Resolution of Geometric Singularities by Complex Detours – Modeling, Complexity and Application

Thorsten Orendt

Zentrum Mathematik
Technische Universität München
D-85747 Garching

2010

TECHNISCHE UNIVERSITÄT MÜNCHEN
Zentrum Mathematik

Resolution of Geometric Singularities by Complex Detours – Modeling, Complexity and Application

Thorsten Orendt

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:		Univ.-Prof. Dr. Rupert Lasser
Prüfer der Dissertation:	1.	Univ.-Prof. Dr. Jürgen Richter-Gebert
	2.	Prof. Dr. Ulrich Kortenkamp Pädagogische Hochschule Karlsruhe
	3.	Univ.-Prof. Dr. Jurij Suris Technische Universität Berlin

Die Dissertation wurde am 15.4.2010 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 14.1.2011 angenommen.

Acknowledgement

"I can no other answer make, but, thanks, and thanks."

William Shakespeare

Completing this thesis has been an invaluable experience for me that would have never been possible, if many persons had not given their support, encouragement and advice.

First and foremost, I would like to thank my advisor Jürgen Richter-Gebert, who offered me the chance to develop in so many different directions. He is the one I have learned almost all of my entire geometric knowledge from. In addition, he is one of the authors of the dynamic geometry software package Cinderella that I have grown to like.

Another person that I want to express my gratitude to is Ulrich Kortenkamp for co-examining this thesis and also developing Cinderella.

I am also indebted to Yuri Suris for co-examining this thesis. He taught me a lot about differential equations and integrable systems.

Besides, I would like to thank Folkmar Bornemann, Ernst Hairer and Oliver Junge for their fruitful discussions and helpful remarks on numerical integration of ordinary differential equations.

There were also the members of the research group M10 that supported me during that time, namely Vanessa Krummeck, Jutta Niebauer, Gerd Fischer, Johann Hartl, Tim Hoffmann, Peter Lebmeir, Stefan Lembach and Hermann Vogel. I thank all of them.

Especially, I feel obliged to my former colleague Michael Schmid. On the one hand, it was only our collaborative work that made this thesis possible. On the other hand, I am glad to know him and consider him a friend.

My family has always been a motivating force in my life – my mother most of all. I am deeply indebted to them for supporting me during all the ups and downs, while working on this thesis.

At last, I would like to thank Judith Meyer, who endured my absence and has always been patient with me.

Contents

Abstract

This thesis deals with problems arising in dynamic geometry and numerical integration, where complex detours play an important role. In the first part, fundamental questions in continuous dynamic geometry are studied. We develop a formal model and investigate, among others, problems of existence. In the second part, complexity theoretic issues from continuous dynamic geometry are addressed. By doing so, new bounds for the so-called Reachability Problem are proved. In the third part, we apply complex detours in the context of numerical integration. It is studied, how this increases the order of convergence of a method.

Zusammenfassung

Die Arbeit besteht aus drei Teilen, welche alle im Zusammenhang mit ebener dynamischen Geometrie stehen und für die komplexe Umwege von zentraler Bedeutung sind. Nachdem wir einleitend auf grundlegende Eigenschaften dynamischer Geometrie und dynamischer Geometrie Systeme eingegangen sind, behandelt der erste Teil stetige dynamische Geometrie. Insbesondere wird ein Model für stetige dynamische Geometrie entwickelt. In einem zweiten Teil beschäftigten wir uns mit einem komplexitätstheoretischen Problem, welches in diesem Kontext entsteht. Der dritte Teil konzentriert sich auf einen geometrischen Effekt, der einen Konvergenzgewinn bei numerischer Integration zur Folge hat.

Stetige dynamische Geometrie entspricht in den meisten Fällen der natürlichen Erwartung, die ein Nutzer einer dynamischen Geometrie Software an das System stellt. Um jedoch mit stetiger dynamischer Geometrie formal arbeiten zu können, müssen wir zuerst einmal statische Geometrie mathematisch fassen. Wir beginnen mit geometrischen Objekten wie z.B. Punkten, Geraden oder Kegelschnitten. Sie werden mittels geeigneter homogener Koordinaten dargestellt, d.h. dass jedes Objekt mit einem Punkt eines speziellen projektiven Raumes identifiziert wird. Der gewählte projektive Rahmen ermöglicht es uns die Interaktion zwischen geometrischen Objekten, so genannter geometrischer Operationen, mittels homogener Polynomgleichungssysteme zu beschreiben. Der Schnittpunkt zweier Geraden ergibt sich z.B. als die Lösung der beiden zugehörigen linearen Geradengleichungen. Geometrische Konstruktionen formalisieren wir anschließend, indem wir ihre Konstruktionsvorschrift mit den eingeführten formalen Objekten und Operationen nachbilden. Ein konkretes Bild, d.h. eine Spezifikation der beteiligten Objekte, einer Konstruktion nennen wir eine Instanz.

Darauf aufbauend wird dynamische Geometrie modelliert. Im Unterschied zu einer gewöhnlichen Konstruktion wie man sie auf einem Blatt Papier schon mal gezeichnet hat, können sich die freien Objekte einer dynamischen Konstruktion bewegen. Hierbei lassen wir aber nicht alle Möglichkeiten an Bewegung zu, sondern beschränken uns auf solche, die letztlich von holomorphen Funktionen herrühren. Das bringt u.a. mit sich, dass die freien Objekte sich stetig bewegen. Eine Abbildung, die nun zu einer gegebenen Konstruktion und einer zugehörigen Startinstanz als Argument die Bewegungen der freien Objekte erhält und daraufhin die Bewegungen der restlichen Objekte zurückgibt, bezeichnen wir als ein dynamisches Geometrie System. Liefert es insbesondere nur stetige Bewegungen, so nennen wir es ferner ein stetiges dynamisches Geometrie System.

Im weiteren Verlauf des Kapitels gelingt es mit Hilfe einiger Hilfsaussagen zu zeigen, dass, wenn man die Bewegungen der freien Objekte fixiert, sich die Koordinaten der restlichen Elemente einer Konstruktion zu Beginn der Bewegungen mittels holomorpher Funktionen lokal beschreiben lassen. Unter Verwendung des Kontinuitätsprinzips folgt, dass die analytischen Fortsetzungen der Koordinatenfunktionen ebenso Instanzen der Konstruktion ergeben, die darüber hinaus ein stetiges dynamisches Geometrie System induzieren. Probleme ergeben sich nur, wenn die Fortsetzung nicht existiert oder der Repräsentant eines Objektes zum Nullvektor wird. Diese Lücken lassen sich jedoch mit Hilfsmitteln der Funktionentheorie schließen.

Mit den eben erwähnten Ideen konstruieren wir für jede zulässige Bewegung der freien Objekte einen stetigen Verlauf von zugehörigen Instanzen. Ferner gelingt es uns zu zeigen, dass die so induzierten stetigen dynamischen Geometrie Systeme die Menge aller möglichen

Systeme tatsächlich auch ausschöpft. Ebenso greifen wir den Punkt auf, dass der eingeführte algebraische Rahmen nicht mächtig genug ist, um transzendente Funktionen darzustellen. Gegen Ende des Kapitels gehen wir dann darauf ein, dass die Forderung nach holomorphen Bewegungen der freien Objekte in gewisser Weise sinnvoll ist. Denn es zeigt sich, dass nur k-fach differenzierbare Bewegungen i.A. nicht zu stetigem Verhalten aller Objekte einer Konstruktion führen.

Im dritten Kapitel dieser Arbeit interessiert uns die intrinsische Komplexität des so genannten Erreichbarkeitsproblems der dynamischen Geometrie. Gefragt wird, ob man zwei gegebene Instanzen einer Konstruktion stetig ineinander überführen kann. Um untere Komplexitätsschranken zu erhalten, ziehen wir uns auf das verwandte Konzept der straight-line Programme und deren Erreichbarkeitsproblem zurück. Im Unterschied zu straight-line Programmen im herkömmlichen Sinn lassen wir hier nicht nur arithmetische, sondern auch andere Operationen zu, welche den algebraischen Äquivalenten gewisser einfacher geometrischer Operationen wie z.B. Schnitt von einer Geraden und eines Kegelschnitts entsprechen. Dies werden komplexe Wurzeln und Logarithmen sein und sind von essentieller Bedeutung für die Komplexitätsschranken. Für die straight-line Programme führen wir ebenso einen Formalismus ein, der ein stetiges Verhalten dieser unter einer stetigen Bewegung der Eingabedaten beschreibt.

In analoger Weise zum vorhergehenden Kapitel stellt sich heraus, dass analytische Fortsetzungen der einzeln verbauten Operationen ein stetiges Verhalten des gesamten Programms induzieren. Somit bewegen wir uns auch hier auf zugrunde liegenden RIEMANNSCHEN Flächen dieser Funktionen.

Wir kodieren klassische Entscheidungsprobleme der Komplexitätstheorie wie 3SAT oder die Lösbarkeit diophantscher Gleichungen in die globale Gestalt RIEMANNSCHER Flächen und komplexer Mannigfaltigkeiten. Dabei ist die Kodierung so gewählt, dass das Ausgangsproblem genau dann mit ja beantwortet werden kann, wenn zwei vermeintliche Punkte der Fläche in der gleichen Zusammenhangskomponente liegen. Abhängig vom ursprünglichen Problem und dessen Lösung kann es passieren, dass die konstruierten Mannigfaltigkeiten in mehrere Zusammenhangskomponenten zerfallen und die Punkte in unterschiedlichen Komponenten liegen oder, dass einer der vorgegebenen Punkte gar nicht auf der Mannigfaltigkeit existiert. In diesen Fällen ist das Ausgangsproblem mit nein zu beantworten.

Im weiteren Verlauf des Kapitels wird die eben beschriebene Konstruktion verwendet, um untere Komplexitätsschranken für verschiedene abgeschwächte Varianten des Erreichbarkeitsproblems für straight-line Programme herzuleiten. Diese implizieren dann eine entsprechende Schranke an die verwandten geometrischen Probleme.

Bevor wir zwei unterschiedliche Einschränkungen des Problems studieren, schränken wir zuerst einmal den Informationsgehalt der Zielinstanz ein, indem wir nur einen Teil dieser vorgeben. Die restlichen Teile behandeln wir als Black Boxes. Von diesen Problemen können wir zeigen, dass sie abhängig von den zugelassenen Operationen NP-schwer, coNP-schwer oder unentscheidbar sind. Die anderen Einschränkung, die wir machen, beziehen sich auf die Bewegungen der Eingabedaten. Hier werden wir einmal die topologische Form dieser vorgeben, indem wir nur JORDAN-Kurven zulassen. Ferner beschränken wir die Länge der Bewegungen in einer zweiten Variante. In beiden Fällen ergibt sich, dass die zugehörigen Entscheidungsprobleme NP-schwer sind.

In Kern des vierten Kapitels stehen ein Superkonvergenzeffekt, der bei der numerischen Integration entlang komplexer Integrationswege gewöhnlicher Differentialgleichungen auftritt, und dessen Interpretation. Sie resultieren aus einem komplex-geometrischen Zugang.

Bevor wir jedoch diese Thematik aufgreifen, wird kurz auf den Zusammenhang zwischen stetiger dynamischer Geometrie und numerischer Integration gewöhnlicher Differentialgleichungen eingegangen.

Anschließend motivieren wir den angesprochen Superkonvergenzeffekt anhand des einfachen Beispiels $\dot{x} = x$. Dieses Beispiel wird uns als Leitfaden für die folgenden Untersuchungen dienen.

Nachdem die nötige Notation eingeführt wurde, werden wir eine auf linearen Differentialgleichungen und Runge-Kutta Verfahren basierende Theorie entwickeln, mit der der zu Beginn entdeckte Effekt erklärt wird. Es stellt sich heraus, dass man durch Wahl einer speziellen komplexen Diskretisierung die Konvergenzordnung am Endpunkt des Integrationsweges steigern kann. Der Schlüssel wird dabei eine geometrische Überlegung sein.

