mathbb{F}$	Arbitrary field
$\square_*$	Pushforward
$\square^*$	Pullback
$\bullet$	Lorentzian inner product
$\mathbf{I}$	First fundamental form
$\mathbf{II}$	Second fundamental form
$\sim$	Equivalence relation
Sp(2,)	Symplectic group

$\mathrm{GL}(n)$	General linear group
Lor	Lorentz group
$\mathrm{O}(1,3)$	Lorentz group (as indefinite orthogonal group)
$\mathrm{SO}(\cdot)$	Special orthogonal group
$\mathrm{SO}^{+}(1,3)$	Restricted Lorentz group
$TM_{\boldsymbol{x}}$	Tangent space to the manifold M at the point $\boldsymbol{x}$
$TM_{\boldsymbol{x}}$	Tangent space to the manifold M at the point $\boldsymbol{x}$
$\lVert\cdot\rVert_L$	Lorentzian norm
$\mathbb{CP}^1$	Complex projective line
$\mathbb{RP}^3$	Three-dimensional real projective space
TM	Tangent bundle of the manifold M
Möb	Möbius group
$\mathrm{PGL}(2,\mathbb{C})$	Projective linear group on the complex numbers
$\mathrm{PSL}(n,\mathbb{F})$	Projective special linear group
$\mathrm{SL}(n,\mathbb{R})$	Special linear group
$\mathrm{SU}(2)$	Special unitary group

Index