Im folgenden Abschnitt wird die entwickelte Theorie auf nicht lineare Differentialgleichungen verallgemeinert. In diesem Zuge wird es uns gelingen, eine Klasse so genannter *composition methods* mit komplexen Koeffizienten herzuleiten und einige der bisher bekannten Verfahren dort einzugliedern. Es stellt sich nämlich heraus, dass diese sich als Iteration eines gewissen Basisverfahrens wie z.B. ein Runge-Kutta Verfahren entlang einer geeigneten Diskretisierung eines komplexen Integrationsweges ergeben.

Diese Methoden können wiederum mittels geschickter Iteration dazu genutzt werden, um Verfahren höherer Ordnung zu konstruieren. Interpretiert man diese dann wieder auf Basis einer komplexen Integration, so ergeben sich interessanterweise Integrationswege mit einer fraktalen Struktur.

1

Introduction

"Es bedarf nur eines Anfangs, dann erledigt sich dass Übrige."

Gaius Sallustius Crispus

150 years ago, there was a great tradition in building mathematical models. At many universities, a student of that time could even attend special seminars, in which building such models was taught and studied. As one may not expect today, the reason for this tradition was not only to produce nice sculptures to look at and put them into a showcase. Indeed, the purpose of building models was to improve the understanding of the mathematical theory behind them. The famous mathematician Felix Klein wrote on this subject in [28, p. 78] once:

> "Wie heute [1926], so war auch damals der Zweck des Modells, nicht etwa die Schwäche der Anschauung auszugleichen, sondern eine lebendige, deutliche Anschauung zu *entwickeln*, ein Ziel, das vor allem durch das Selbstanfertigen von Modellen am besten erreicht wird."

In his opinion, building mathematical models is not only a tool for visualization, but also a method in order to enhance one's mathematical skills and knowledge. It is the deeper insight in mathematics that follows a construction process.

Nowadays, mathematicians are still building mathematical models. Most of them, however, are produced on a computer, but the fact that they improve the understanding of mathematics has not changed over the years.

1.1 Dynamic Geometry and Dynamic Geometry Systems

Imagine a construction of three angular bisectors of a triangle that meet in a point. This is an example of a well-known and elementary geometric construction, where the positions of the three vertices of the triangle do not play any role for the concurrence of the corresponding angular bisectors. Due to this fact we can regard the vertices as *free* points and the other objects as *dependent* elements of this construction. For each choice of the free points, one gets an associated drawing, for which the underlying angular bisectors will still meet in a point. All these drawings are static *instances* of the same elementary geometric construction. Starting from an instance, we can move the free points continuously in the plane. It is possible to move the dependent elements also in a continuous way so that the dependent elements complete them to an instance of this construction at each position of the free points during their movement.

The picture just sketched is an example of a dynamic construction[1]. *Dynamic geometry* is the field that deals with dynamic constructions. In contrast to usual elementary geometry, where a construction is static, dynamic geometry is equipped with an additional parameter,

[1] This and another one are illustrated in Figure 1.1.

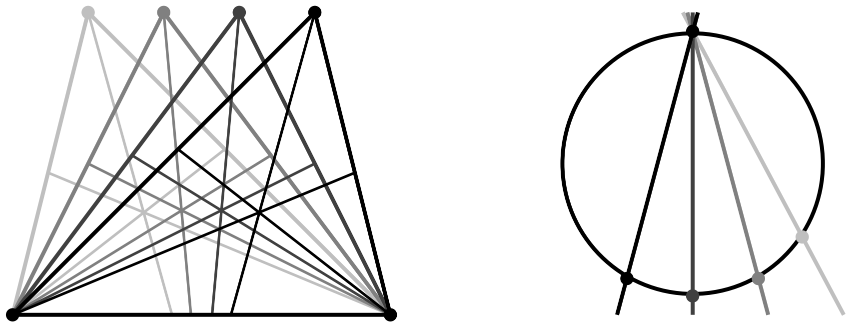

Figure 1.1: Examples of dynamic constructions.

which enables us to deform instances as time goes by. Nowadays, there are several software packages such as Cinderella [47] providing dynamic geometry on a computer. They are implementations of a formal model that we call a *dynamic geometry system*. The idea of a dynamic geometry system is as follows: Starting from a construction and a corresponding starting instance, a dynamic geometry system gets the movements of the free elements as input and outputs the associated movements of the dependent objects.

Since a dynamic geometry system should provide a reasonable behavior of the dependent elements, it has to tackle several problems arising in dynamic geometry. In the first place, we would like to have *consistency*, this is, preserving that the geometric objects form an instance of an underlying construction during a moving process. To accomplish consistency, each dependent element has to satisfy two conditions. On the one hand, the objects have to keep their object identities. For instance, a point should be a point during the entire movement and not become a line. On the other hand, the objects have to satisfy several relations given by an underlying construction during the movement. For example, if a point is the intersection of a line and a circle, then, in fact, it has to be one of the two possible points of intersection, but not a point somewhere else in the plane.

This leads us directly to another problematic issue that is already inherited from static geometry – *ambiguity*. Therefore, we recall the geometric situation of a circle and a line that intersect. At it, we can understand the occurring problem easily. There are obviously constellations so that two points of intersection exist. They are consistently equivalent. Due to this fact it is natural to ask, which of them to choose, if we have to choose only one of them. This may become difficult in the presence of continuity requirements for a dynamic geometry system, when the line passes through a tangent situation. As a result, resolving ambiguity is crucial for a dynamic geometry system.

Apart from consistency and the problem of ambiguity, another desirable and intuitive feature, which has not yet been mentioned, is *determinism*. This concept ensures that performing the same movements from the same starting instance yields always the same motion of the dependent elements.

Based on determinism, there are two more useful features, which one can claim as well, this is, *conservatism* and *continuity*. Conservatism means roughly that same terminal points of the free elements result in same terminal positions of the dependent objects independent of the path they have taken. This concept simplifies the implementation of a dynamic geometry system a lot, since the results only depend on the terminal positions of the free elements, but not on the movements themselves. Continuity, however, means that the motions of the dependent objects should also be continuous during a continuous movement of the free

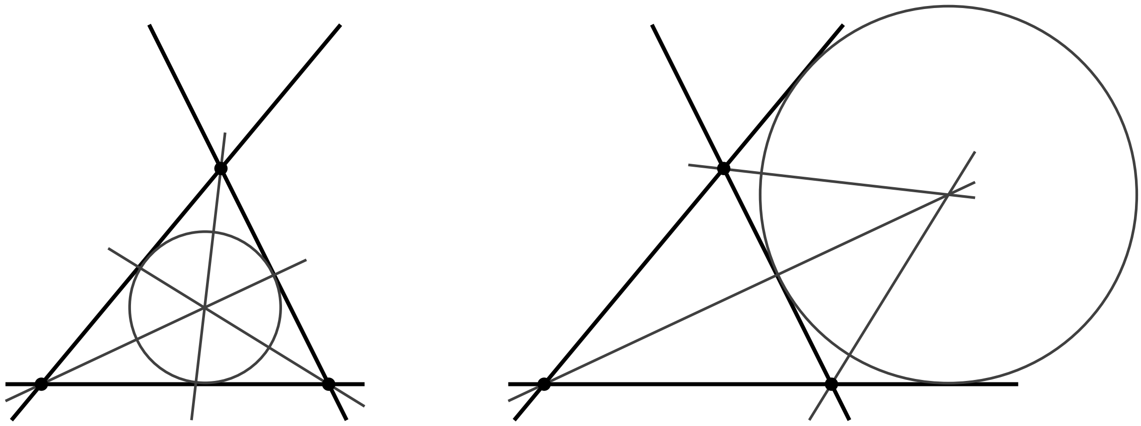

Figure 1.2: Variations of an “incircle” of a triangle.

elements. Like determinism, continuity is not automatically implied by consistency. This can be understood at the example mentioned before of a triangle and its angular bisectors. As a first step, we observe that, similar to the case of an intersecting circle and a line, there are two consistent choices of an angular bisector – an inner one that proceeds through the interior of the triangle and an outer one, which does not. The reason for this is that two intersecting lines always form an inner and an outer angle. So at each corner of the triangle, we can choose between an inner and an outer angular bisector. Now starting from an instance, where all three bisectors are inner ones (compare Fig. 1.2 (left)), we assume to perform an infinitesimally small movement with one of the vertices. By doing so, the associated dynamic geometry system could output two of the outer angular bisectors that do not proceed through the inside of the triangle at the end of the movement. Hence we end up with a drawing as illustrated in Figure 1.2 (right). The three angular bisectors still meet in a point implying the consistency of the instance, but their behavior cannot have been continuous, since a continuous change from an inner bisector to an outer one or vice versa is only possible, if they coincide during the movement. Otherwise, the transition cannot have been continuous. However, this is impossible, since an inner and a corresponding outer angular bisector are always orthogonal.[2] As a result, the behavior of the angular bisectors cannot have been continuous during the movement of the free point.

Furthermore, it is worthwhile to mention that conservatism and continuity are mutually exclusive. We would like to exemplify this point, now. Therefore, we claim continuity and assume that we have got two intersecting lines, one of their angular bisectors and a point on this bisector different from the point of intersection of the lines. The first two lines serve as free elements and the remaining objects are dependent. Rotating one of the free lines around their point of intersection by an integer multiple of 2π results in a rotation of the angular bisector around the same center, but with half of the angular velocity of the rotating line. At the end of the rotation, the two lines and the angular bisector are again in their initial positions, though the point on the angular bisector is not. Like the bisector, it moves with half of the angular velocity of the free line. Hence the point is at the position that one gets by mirroring the point in its initial position at the rotation center. It is clear that performing this movement once more yields the initial configuration of all objects. So essentially, there are two different movements of the free line, when it is rotated

[2] A corresponding proof can be found in [46]

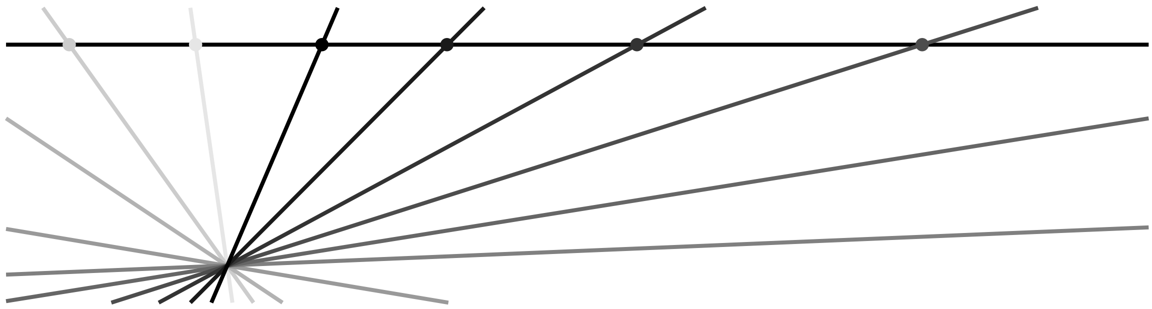

Figure 1.3: Intersecting lines in motion.

by an integer multiple k of 2π, this is, k is either odd or even. Both rotations have the same terminal point, but the dependent elements, particularly the point on the angular bisector, do not end up under a continuous motion at the same positions for both movements. This is in contradiction to conservatism.

The last investigation shows another characteristic of a continuous dynamic geometry system. Roughly speaking, the terminal positions of the dependent elements depend on the movements we have chosen. Such an effect is called *monodromy* and occurs naturally in complex function theory as well. It is related to analytic continuations and RIEMANN surfaces.

1.2 A Complex and Projective Setup

The last paragraph already indicates that a complex framework is appropriate to deal with continuous dynamic geometry. In addition, we will exemplify below that a projective point of view is also reasonable. It is well-known that projective geometry, particularly complex projective geometry, is excellently suited for working with geometry, in general. Nevertheless, we would like to exemplify briefly the benefit of this point of view. For further readings on this topic, we refer to [46, 51]

A great advantage of projective geometry is its generality. It is projective geometry that removes many special cases that arise in affine geometry. A standard example deals with two intersecting lines in the plane that become parallel. Imagine these two intersecting lines. Now, we choose a point on one of the lines that is not their point of intersection and start rotating this line around this chosen point. By doing so, the point of intersection of these two lines slides along the other resting line until the lines become parallel. In this position, the point vanishes and does not exist in affine geometry causing a special case. However, if one keeps on rotating, the point of intersection will occur again and return from the other end of the resting line. Contrary to the just drawn picture, there is always a well-defined point of intersection in projective geometry. Even when the two lines are parallel. Loosely speaking, one can think of the point of intersection in case of parallelism as a point that lies infinitely far away in an associated direction. Moreover, it is not hard to show that the motion of this point completed by a point at infinity is continuous, even in the situation of parallelism, in a properly chosen projective space equipped with a suitable metric.

A further point, which we should address, is that *complex numbers* embed our studies in more general framework. Their use is also exemplified easily. This time, we consider two intersecting circles. Their two points of intersection define a unique line. If the circles

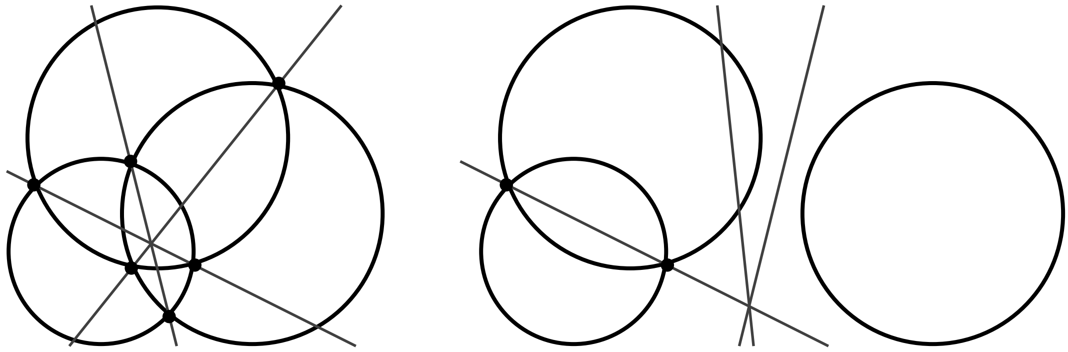

Figure 1.4: An illustration that theorems also holds over the complex numbers.

are moved sufficient far away from each other, the two points of intersection will vanish. A convenient way to describe these points is by the solutions of a system of polynomial equations. In other words, the points of intersection and the solutions of the polynomial system are in one-to-one correspondence. Over the complex numbers $\mathbb{C}$, the solutions of this system still exist, even when two circles do not intersect anymore. As a result, they still provide a reasonable geometric meaning in a proper complex space. The reason for their disappearing is that their coordinates have become complex. A surprising fact is that, though the points have complex coordinates, the line joining them can be, in turn, identified as a real line and is therefore visible and well-defined in the affine plane again.

1.3 Outline and Summary

This thesis consists of three parts concerning dynamic geometry, to which complex detours are central. After we have introductory dealt with basic features of dynamic geometry and associated systems, the first part is about continuous dynamic geometry. Particularly, we will develop a model for continuous dynamic geometry (compare [32, 48]). In a second part, we are going to study a complexity theoretic issue emerging in this field, this is, the so-called Reachability Problem in dynamic geometry. The third part is a joint work with Jürgen Richter-Gebert and Michael Schmid (compare [39]). It concentrates on a geometric effect that occurs at numerical integration along certain complex paths of integration and improves the order of convergence of the used method.

1.3.1 Continuity and Dynamic Geometry

In the majority of cases, continuous dynamic geometry meets the expectations that a user demands from a dynamic geometry software package. A very important question that arises in this context is the following: Given a construction and an instance, is there always a possibility to adjust the dependent elements continuously under continuous movements of the free objects? Or in other words, does there always exist a continuous dynamic geometry system to a given construction and a corresponding starting instance? In [32], Kortenkamp and Richter-Gebert showed that, in a precise sense, the answer to this question is yes, if the movements of the free objects are linear and the construction steps of the underlying construction are of algebraic nature. Some further works that also deal with fundamentals in continuous dynamic geometry are [30, 50]. We use them as a basis for the second chapter, in which we extend the results presented in [32]. The outline of this chapter is as follows:

To work with continuous dynamic geometry formally, we are going to develop a proper mathematical model of static geometry in Subsection 2.2.1. Introducing homogenous coordinates will yield an appropriate representation of geometric objects such as points, lines or conics. The interaction between them can be described by suitable homogenous polynomial systems due to the chosen projective viewpoint. For example, the point of intersection of two lines is obtained by solving two linear equations simultaneously. Next, we are going to formalize geometric constructions by stating a sequence of construction steps and associated references. Our formal concept of a construction will not correspond to a concrete drawing, but to an instruction how to produce it. The notion of a concrete drawing is, in turn, captured by our concept of instances of a construction.

The next step is going to be modeling dynamic geometry. This will be done in Subsection 2.2.2. In contrast to usual constructions as one has already drawn them on a sheet of paper, the free objects of a dynamic construction can move. Hereby, we will not admit all possible movements, but we allow those that are restrictions of holomorphic functions. This implies, among others, certain continuous movements of the free elements. After that, we are going to define (continuous) dynamic geometry systems formally.

In Section 2.3, we are going to provide some lemmata that are required to show the main results of this chapter. Based on fixed movements of the free elements, we are going to constitute holomorphic functions that will describe the coordinates of the involved objects (Subsection 2.3.1). Analytic continuations of these coordinate functions will induce further instances of the underlying construction, which, in turn, will induce a continuous dynamic geometry system. Problems can only occur when either an analytic continuation does not exit or a representative becomes the all-zero vector, but they can be solved by techniques from function theory.

The just mentioned idea will always enable us to find continuously behaving instances along continuously moving free objects (Subsection 2.4.1). Moreover, we will be able to prove that all possible continuos dynamic geometry systems are captured by our approach. At the end of this chapter, i.e. in Subsection 2.4.2, we are going to demonstrate that transcendental functions are not constructible by dynamic algebraic constructions. Finally, we will close this chapter with Subsection 2.4.3, in which it will be proved that only k-times differentiable movements of the free objects do not imply a continuous behavior of all objects, in general.

1.3.2 Complexity and Dynamic Geometry

Apart from the fundamental questions on continuous dynamic geometry systems discussed in the previous chapter, complexity theoretic questions arise in dynamic geometry as well, particularly the so-called *Reachability Problem*. Roughly speaking, the problem is to decide, whether a given instance of a construction can be deformed continuously into another specific one. An early formulation can be found in [30], whereas a more common one is stated in [48]. The complexity of this problem depends closely on the power of the allowed geometric operations to form a construction and the restrictions that are imposed on the movements of the free objects. In [11], it is basically shown that the Reachability Problem is decidable, if construction steps are used that are at most algebraically equivalent to solving quadratic polynomials. What is more, Richter-Gebert and Kortenkamp proved in [48] that the problem is at least NP-hard in $\mathbb{R}$ for a similar setup and even PSPACE-hard or undecidable for a more powerful setup, which also includes construction steps of transcendental power. Besides, they showed in [31] that the Reachability Problem in this setup in $\mathbb{C}$ is at least as hard as testing, whether a corresponding straight-line program built of arithmetic operations and square roots vanishes identically. However, lower complexity bounds in $\mathbb{C}$ concerning these set of operations have not been established for at least ten years. Several results concerning the complexity of variants of the Reachability Problem in $\mathbb{C}$ will be derived in the third chapter of this thesis.

Thereby, we are going to rely on the related concept of straight-line programs and its corresponding reachability problem implying lower complexity bounds for the geometric problem as well (Section 3.2). In contrast to usual straight-line programs only consisting of arithmetic operations, we are going to admit operations that are induced by certain geometric operations such as the intersection of a conic and a line. They will be complex roots and logarithms.

In Section 3.3, a formal concept is going to be introduced, which will formalize a continuous behavior of a straight-line program under a continuous change of the input data. Similar to the previous chapter, it will turn out that analytic continuations of the used operations will induce a continuous behavior of the entire program. Roughly speaking, walking on Riemann surfaces induced by the operations of a program is continuous.

The main idea will be to encode classical problems from complexity theory such as 3SAT or Hilbert's 10th in global shapes of Riemann surfaces and complex manifolds. Thereby, the encoding will be chosen so that an original problem from complexity theory will have the answer yes, if and only if two supposed points of the constructed manifold lie in the same path component. Dependent on the chosen instance of the original decision problem, it can either be that the manifold decomposes into several components and the two points lie in different components or one of the supposed points does not even exist on the manifold. In this cases, the instance of the original problem will have the answer no.

In the subsequent paragraphs, we are going to study related problems and several weakened variants of the reachability problem concerning straight-line programs. On the one hand, we are going to reduce the information content of the terminal instance by specifying only a part of it (Section 3.4). The remaining parts are going to be treated as black boxes. These problems are going to be proved NP-hard, coNP-hard or even undecidable depending on the admitted operations. On the other hand, we are going to restrict the movements of the input data in Section 3.5. In this way, we are going to derive two further lower bounds. If we only admit Jordan curves or bound the length of the movements of the input data, then the problems will be NP-hard.

1.3.3 Geometry and Numerical Integration

On the first sight, there may seem to be no connection between dynamic geometry and numerical integration, but there is, in fact. On the one hand, the motion of a continuous dynamic construction under certain movements of the free elements can be modeled as an initial value problem, which must be integrated numerically, in general.

As already mentioned, it will turn out in the second chapter that the used complex approach is reasonable, since it provides continuous dynamic geometry. It is against this background that one can ask, whether this philosophy also offers some benefits concerning numerical integration of initial value problems.

Basically, one possible benefit is that we additionally have the opportunity to make a detour that avoids occurring singularities. For instance, imagine a two-body problem with one fixed and one moving body, whose initial velocity points directly towards the other body. Then they are going to collide at some point in time. Detour strategies concerning this situation motivated my research in this field initially and were studied in [38]. In addition, other examples of detouring singularities were discussed in [9]. It is this additional freedom that enables one to gain a computational benefit as well. Obviously it holds: The farer a path of integration stays away from singular situations the lower the resolution of an adaptive integrator will be (see [9]). However, there can also be achieved a second computational benefit, i.e. a superconvergence effect[3]. It can be obtained in the presence of certain complex time grids. Mainly, this effect will be the subject of the 4th chapter. Furthermore, it will turn out that this will be the key to gain a geometric access to an entire class of composition methods with complex coefficients apart from the usual way of solving a suitable set of order conditions over the reals or complex numbers such as done in [23, 24, 58]. In detail, the chapter is organized as follows:

In Section 4.1, we are going to present a simplest possible scenario, in which the superconvergence effect is studied. This example, namely complex detours for $\dot{x} = x$, will serve as a motivating paradigm for our further considerations.

Section 4.2 will introduce the necessary setup of complex time grids in relation to Runge-Kutta methods.

The main technical part of this chapter is going to be developed in Section 4.3. Hereby, we are going to rely on linear systems of ordinary differential equations with constant coefficients. At first, we are going to deal with the problem, when a Runge-Kutta method applied to a complex path of integration yields a real terminal point (Subsection 4.3.1). Thereafter, we are going to prove that one can achieve a superconvergence effect concerning a Runge-Kutta method applied to a linear ODE by choosing a suitable complex path of integration (Subsection 4.3.2). By doing so, we will gain a lot of geometrical insight in the structure of possible paths and can derive explicit criteria that have to be satisfied in order to obtain superconvergence. These criteria will be closely related to the multiplicative structure of roots of unity in the complex plane.

Section 4.4 is going to link complex detours with composition methods. This connection will enable us to extend our method to the case of a nonlinear right-hand side of an ordinary differential equation. A composition method is obtained by applying an integration method, say e.g. a Runge-Kutta method, consecutively in a controlled way. The sequence of these applications can again be considered as a single step of a more complicated integration method. If this sequence is chosen properly, one can increase the order of the original method by at least one. It will turn out that this is closely related to the condition derived in the previous section achieving superconvergence in the linear case. A geometric approach will allow us to solve these conditions canonically and to interpret some already known composition methods entirely on the level of an integration along a complex path of integration. In other words, these composition methods will correspond to an application of a basic method along a suitably chosen complex detour. We can even iterate[4] this process and by doing so, we are going to obtain methods of arbitrary high order (Subsection 4.4.2). Surprisingly, the corresponding paths that encode the iterative methods will exhibit a fractal structure. At the end of this section, we are going to illustrate our approach by the sophisticated problem of computing the Arenstorf orbit.

[3]means that the method converges faster than expected at the terminal point.
[4]This iteration is also known as the "Yoshida trick" [61].

2

Continuity and Dynamic Geometry

> *"Es [das Prinzip der Kontinuität] sagt aus, daß eine an einer Figur mit hinreichender Allgemeinheit erkannte Beziehung auch für alle anderen Figuren gilt, die sich aus ihr durch kontinuierliche Lagenveränderung ableiten lassen."*
>
> *Found in Klein's book [28]*

The quote stated at the beginning is a description of PONCELET's famous principle of continuity published in his work [44] on projective geometry. Although, his formulation was vague and he did not provide any proof of his principle (compare [22, 28]), it is well-known today and there are other areas in mathematics such as analysis that make use of it (see [29, 52]). A more modern formulation is given by:

> If an analytic identity in a finite number of variables holds on an infinite set with accumulation point, then it also holds by all possible analytic continuations.

It is this principle that is going to build the foundation of the results in this chapter, since we are going to encounter that geometric objects and their continuous motions are described by analytic functions and their analytic continuations. But prior to this, we discuss similarities and distinctions between the present chapter and other works.

2.1 Comparison to Other Works

The comparison focuses on [30, 32]. In there, a model for continuous dynamic geometry was already developed and many of our definitions and ideas follow those presented in these works.

Similarities and Differences Between the Models

Basically, projective geometry is used in all three writings to describe geometric objects. In addition, corresponding geometric operations are formalized by suitable relations.

However in contrast to [30, 32], we additionally allow complex homogenous coordinates in our model. For this reason, we can always perform a geometric operation and thus we always obtain a well-defined geometric object as output of a geometric operation. That this is not the case, if one only allows real coordinates, can be seen at the following simple consideration: For instance, consider an intersection of a line and a circle such as in Example 2.1. Depending on their positions, there are two real points of intersection or not. The restriction to real coordinates causes an empty output set of a corresponding geometric operation in situations, where only points of intersection with complex coordinates exist. To cover cases like these, the model introduced in [32] is adapted in several ways. Each set of geometric elements of equal type is extended by another object $*$. It is defined to be the output of a geometric operation, where either the geometric situation is degenerated or

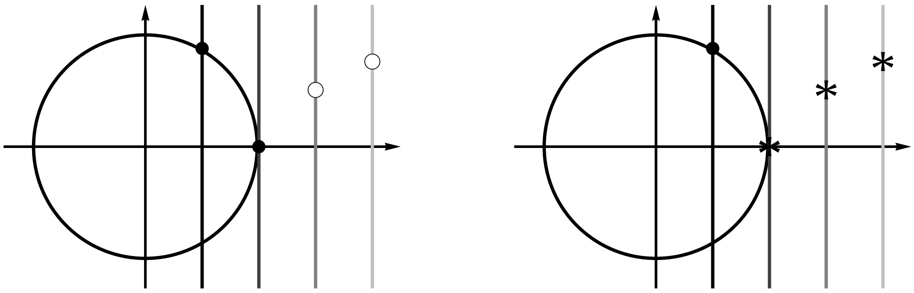

Figure 2.5: Illustrating Example 2.1. Real (•) and complex (∘) points of intersection in our model are shown on the left, while their pendants from [32] can be seen on the right.

at least one output object with complex coordinates exists (compare Fig. 2.5). Similarly, geometric operations are also extended by additional relations capturing the cases, in which $*$ is involved. Moreover, we should mention that the other definitions in [32] also rely on this way of dealing with degenerate cases and complex objects. Such a concept is obsolete in our model. To illustrate this, we consider a simple example at the end of this paragraph.

Apart from the differences arising from the fact that we admit complex instead of real numbers, our subsequent definitions of algebraic operations[5], geometric constructions, instances of a geometric construction[6] and continuous dynamic geometry systems basically follow again those in [30, 32].

The last considerable difference between the models concerns the movements of the free elements of a construction. While in [32] the movements of free elements are restricted to linear functions, we allow holomorphic functions in our model.

Example 2.1. At this point, we point out the difference between the underlying model in [32] and our ones used in this thesis. We consider the following geometric situation: Given the unit circle defined by $x^2 + y^2 = 1$ and a line $x = t$, where $t \in [0, 2]$. The coordinates of their two points of intersection are determined by

$$\begin{pmatrix} t \\ \sqrt{1-t^2} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} t \\ -\sqrt{1-t^2} \end{pmatrix}.$$

Obviously, there are two distinguished real points for $t \in [0, 1[$, exactly one real point for $t = 1$ and two distinguished complex points for $t \in]1, 2]$ (compare Fig. 2.5). Now, we compare the output of a geometric operation that yields one of these points of intersection in the different models. In [32], such an operation outputs one of the points for $t \in [0, 1[$ and $*$ for $t \in [1, 2]$. This also implies that the object $*$ can occur over an entire interval or movement. In contrast, we obtain always a point for all $t \in [0, 2]$ in our model.

Comparing the Results

As a first step, we shall remark that we use, to some extent, the same ideas and techniques as in [32]. Examples are resultants, series expansions or the idea used in the proof of Lemma 2.9. In doing so, our model enables us to confirm several results presented in [30, 32] in greater detail and to extend them.

[5]The formulation in [32] is somewhat imprecise. We present an exact definition.
[6]They are also found as *geometric straight-line programs* in [12, 48].

As in [32], it basically turns out in this chapter that the coordinates of geometric objects and their motions are described by holomorphic functions and their analytic continuations. However in contrast to [32], we can prove the existence of continuous dynamic geometry systems not only for linear movements, but also for movements that are induced by general holomorphic functions. Furthermore, the results are extended by showing that all possible continuous dynamic geometry systems are indeed captured by our model. Moreover, we provide a proof for a remark in [32] that a continuous dynamic geometry systems cannot be based on movements induced by only finitely differentiable functions. Last but not least, we take up the concept of constructible functions from [30]. In this way, we can additionally show that some fundamental transcendental functions like the exponential function or the logarithm are not constructible in the underlying algebraic setting.

2.2 Modeling Dynamic Geometry

In the following section, we will develop a reasonable framework that model geometric constructions as well as the dynamics of them. Thereby, our point of view will be a projective one inducing that all geometric objects are represented in homogeneous coordinates and their interactions are described by relations that can mainly be represented by zero sets of multi-homogeneous polynomial systems. As mentioned in the introduction, we would like to avoid case distinctions and to derive a uniform theory, so we are going to admit complex numbers as well. Starting from modeling a static geometric construction, we are going to continue with capturing the notion of dynamic constructions.

2.2.1 Objects, Operations and Constructions

A geometric construction consists of several parts, namely objects and relations describing their interplay. Examples of geometric objects are points, lines or circles. In our model, they are given formally by their homogeneous coordinates.

Definition 2.1 (geometric objects)**.** *Let*

$$\mathbb{CP}^d := \frac{\mathbb{C}^{d+1} \setminus \{(0,\ldots,0)^T\}}{\mathbb{C} \setminus \{0\}}.$$

An element of $\mathbb{CP}^d$ *is an equivalence class* $[p]$ *with representative* $p \in \mathbb{C}^{d+1} \setminus \{(0,\ldots,0)^T\}$*. To simplify the notation, we also denote* $[p]$ *by* $\mathbf{p}$*. Furthermore, we define the set of points* $\mathcal{P}$ *and lines* $\mathcal{L}$ *as two disjoint copies of* $\mathbb{CP}^2$*, the set of conics* $\mathcal{C}$ *as a copy of* $\mathbb{CP}^5$ *and the set* $\mathcal{K}$ *of algebraic curves of degree 3 as a copy of* $\mathbb{CP}^{10}$*. We denote* $\mathcal{T} := \{\mathcal{P}, \mathcal{L}, \mathcal{C}, \mathcal{K}\}$ *as the set of geometric types.*

On the basis of the introduced geometric objects, the (static) interaction between them will be encapsulated by adequate relations, which are most often described by special multi-homogeneous polynomial systems. The use of relations in this context is necessary to capture the introductory mentioned ambiguity, which occurs naturally in geometry.

Definition 2.2 (geometric operations)**.** *We call a relation* $\omega \subseteq I_1 \times \ldots \times I_k \times O,\ k \in \mathbb{N}$*, an* operation*, if* $I_1, \ldots, I_k, O \in \mathcal{T}$*. Throughout this article, we set* $\mathrm{in}(\omega) := k$*. If* O *is a copy of* $\mathbb{CP}^d$*, we call* ω *algebraic, if*

$$\omega = \Big\{ \big([i_1], \ldots, [i_k], [o]\big) \in I_1 \times \ldots \times I_k \times O \;\Big|\; P_\omega(i_{1,1}, \ldots, i_{k,d_k+1}, o_1, \ldots, o_{d+1}) = 0 \Big\},$$

where $P_\omega \subset \mathbb{C}[X_{1,1}, \ldots, X_{k,d_k+1}, X_1, \ldots, X_{d+1}]$ *is a system of* d *multi-homogeneous polynomials. Furthermore we define*

$$\omega(\mathbf{i}_1, \ldots, \mathbf{i}_k) := \left\{\mathbf{o} \in O \mid (\mathbf{i}_1, \ldots, \mathbf{i}_k, \mathbf{o}) \in \omega\right\}.$$

As extension, we call $\omega \in \mathcal{T}$ *a* free operation. *In this case, we set* $\mathrm{in}(\omega) := 0$.

Remark 2.1. Since we admit complex numbers, it holds: If ω is an algebraic operation, Bezout's theorem yields $\omega(\mathbf{i}_1, \ldots, \mathbf{i}_k) \neq \emptyset$ guaranteeing that an algebraic operation can always be performed, since it always provides an adequate geometric object as output. Observe that there is no need to stop at curves of degree 3, since all plane algebraic curves of a fixed degree can be represented by a complex projective space of adequate dimension and Bezout's theorem is also applicable to them. In addition, the results presented in this chapter generalize straight forward to algebraic curves of a higher degree as well. However in order not to complicate the following considerations and since we will only need the objects from the set $\mathcal{T}$ in the second chapter, we focus on them.

In the following, we introduce some fundamental algebraic operations, which will be used later on in this article. A derivation of the describing polynomial systems for each operation is mostly self-explanatory, but one can find most of them in [46].

Example 2.2. (MEET & JOIN) Two of the most fundamental (and self-explanatory) algebraic operations are given by the intersection of two lines

$$\mathrm{MEET} := \Big\{\big([g], [h], [p]\big) \in \mathcal{L} \times \mathcal{L} \times \mathcal{P} \mid g^T p = 0 \text{ and } h^T p = 0\Big\}$$

and by the join of two points

$$\mathrm{JOIN} := \Big\{\big([p], [q], [l]\big) \in \mathcal{P} \times \mathcal{P} \times \mathcal{L} \mid l^T p = 0 \text{ and } l^T q = 0\Big\}.$$

An issue that we should also address here are degenerate cases. Since they are not problematic at all, even for the subsequent examples, we discuss them exemplarily at the operation MEET. For that, denote the all-zero vector $(0,0,0)^T \in \mathbb{R}^3$ by o. Although, o satisfies $g^T o = 0$ and $h^T o = 0$ for all $g, h \in \mathbb{R}^3 \setminus \{(0,0,0)^T\}$, it holds $\big([g], [h], [o]\big) \notin \mathrm{MEET}$, since o does not represent a point from $\mathcal{P}$. Lastly, we have to consider the intersection of two lines $[g], [h] \in \mathcal{L}$ with $[g] = [h]$. But in this case, it is easy to observe that each point $\mathbf{p} \in \mathcal{P}$ on the line $[g]$ satisfies the relation MEET, i.e. $\big([g], [h], \mathbf{p}\big) \in \mathrm{MEET}$.

Example 2.3 (constants). Also constant objects can be represented by a suitable chosen algebraic operation. For example, let $\mathbf{p} \in \mathcal{P}$ and $[g], [h] \in \mathcal{L}$ two distinct lines through $\mathbf{p}$. Then

$$\mathrm{CONST}_{\mathbf{p}} := \Big\{\big([q]\big) \in \mathcal{P} \mid g^T q = 0 \text{ and } h^T q = 0\Big\}$$

represents the constant point $\mathbf{p}$.

Example 2.4 (intersection of two conics). A more sophisticated algebraic operation is given by the intersection of two conics. Let $\mathbf{c} = [c] \in \mathcal{C}$ be a conic. Then $\mathbf{c}$ is determined by the zero set of a polynomial

$$P_{\mathbf{c}} = c_1 X^2 + c_2 Y^2 + c_3 Z^2 + c_4 XY + c_5 XZ + c_6 YZ.$$

Furthermore, we denote the entries of a vector p by p_1, p_2 and p_3. By using this conventions, we define the algebraic operation

$$\mathrm{MEETCC} := \Big\{\big(\mathbf{c}_1, \mathbf{c}_2, [p]\big) \in \mathcal{C} \times \mathcal{C} \times \mathcal{P} \mid P_{\mathbf{c}_k}(p_1, p_2, p_3) = 0 \text{ for } k = 1, 2\Big\}.$$

Thereby, $\mathrm{MEETCC}(\mathbf{c}_1, \mathbf{c}_2)$ is the set of intersection points of $\mathbf{c}_1$ and $\mathbf{c}_2$.

Example 2.5 (intersection of a conic & a line). The intersection of a conic and a line is represented by

$$\mathrm{MEETCL} := \Big\{ \big(\mathbf{c}, [l], [p]\big) \in \mathcal{C} \times \mathcal{L} \times \mathcal{P} \;\big|\; P_{\mathbf{c}}(p_1, p_2, p_3) = 0 \text{ and } l^T p = 0 \Big\}.$$

Example 2.6 (intersection of an algebraic curve of degree 3 & a line). Similar to a conic, an algebraic curve $\mathbf{k}$ of degree 3 can be described by a homogeneous polynomial

$$P_{\mathbf{k}} = \sum_{\substack{i,j,k \in \mathbb{N} \\ i+j+k=3}} c_{i,j,k}\, X^i Y^j Z^k.$$

Analogously, we define the algebraic operation

$$\mathrm{MEETKL} := \Big\{ \big(\mathbf{k}, [l], [p]\big) \in \mathcal{K} \times \mathcal{L} \times \mathcal{P} \;\big|\; P_{\mathbf{k}}(p_1, p_2, p_3) = 0 \text{ and } l^T p = 0 \Big\}.$$

The operation MEETKL outputs the points of intersection $[p]$ of an algebraic curve $\mathbf{k}$ of degree 3 and a line $[l]$.

Example 2.7 (circle through three points). Let $p_4 := (-i, 1, 0)^T$ and $p_5 := (i, 1, 0)^T$. The corresponding algebraic operation that provides a circle through three given points is described by the relation

$$\mathrm{CIRCLE} := \Big\{ \big([p_1], [p_2], [p_3], \mathbf{c}\big) \in \mathcal{P}^3 \times \mathcal{C} \;\big|\; P_{\mathbf{c}}\big(p_{j,1},\, p_{j,2},\, p_{j,3}\big) = 0 \text{ for all } j \in \{1, \ldots, 5\} \Big\},$$

where $p_{j,k}$ denotes the k-th component of the vector p_j for $j \in \{1, \ldots, 5\}$ and $k \in \{1, 2, 3\}$. The points $[p_4]$ and $[p_5]$ are essential for circles, since one can show that a conic is a circle, if and only if the points $[p_4]$ and $[p_5]$ lie on this conic. So we denote them by $\mathtt{I} := p_4$ and $\mathtt{J} := p_5$.

However, there is a question left: What happens, if at least three of the five points are collinear? We start with three collinear points, say p_1, p_2, p_3. Since CIRCLE provides a conic $\mathbf{c}$ through the five points $p_1, \ldots, p_5$, the conic $\mathbf{c}$ splits up into the two lines joining p_1, p_2, p_3 and p_4, p_5. In case of four collinear points, say $p_2, \ldots, p_5$, $\mathbf{c}$ splits up into two lines again. One of them is the line joining $p_2, \ldots, p_5$ and the other one is an arbitrary line through p_1. The last case to be considered depends to five collinear points. Here, $\mathbf{c}$ becomes the line through all five points.

Until now, we have only stated geometric operations that are algebraic. However, an example of a non-algebraic operation is the following one.

Example 2.8 (WHEEL). Let $[p], [q] \in \mathcal{P}$ be two points with representatives $p = (p_1, p_2, p_3)^T$ and $q = (q_1, q_2, q_3)^T$. Then we define the relation WHEEL to be

$$\mathrm{WHEEL} := \Big\{ \big([p], [q]\big) \in \mathcal{P} \times \mathcal{P} \;\big|\; p_1 = e^{q_1}, p_1 \neq 0, q_2 = 0, p_3 = q_3 = 1 \Big\}.$$

First of all, we notice that the relation WHEEL is well-defined, since the equation $p_1 = e^{q_1}$ has a solution, if and only if $p_1 \neq 0$. More precisely, q_1 is a complex logarithm of p_1. Hence the operation WHEEL cannot be described by polynomials. The geometric purpose of this operation is the following one: Given a complex number $z = r \cdot e^{i \cdot \varphi}$ in polar coordinates, this is, a real number $r > 0$ and an angle $\varphi \in]-\pi, \pi]$, then the imaginary part of the logarithm of z equals $i \cdot (\varphi + 2k\pi)$ for some $k \in \mathbb{Z}$. We can consider the imaginary part as a wheel of radius 1 that rolls along the imaginary axis in the complex plane. When the wheel is rotated by φ or $\varphi + 2k\pi$, it trivially covers a distance of φ or $\varphi + 2k\pi$, respectively. So in a way, the operation WHEEL transfers angles to distances.

Up to here, we have formalized certain geometric objects and a possibility to represent their interactions. The next step is to establish a concept of a (static) geometric construction. This will be done by

Definition 2.3 (geometric constructions). *Let $m \in \mathbb{N}$ and $n \in \mathbb{N}_0$. A* geometric construction *$\mathscr{C}$ consists of*

(i) a sequence $O_{1-m}, \ldots, O_0, O_1, \ldots, O_n \in \mathcal{T}$ of geometric types,

(ii) a sequence $\omega_{1-m}, \ldots, \omega_n$ of operations with $\omega_k \subseteq I_1 \times \ldots \times I_{\mathrm{in}(\omega_k)} \times O_k$ and $\mathrm{in}(\omega_k) = 0$ for all $k \in \{1-m, \ldots, 0\}$ (free operations),

(iii) input assignments *$\eta_k = (s_1^k, \ldots, s_{\mathrm{in}(\omega_k)}^k) \in \mathbb{Z}^{\mathrm{in}(\omega_k)}$ $(k \in \{1, \ldots, n\})$ such that $s_l^k < k$ and $O_{s_l^k} = I_l$, for $l \in \{1, \ldots, \mathrm{in}(\omega_k)\}$.*

A construction is called algebraic*, if $\omega_1, \ldots, \omega_n$ are algebraic operations.*

Condition (*i*) and (*ii*) specify the types of geometric objects that are used for the construction $\mathscr{C}$ and their already mentioned interplay, respectively; condition (*iii*) ensures two things: Firstly, the required types of the input objects and the output object of each operation have to be appropriate and secondly, only free or already constructed objects are used as input for an operation.

Remark 2.2. We observe that our concept of algebraic construction is reasonable, since it captures compass and ruler constructions in the sense of [36]. There, compass and ruler constructions are formalized. Roughly speaking, they consist of a finite sequence of construction steps that start from the two fixed points $(0,0)^T$ and $(1,0)^T$. In addition, each construction steps has to be one of the following: The join of two points, the intersection of two lines, the intersection of a circle and a line, the intersection of two circles and a step that provides a circle to a given center and a further point on the circle.

We can recreate all of these steps in our model. Obviously, there are points representing $(0,0)^T$ and $(1,0)^T$. Furthermore, we have already introduced the algebraic operations JOIN and MEET that cover the join of two points and the intersection of two lines. As mentioned, circles are conics through the points I and J. For this reason, the intersection of a circle and a line and the intersection of two circles is modeled by MEETCL and MEETCC. Lastly, we have to recreate the step that provides a circle to a given center and a further point on the circle. An algebraic construction providing this is stated in Example 2.12.

Now, we would like to exemplify the last definition. For this reason, we introduce a short example of a construction. It yields a point of intersection of two given conics.

Example 2.9. We are not going to state a sequence of geometric types and so on. Instead, we write a construction $\mathscr{C}$ in the following informal way, which should be self-explanatory:

$$\begin{aligned} \mathbf{c}_{-1} &\in \mathcal{C} \\ \mathbf{c}_0 &\in \mathcal{C} \\ \mathbf{p}_1 &\in \mathrm{MEETCC}(\mathbf{c}_{-1}, \mathbf{c}_0) \end{aligned}$$

The notion of a concrete geometric construction that one has usually in mind is not equivalent to a construction in terms of Definition 2.3. Normally, one imagines a construction as a specific picture, where e.g. concrete points and lines have special positions. But this is not our notion of a construction. Our definition is more like an instruction of a construction process rather than a concrete geometric picture. However, the idea of concrete situations is captured by the next definition of *geometric instances*. Ultimately, this definition ensures the consistency of a dynamic geometry system.

Definition 2.4 (geometric instances)**.** *Let $\mathscr{C}$ be a construction as defined in Definition 2.3. Then a* geometric instance *of $\mathscr{C}$ is an element $\vec{\mathbf{o}} = (\mathbf{o}_{1-m}, \ldots, \mathbf{o}_n) \in O_{1-m} \times \ldots \times O_n$ such that*

$$\left(\mathbf{o}_{s_1^k}, \ldots, \mathbf{o}_{s^k_{\mathrm{in}(\omega_k)}}, \mathbf{o}_k\right) \in \omega_k$$

for all $k \in \{1, \ldots, n\}$. By definition, $\mathbf{o}_{1-m}, \ldots, \mathbf{o}_0$ are called free *and $\mathbf{o}_1, \ldots, \mathbf{o}_n$ are called* dependent*. Furthermore, we denote the set of all instances of $\mathscr{C}$ by $\mathcal{I}_{\mathscr{C}}$. We call $\vec{\mathbf{o}}$ a* regular instance *of $\mathscr{C}$, if for each $k \in \{1, \ldots, n\}$*

$$\left|\omega_k\left(\mathbf{o}_{s_1^k}, \ldots, \mathbf{o}_{s^k_{\mathrm{in}(\omega_k)}}\right)\right| = B_k,$$

where B_k is the product of degrees of the polynomials of the underlying polynomial system associated to ω_k.

Remark 2.3. Note that defining the regularity of an instance by the use of the number $B_k \in \mathbb{N}$ is well-defined according to BEZOUT's theorem. It does not only state that the set of possible output objects of an operation is not empty, but also the number of these objects, counted with multiplicities, is equal B_k in generic cases.

Example 2.10. Let $\mathscr{C}$ be the construction from Example 2.9. A regular and two non-regular instances of $\mathscr{C}$ are shown in Figure 2.6. There are two qualitatively different types of non-regular instances. On the one hand, the number of output objects of an operation can be less than B_k (shown in the lower left drawing). Or on the other hand, there can be infinitely many output objects. For instance, when objects coincide as shown in the lower right drawing. Last but not least, we would like to state the corresponding formal objects $\vec{\mathbf{o}} = (\mathbf{o}_{-1}, \mathbf{o}_0, \mathbf{o}_1)^T$ for the instance shown in Fig. 2.6. The regular instance is given by

$$\mathbf{o}_{-1} := \left[(1,1,-1,0,0,0)^T\right], \quad \mathbf{o}_0 := \left[(1,1,0,0,-2,0)^T\right], \quad \mathbf{o}_1 := \left[\left(\frac{1}{2}, \frac{\sqrt{3}}{2}, 1\right)^T\right]$$

and the two non-regular instances by

$$\mathbf{o}_{-1} := \left[(1,1,-1,0,0,0)^T\right], \quad \mathbf{o}_0 := \left[(1,1,3,0,-4,0)^T\right], \quad \mathbf{o}_1 := \left[(1,0,1)^T\right]$$

and

$$\mathbf{o}_{-1} := \left[(1,1,-1,0,0,0)^T\right], \quad \mathbf{o}_0 := \left[(1,1,-1,0,0,0)^T\right], \quad \mathbf{o}_1 := \left[\left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, 1\right)^T\right].$$

2.2.2 Movements and Dynamic Geometry Systems

So far, we have just modeled geometric constructions that are static. Dynamics has not yet been included. However, we have to extend our formal model, since we are interested in dynamic geometry. As a first step, continuous motions of the free objects have to be formalized. As to do so, we need a topology on $\mathbb{CP}^d$. Since $\mathbb{CP}^d$ is a complex manifold of dimension $d \in \mathbb{N}$, we have got a natural topology on it. Furthermore, if X is a topological space, one can speak of continuous mappings from X to $\mathbb{CP}^d$ in a natural way. Holomorphic maps can be studied as well.

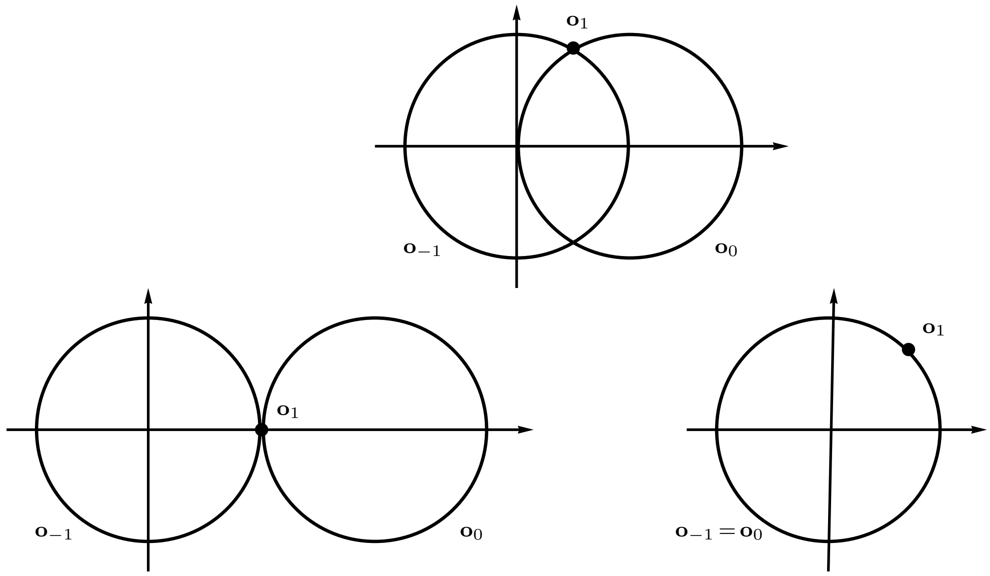

Figure 2.6: A regular (upper row) and two non-regular (lower row) instances of $\mathscr{C}$.

Remark 2.4. Let X be an open subset of $\mathbb{C}$. By definition, a mapping $\varphi : X \to \mathbb{CP}^d$ is called holomorphic at a point $x \in X$, if $\kappa \circ \varphi : X \to \mathbb{C}^{d+1}$ is holomorphic (component-wise) at x, where κ is an adequate chart of $\mathbb{CP}^d$.

After we have clarified the meaning of continuity and holomorphy with respect to $\mathbb{CP}^d$, we are ready to introduce our concept of a continuously moving element.

Definition 2.5 (h-movements)**.** *Given a continuous mapping $\mu : [0,1] \to M$, where $M \in \{\mathcal{P}, \mathcal{L}, \mathcal{C}, \mathcal{K}\}$. Then, we call μ an* h-movement*, if there exists a holomorphic map $\varphi : U \to M$, where $U \subseteq \mathbb{C}$ is an open neighborhood of $[0,1]$ such that $\varphi\big|_{[0,1]} \equiv \mu$ holds.*

Furthermore, let $\mathcal{O}(U)$ be the set of all holomorphic functions defined on an open set $U \subseteq \mathbb{C}$ and we define for every $\varepsilon > 0$ the sets $[0,1]_\varepsilon := \{z \in \mathbb{C} \mid \exists\, w \in [0,1] : \|z - w\| < \varepsilon\}$. Then if the supremum

$$\sup\left\{\varepsilon \in \mathbb{R} \;\middle|\; \exists\, \varphi \in \mathcal{O}\big([0,1]_\varepsilon\big) :\ \varphi\big|_{[0,1]} \equiv \mu\right\}$$

exists, we denote it by ε^* and set $U_\mu := [0,1]_{\varepsilon^*}$. If it does not exist, we define U_μ to be $\mathbb{C}$.

Remark 2.5. We emphasize that an h-movement always corresponds to a holomorphic function. For this reason, an h-movement induces a continuous motion of a free object in the underlying space M.

An illustration of the sets $[0,1]_\varepsilon$ are given in Fig. 2.7. They are tubes in the complex plane that contain the interval $[0,1]$. We should also mention the sets $[0,1]_\varepsilon$ are well-defined for all $\varepsilon > 0$, since they are not empty. As every non-empty subset of the real numbers with an upper bound has a supremum, the set U_μ is well-defined as well. We can think of U_μ as the maximal ε-tube that admits a holomorphic function that describes μ. Besides, it holds that if there exists an h-movement μ and two holomorphic functions $\varphi_k : U_k \to M$ so that $\varphi_k\big|_{[0,1]} \equiv \mu$ on adequate neighborhoods U_k of $[0,1]$ ($k \in \{1,2\}$), then it follows by the identity theorem for holomorphic functions that $\varphi_1 \equiv \varphi_2$ on $U_1 \cap U_2$.

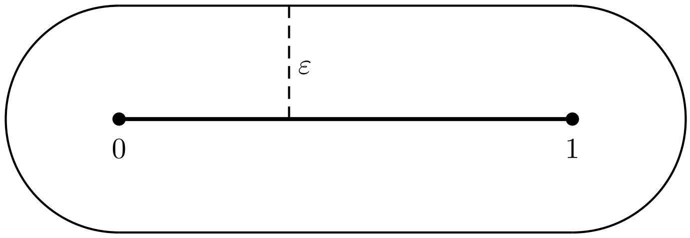

Figure 2.7: An ε-tube $[0,1]_\varepsilon$ of the interval $[0,1]$.

Now, all things are developed to define the last part of our formal model. The next definition captures both dynamic geometry and continuous dynamic geometry.

Definition 2.6 (dynamic geometry systems)**.** *Given a construction $\mathscr{C}$ with $m \in \mathbb{N}$ free objects and an instance $\vec{\mathbf{o}} \in \mathcal{I}_{\mathscr{C}}$. Let Φ be the set of all functions $\mu : [0,1] \to O_{1-m} \times \ldots \times O_0$ with $\mu_j(0) = \mathbf{o}_j$ and μ_j is an h-movement for all $j \in \{1-m, \ldots, 0\}$. Then a function $\mathcal{D} : \Phi \to \Psi$, where Ψ is the set of all functions $\nu : [0,1] \to O_1 \times \ldots \times O_n$ with $\left(\mu(t)^T, \nu(t)^T\right)^T \in \mathcal{I}_{\mathscr{C}}$ for all $t \in [0,1]$, is called a* dynamic geometry system (DGS)*.*

In addition, if for all $j \in \{1, \ldots, n\}$, the j-th component $\nu_j(\mu)$ of $\mathcal{D}(\mu)$ is a continuous mapping for every $\mu \in \Phi$, we call $\mathcal{D}$ a continuous dynamic geometry system (CDGS)*.*

At first, we observe that the motion of a free object must be described by a holomorphic function, since we only admit h-movements. Starting from a construction $\mathscr{C}$ and a given instance $\vec{\mathbf{o}} \in \mathcal{I}_{\mathscr{C}}$, a DGS maps h-movements of the free objects that are encapsulated in μ to a non-constant vector ν of geometric objects in general. The several functions in ν describe the motion of the dependent objects during the motions of the free objects. Thereby, it is claimed that the free as well as the dependent objects form an instance of the underlying construction $\mathscr{C}$ at each point in time. Although the motion of the free objects is continuous, the motion of the dependent objects has not to be automatically continuous. This feature must be claimed in addition.

Remark 2.6. A CDGS in our sense has to be continuous, consistent and deterministic. Continuity is obvious; consistency follows because of the fact that only functions forming an instance are output by a DGS; determinism is obtained, since ν only consists of functions.

Remark 2.7. We have not yet mentioned semi-free objects like a point on a line. However, they are also included in our framework. We have to realize them by a sub-CDGS. For instance, a semi-free point on a line can be implemented by two intersecting lines, where one of them is the line, on which the semi-free point shall move and the other one is needed to generate the movement. While the line, on which the point moves stays constant, the movement of the additional line is chosen properly to generate the desired movement.

From Definition 2.6, we can derive a first proposition stating that a DGS does always exist in our algebraic setting.

Proposition 2.1. *Given an algebraic construction $\mathscr{C}$ and an instance $\vec{\mathbf{o}} \in \mathcal{I}_{\mathscr{C}}$. Then there exists a DGS.*

Proof. Follows directly from Remark 2.1. $\square$

2.3 Required Tools

In the present section, we will derive the necessary technical lemmata in order to prove among others the existence CDGSs. To do so, we divide this section into two parts. In the first one, we will deal with the starting situation. In this case, we can show that in the generic case, the initial behavior of all objects of an algebraic construction corresponding to a set of given h-movements of the corresponding free objects can be locally described by holomorphic coordinate functions. The second subsection studies the properties of these coordinate functions. Thereby, it will turn out that analytic continuations of the derived holomorphic coordinate functions are the key to construct an associated CDGS. For this reason, we are going to study the corresponding RIEMANN surfaces. Moreover, a continuation theorem will be proved. It deals with singularities of coordinate functions and the case, when a representative becomes the all-zero vector.

2.3.1 Local Holomorphic Coordinates

The goal of this subsection is to prove a lemma, which informally states that the initial behavior of all objects of a regular instance of a corresponding algebraic construction can be locally described by holomorphic coordinate functions. To prove this, we have to tackle some technical issues, which is done by the next three lemmata.

The first one of the three lemmata shows that a finite number of elements of $\mathbb{CP}^d$ can always be transformed so that the underlying representatives of these elements do not lie in a coordinate plane of the vector space $\mathbb{C}^{d+1}$.

Lemma 2.1. *Let $n, d \in \mathbb{N}$ and $\mathbf{o}_1, \ldots, \mathbf{o}_n \in \mathbb{CP}^d$. Then there exists a projective transformation $M \in \mathbb{C}^{(d+1)\times(d+1)}$ so that no representative $v_k := M \cdot o_k$ of the transformed object $\mathbf{M}\cdot\mathbf{o}_k$ has a vanishing coordinate. In symbols, this means that $v_{k,l} \neq 0$ for all $k \in \{1, \ldots, n\}$ and $l \in \{1, \ldots, d+1\}$, where $v_{k,l}$ denotes the l-th coordinate of the representative v_k.*

Proof. Let us assume that the l-th entry of the vector o_k equals zero, this is, $o_{k,l} = 0$. Since $\mathbf{o}_k \in \mathbb{CP}^d$, there is a non-vanishing coordinate of o_k, say $o_{k,j}$. Now, we set $M := I + \varepsilon_{l,j}$, where I is the identity matrix and $\varepsilon_{l,j} \in \mathbb{C}^{(d+1)\times(d+1)}$ is the zero matrix besides the (l, j)-th entry that equals $\varepsilon \in \mathbb{R}$. It follows that $v_{k,l} \neq 0$, if and only if $\varepsilon \neq 0$. For continuity reasons, the entry ε can be chosen so that $v_{k,l} \neq 0$ and $v_{i,j} \neq 0$, if $o_{i,j} \neq 0$ for all $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, d+1\}$. As a result, the claims follow inductively. □

The next lemma states that we can always choose an affine plane for dehomogenizing these elements so that none of them lies at infinity according to this dehomogenization. In addition, the dehomogenization can be chosen so that all k-th coordinates of all underlying objects are pairwise different. More precisely, we get

Lemma 2.2. *Let $n, d \in \mathbb{N}$ and $\mathbf{o}_1, \ldots, \mathbf{o}_n \in \mathbb{CP}^d$ pairwise different, and no representative has a vanishing coordinate, i.e. $o_{k,j} \neq 0$ for all $k \in \{1, \ldots, n\}$ and all $j \in \{1, \ldots, d+1\}$. Then there exist $a_1, \ldots, a_{d+1} \in \mathbb{C}$ and $b \in \mathbb{C} \setminus \{0\}$, inducing an affine plane*

$$H := \left\{(x_1, \ldots, x_{d+1}) \in \mathbb{C}^{d+1} \;\middle|\; a_1 x_1 + \ldots + a_{d+1} x_{d+1} = b\right\},$$

so that there exist representatives $o_k \in H$ for $k \in \{1, \ldots, n\}$. Furthermore, H can be chosen so that $o_{k,j} \neq o_{l,j}$ holds for all $k, l \in \{1, \ldots, n\}$ and all $j \in \{1, \ldots, d+1\}$.

Proof. For continuity reasons, it holds that H can be chosen in such a way, that $H \cap \mathbf{o}_k \neq \emptyset$ for all $k \in \{1, \dots, n\}$. Now, we prove the rest of the statement. Therefore, let us assume that there exist two representatives with equal coordinates, i.e. there are $k, l \in \{1, \dots, n\}$ and an $r \in \{1, \dots, d+1\}$ with $o_{k,r} = o_{l,r}$. Since $\mathbf{o}_k \neq \mathbf{o}_l$, there exists an index $j \in \{1, \dots, d+1\} \setminus \{r\}$ with $o_{k,j} \neq o_{l,j}$. We now construct another plane H', which has the desired property. Therefore, we replace the coefficient a_j of H by $a_j + \varepsilon$ to define a plane H'. Hereby, we can choose $\varepsilon > 0$ small enough so that H' still contains representatives $o'_1, \dots, o'_n$ of the objects $\mathbf{o}_1, \dots, \mathbf{o}_n$ and no further coordinates of them coincide. In other words, it holds that $H' \cap \mathbf{o}_k \neq \emptyset$ for all $k \in \{1, \dots, n\}$ and that $o'_{k,s} \neq o'_{l,s}$, if $o_{k,s} \neq o_{l,s}$. As o_k and o'_k are representatives of the element $\mathbf{o}_k \in \mathbb{CP}^d$, there is a non-vanishing factor λ_ε so that $o'_k = \lambda_\varepsilon \cdot o_k$. This factor depends continuously on ε and is given by

$$\lambda_\varepsilon = \frac{b}{a_1 \cdot o_{k,1} + \ldots + (a_j + \varepsilon) \cdot o_{k,j} + \ldots + a_{d+1} \cdot o_{k,d+1}}.$$

Analogously, there is a factor

$$\mu_\varepsilon = \frac{b}{a_1 \cdot o_{l,1} + \ldots + (a_j + \varepsilon) \cdot o_{l,j} + \ldots + a_{d+1} \cdot o_{l,d+1}},$$

that satisfies $o'_l = \mu_\varepsilon \cdot o_l$. It holds that $\lambda_\varepsilon \neq \mu_\varepsilon$ for all ε in an adequate neighborhood of $0 \in \mathbb{R}$ and therefore $o'_{k,r} \neq o'_{l,r}$. Finally, the continuity of λ_ε and μ_ε ensures that the previous construction decreases the finite number of equal coordinates. $\square$

Each representative of a dependent object of a regular instance has to satisfy a corresponding multi-homogeneous polynomial system that is specified by the associated construction. Using some basic resultant techniques enables us to reduce each such polynomial system to a set of univariate polynomials that describes the same solutions. Formally, we get

Lemma 2.3 (resultants). *Let $f_1, \dots, f_n \in \mathbb{C}[X_1, \dots, X_n]$. Then for each X_i, there exists a polynomial $p_i \in \mathbb{C}[X_i]$, whose roots are the x_i-coordinates of the solutions of $f_1 = \ldots = f_n = 0$.*

Proof. Follows by [10, p. 116, Proposition (5.15)]. $\square$

We are not so much interested in the result of the previous lemma as in the derivation of the mentioned univariate polynomials, since we are going to refer to the same construction in the next proof.

With these three lemmata in mind, we are now able to prove that the initial behavior of all objects corresponding to a given algebraic construction and an associated regular instance can be locally described by holomorphic coordinate functions.

Lemma 2.4 (local holomorphic coordinates). *Given an algebraic construction $\mathscr{C}$ and a regular instance $\vec{\mathbf{o}} = (\mathbf{o}_{1-m}, \dots, \mathbf{o}_n)^T \in \mathcal{I}_{\mathscr{C}}$. Furthermore, let for all $k \in \{1-m, \dots, 0\}$, $\mu_k : [0,1] \to \mathbb{CP}^{d_k}$ be an h-movement so that $\mu_k(0) = \mathbf{o}_k$. Then, for each object $\mathbf{o}_k \in \mathbb{CP}^{d_k}$, $k \in \{1-m, \dots, n\}$, there exist holomorphic coordinate functions*

$$c_{k,1}(t), \dots, c_{k,d_k+1}(t) : V \to \mathbb{C}$$

defined on an appropriate open neighborhood $V \subseteq \bigcap_{k \in \{1-m,\dots,0\}} U_{\mu_k}$ of $0 \in \mathbb{C}$ with $\mathbf{o}_k(0) = \mathbf{o}_k$ and $\left(\mathbf{o}_{1-m}(t), \dots, \mathbf{o}_n(t)\right)^T \in \mathcal{I}_{\mathscr{C}}$ for all $t \in V$, where $\mathbf{o}_k(t) := \left[\left(c_{k,1}(t), \dots, c_{k,d_k+1}(t)\right)^T\right]$.

Proof. We prove the lemma by induction over the length $k = m + n$ of the construction $\mathscr{C}$. For $k = 1$, $\mathscr{C}$ only consists of one free operation ω_0. Then the h-movement μ_0 induces the claimed holomorphic coordinate functions by definition.

Now, let $k > 1$. If the last operation ω_n of $\mathscr{C}$ is a free operation, there is nothing to show (only free operations, analogous argument as for $k = 1$). Otherwise, ω_n is an algebraic operation and for each operation ω_l, $l \in \{1 - m, \ldots, n - 1\}$, there exist the claimed holomorphic coordinate functions defined on an appropriate open neighborhood U of $0 \in \mathbb{C}$. By substituting every formal unknown

$$X_{1,1}, \ldots, X_{\mathrm{in}(\omega_n), d_{s^n_{\mathrm{in}(\omega_n)}}+1}$$

associated to the input signature of ω_n in its describing polynomial system P_{ω_n}, we get a polynomial system $P(t) \subseteq \mathcal{O}(U)[X_1, \ldots, X_{d_n+1}]$ with holomorphic coefficients dependent on t. Since $\vec{\mathbf{o}}$ is a regular instance, BEZOUT's theorem states that the number of one-dimensional solution spaces of $P(t_0)$ is B_n and each of them has multiplicity one. Due to the fact that the solutions of $P(t)$ depend continuously on t, there is an open neighborhood $W \subset U$ of $0 \in \mathbb{C}$, on which $P(t)$ still has B_n simple solution space. From Lemma 2.1, we conclude that there exists a projective transformation T of $\mathbb{CP}^{d_n}$ so that none of the transformed solutions of the system $P(t)$ lies in a coordinate plane of $\mathbb{C}^{d_n+1}$. This transformation also indues a transformed multi-homogenous polynomial system $\tilde{P}(t)$, whose solutions are the transformed solutions of $P(t)$. Thereby, we can choose one fixed projective transformation T for all $t \in W$. For this reason the coefficients of $\tilde{P}(t)$ are holomorphic functions on W as well.

Now, an affine equation is added to the polynomial system $\tilde{P}(t)$. Then the solutions $x_1(t), \ldots, x_{B_n}(t)$ of this extended affine polynomial system are special representatives of the former one-dimensional solution spaces, which also depends continuously on $t \in W$. Lemma 2.2 guarantees the existence of an affine equation so that any two of the solutions $x_1(0), \ldots, x_{B_n}(0)$ do not have an equal coordinate. In addition, we can also claim that $x_1(t), \ldots, x_{B_n}(t)$ do not have an equal coordinate for all $t \in V$ due to their continuity.

Since the number of variables and equations of the extended polynomial system are equal[7], we can construct polynomials $p_1(t, X), \ldots, p_{d_n+1}(t, X) \in \mathcal{O}(V)[X]$ with holomorphic coefficients defined on W for each unknown $X_1, \ldots, X_{d_n+1}$ (resultants according to Lemma 2.3). For $l \in \{1, \ldots d_n + 1\}$, the roots of $p_l(t, X)$ are the l-th coordinates of the solutions $x_1(t), \ldots, x_{B_n}(t)$. Since the degree of $p_l(t, X)$ is equal to B_n, it follows that $p_l(t, X)$ only has roots of multiplicity one. By using [34, p. 9, 1.2.4], there exist holomorphic coordinate functions $\tilde{c}_1, \ldots, \tilde{c}_{d_n+1} : V \to \mathbb{C}$, where $V \subset W$ is an adequate neighborhood of $0 \in \mathbb{C}$. Applying the inverse transformation T^{-1} on $\left(\tilde{c}_1(t), \ldots, \tilde{c}_{d_n+1}(t)\right)^T$ yields holomorphic functions $c_1, \ldots, c_{d_n+1} : V \to \mathbb{C}$ so that

$$[o_n(0)] = \mathbf{o}_n \quad \text{and} \quad \left(\mathbf{o}_1(t), \ldots, \mathbf{o}_n(t)\right)^T \in \mathcal{I}_{\mathscr{C}} \quad \text{for all } t \in V,$$

if $o_n(t) := \left(c_1(t), \ldots, c_{d_n+1}(t)\right)^T$. □

For further use, let us fix an adapted notation that reflects the situation motivated by Lemma 2.4 throughout the rest of this chapter. In detail, the mentioned lemma guarantees the existence of holomorphic (coordinate) functions $c_{1-M}, \ldots c_0, c_1, \ldots, c_N$, defined on an

[7] due to ω_n is algebraic.

appropriate neighborhood V of $0 \in \mathbb{C}$, where

$$M = \sum_{k=1-m}^{0} (d_k + 1) \quad \text{and} \quad N = \sum_{k=1}^{n} (d_k + 1).$$

According to the proof of this lemma, it holds that for each $l \in \{1, \ldots, N\}$, there exists a polynomial

$$P_l \in \mathbb{C}[C_{1-M}, \ldots, C_{l-1}, X]$$

so that $P_l\big(c_{1-M}(t), \ldots, c_{l-1}(t), c_l(t)\big) = 0$ for all $t \in V$. We will see that the analytic continuation of the coordinate functions $c_{1-M}, \ldots c_0, c_1, \ldots, c_N$ will be a main ingredient to construct a CDGS related to an algebraic construction and a corresponding regular starting instance. In order to deal with the field of analytic continuation, we therefore introduce some helpful apparatus at the beginning of the next subsection.

2.3.2 Global Viewpoint on Local Coordinates

Until now, we have shown that the initial behavior of all objects of an algebraic construction corresponding to a set of given h-movements of the corresponding free objects can be locally described by holomorphic coordinate functions $c_{1-M}, \ldots, c_N$. Starting from this situation, it will turn out that we can construct a corresponding CDGS by continuing the coordinate functions $c_{1-M}, \ldots, c_N$ of the objects analytically. This construction works as long as all analytic continuations of $c_{1-M}, \ldots, c_N$ along a path exist. For this reason, we are going to study their analytic continuations in this subsection.

Problems occur only in two cases – either one of the analytic continuations may not exist or all coordinate functions of an object become the all-zero vector. To deal with the problem concerning analytic continuations, we are going to characterize *critical paths* and *points* first (see Definition 2.8). Along a non-critical path, we can show that the analytic continuations of $c_{1-M}, \ldots, c_N$ exist and thus we can use them to construct a CDGS (see Lemma 2.6 and 2.7). In cases, where the analytic continuations may not exist, i.e. at isolated critical points like singularities or branch points, we can prove that LAURENT series and so-called uniformizing variables describe the behavior of $c_{1-M}, \ldots, c_N$ in a neighborhood of such points (see Lemma 2.8). This will enable us to construct a CDGS except for a finite number of such critical points. Lastly, we are going to prove a continuation lemma that closes the remaining gaps at the critical points (see Lemma 2.9). This lemma will also captures the cases, when all coordinate functions of an object become the all-zero vector. But prior to this, we fix the necessary function theory.

Excursion to Riemann Surfaces

A standard approach to deal with local properties of holomorphic functions is done by studying their power series expansions. A formal way of doing this is given by

Definition 2.7 (germs). *Let $z_0, \alpha_j \in \mathbb{C}$, where $j \in \mathbb{N}_0$. We denote*

$$g := (z_0; \alpha_0, \alpha_1, \ldots)$$

as a germ *with* base $\underset{\sim}{g} := z_0$ *and* top $\widehat{g} := \alpha_0$*, if*

$$\sum_{k=0}^{\infty} \alpha_k (z - z_0)^k$$

is a complex power series with positive radius of convergence. In this case let

$$g[z] := \sum_{k=0}^{\infty} \alpha_k (z - z_0)^k,$$

for all z sufficiently close to $\underline{g}$. Furthermore let $\mathbb{O}_{z_0}$ be the set of all germs with base $z_0 \in \mathbb{C}$ and

$$\mathbb{O}_U := \bigcup_{z \in U} \mathbb{O}_z,$$

for every open set $U \subseteq \mathbb{C}$.

Remark 2.8. Let $U \subseteq \mathbb{C}$ be an open set. A germ $g \in \mathbb{O}_U$ can also be identified with a unique determined holomorphic function $g : D \to \mathbb{C}$ defined on an appropriate open disc $D \subseteq U$. Since it should become clear by the context, whether the function or the germ is meant, we will denote both variants, the germ as well as the function, by the same symbol g.

In the following, let $U \subseteq \mathbb{C}$ be a fixed open set. In this situation, one has got the canonical projection

$$\pi_U : \mathbb{O}_U \to U, \ g \mapsto \underline{g}$$

that maps a germ to its base. If $f \in \mathcal{O}(V)$, where $V \subseteq U$ is open, there exists for every $z \in V$ a corresponding germ denoted by $\rho_z(f)$. In addition, each holomorphic map $f \in \mathcal{O}(V)$, where $V \subseteq U$ is open, induces the set of germs

$$\sigma(f, V) := \{\rho_z(f) \mid z \in V\} \subseteq \mathbb{O}_U.$$

All these sets induce, in turn, a special topology on $\mathbb{O}_U$.[8]

Lemma 2.5. *There is exactly one topology on $\mathbb{O}_U$ such that π_U is local topological. Moreover, $\mathbb{O}_U$ becomes a* Hausdorff *space.*

Proof. We regard the topology $\mathfrak{T}$ induced by the basis

$$\mathfrak{B} := \{\sigma(f, V) \mid V \subseteq U \text{ open and } f \in \mathcal{O}(V)\}.$$

Due to this fact $\mathfrak{T}$ consists of all sets $M \subseteq \mathbb{O}_U$, where M is obtained as a union of sets from the basis $\mathfrak{B}$. The only fact we have to prove in order to show that $\mathfrak{T}$ is a topology on $\mathbb{O}_U$ is that the intersection of two sets $\sigma(f, V), \sigma(h, W) \in \mathfrak{B}$ is in $\mathfrak{T}$. To do so, let $g \in \sigma(f, V) \cap \sigma(h, W)$.[9] Then there is a $z_0 \in V \cap W$ and a neighborhood $N \subseteq V \cap W$ of z_0 so that $f(z) = g[z] = h(z)$ for all $z \in N$. As a result, $g \in \sigma(f\big|_N, N)$ implying that $\sigma(f, V) \cap \sigma(h, W)$ is open. That the map π_U is a local topological map is a consequence of the definition of the basis $\mathfrak{B}$. To complete the proof, we have to show that $\mathbb{O}_U$ is a Hausdorff space. For this reason, let $f, g \in \mathbb{O}_U$. The case of $\underline{f} \neq \underline{g}$ is trivial. If $\underline{f} = \underline{g}$, there exists an open disc $D \subseteq U$ centered at $\underline{f}$ so that f and g induce holomorphic functions $\tilde{f}, \tilde{g} \in \mathcal{O}(D)$. In the case of

$$\underbrace{\sigma(\tilde{f}, D)}_{f \in} \cap \underbrace{\sigma(\tilde{g}, D)}_{g \in} = \emptyset,$$

we are done, since this is the definition of the Hausdorff axiom. And otherwise, $f = g$ follows by the identity theorem. $\square$

[8] The reader, who is not familiar with the required topological theory is referred to [26].

[9] If this intersection is empty, we have nothing to show, as $\emptyset \in \mathfrak{T}$.

In addition, one can even prove that $\mathbb{O}_U$ is not only a HAUSDORFF space, but also a RIEMANN surface.

Theorem 2.1. *There exists exactly one holomorphic structure on $\mathbb{O}_U$ so that $\mathbb{O}_U$ becomes a* RIEMANN *surface. Furthermore, π_U is a local biholomorphic map according to this holomorphic structure.*

Proof. Follows by [34, p. 7, 1.2.1]. □

The following theorem links the concept of analytic continuation to the abstract HAUSDORFF space $\mathbb{O}_U$. Informally, an analytic continuation of a holomorphic function corresponds to a path in $\mathbb{O}_U$ and vice versa.

Theorem 2.2. *Let $\tilde{\gamma} : [0,1] \to \mathbb{O}_U$ be a path. Then $\tilde{\gamma}(1)$ is obtained by analytic continuation of $\tilde{\gamma}(0)$ along the path*

$$\gamma : [0,1] \to U;\ t \mapsto \pi_U\big(\tilde{\gamma}(t)\big).$$

Conversely, let $\gamma : [0,1] \to U$ be a path, $f : G \to U$ a holomorphic map defined on a region $G \subseteq U$ containing $\gamma(0)$ and $f^\gamma : R \to U$ the analytic continuation of f along γ, where $R \subseteq U$ is a region that contains $\gamma(1)$. Then it holds that the germs corresponding to f and f^γ, this is, $\rho_{\gamma(0)}(f)$ and $\rho_{\gamma(1)}\big(f^\gamma\big)$, are in the same path-component of $\mathbb{O}_U$.

Proof. For each $t \in [0,1]$, the germ $\tilde{\gamma}(t)$ is defined on an associated open disc $D_t \subset U$. The collection

$$\big\{\sigma(\tilde{\gamma}(t), D_t) \subset \mathbb{O}_U \bigm| t \in [0,1]\big\}$$

is an open cover of $\tilde{\gamma}([0,1])$. As $\tilde{\gamma}$ is a continuous map, $\tilde{\gamma}([0,1])$ is compact in $\mathbb{O}_U$. Therefore, there is a finite cover of open sets

$$\sigma\big(\tilde{\gamma}(t_0), D_0\big), \ldots, \sigma\big(\tilde{\gamma}(t_r), D_r\big) \subset \mathbb{O}_U$$

still covering $\tilde{\gamma}([0,1])$. These sets induce a decomposition $0 = t_0 < \ldots < t_r = 1$ of $[0,1]$ and associated open discs $D_0, \ldots, D_r \subset U$ covering $[0,1]$. In addition, D_j is centered at t_j and the germ $\tilde{\gamma}(t_j)$ is defined on D_j for all $j \in \{0, \ldots, r\}$. Moreover, it holds $\sigma\big(\tilde{\gamma}(t_j), D_j\big) \cap \sigma\big(\tilde{\gamma}(t_{j+1}), D_{j+1}\big) \neq \emptyset$ for all $j \in \{0, \ldots, r-1\}$. Now, the identity theorem provides that $\tilde{\gamma}(1)$ is yielded by analytic continuation of $\tilde{\gamma}(0)$ along the path γ.

Conversely, regard a finite chain of open discs $D_0, \ldots, D_r \subset U$ induced by the analytic continuation of f along γ. It is obvious that they cover $[0,1]$. The corresponding holomorphic functions $f\big|_{D_0} \equiv f_0, \ldots, f_r \equiv f^\gamma\big|_{D_r}$ induce open sets $\sigma(f_0, D_0), \ldots \sigma(f_r, D_r) \subset \mathbb{O}_U$ so that $\sigma(f_j, D_j) \cap \sigma(f_{j+1}, D_{j+1}) \neq \emptyset$ for all $j \in \{0, \ldots, r-1\}$. As these sets are path-connected, there exists a path $\tilde{\gamma} : [0,1] \to \mathbb{O}_U$ connecting $\rho_{\gamma(0)}(f)$ and $\rho_{\gamma(1)}\big(f^\gamma\big)$. □

As an immediate consequence, one obtains

Corollary 2.1. *The path-components of $\mathbb{O}_U$ are* RIEMANN *surfaces. Furthermore, if $f : G \to \mathbb{C}$ is a holomorphic function, defined on a region $G \subseteq U$ and $z_0 \in G$. Then the path-component $X_U(f)$ containing $\rho_{z_0}(f)$ consists of all possible analytic continuations of f along paths in U.*

Back to Local Coordinates

At this point, we come back to the local holomorphic coordinate functions and continue our observations concerning them. As a first step, we are going to transfer the function theoretic concepts of the last paragraph. They have been constituted in Lemma 2.4 and they correspond to fixed h-movements of the free objects of an underlying algebraic construction $\mathscr{C}$. For $k \in \{1-m, \ldots, 0\}$, we have an h-movement μ_k and an associated open set U_{μ_k}. Roughly speaking, the set U_{μ_k} is the maximal[10] ε-tube, on which a holomorphic function φ can be defined that describes μ_k, i.e. $\varphi|_{[0,1]} \equiv \mu_k$. Throughout the rest of this chapter, we denote their intersection by

$$\mathbb{U} := \bigcap_{k \in \{1-m,\ldots,0\}} U_{\mu_k}.$$

It is clear that $\mathbb{U}$ is open, again. In Lemma 2.4, we have proved the existence of coordinate functions $c_{1-M}, \ldots, c_N$. So for each $l \in \{1-M, \ldots, N\}$, we have a corresponding germ $g_l := \rho_0(c_l)$, which also induces a Riemann surface $X_l := X_{\mathbb{U}}(c_l)$ and a projection $\pi_l := \pi_{\mathbb{U}}\big|_{X_l}$, where the image $\pi_l(X_l) \subseteq \mathbb{U}$ will be denoted by Π_l. If c_l can be continued analytically along a path $\gamma : [0,1] \to \mathbb{U}$, the analytic continuation of c_l induces a germ $g_l^\gamma \in X_l$. We are going to use γ as index only, if the corresponding continuation is possible implying that $\gamma([0,1]) \subseteq \Pi_l$. However, if g_l is not continuable analytically along a path γ, let $t^* \in]0,1]$ be the supremum of all $t \in [0,1]$ so that g_l can be continued analytically along the path

$$\gamma_{t^*} : [0, t^*[\to \mathbb{C}, \ t \mapsto \gamma(t).$$

In this case, we call $\gamma(t^*)$ a *singular point* of g_l. Our following studies are based on the set $\mathbb{U}$. For this reason, we will only consider paths $\gamma : [0,1] \to \mathbb{C}$ with $\gamma\big([0,1]\big) \subseteq \mathbb{U}$ throughout the rest of this chapter.

For each $l \in \{1, \ldots, N\}$, the polynomial $P_l \in \mathbb{C}[C_{1-M}, \ldots, C_{l-1}, X]$ induces a polynomial $P_l \in \mathbb{O}_0[X]$ by replacing C_j by g_j for all $j \in \{1-M, \ldots, l-1\}$. Then it holds that $P_l(g_{1-M}, \ldots, g_{l-1}, g_l)$ is the zero polynomial. Furthermore, let γ be a path so that the analytic continuations $g_{1-M}^\gamma, \ldots, g_l^\gamma$ exist. The identity theorem yields trivially

$$P_l\big(g_{1-M}^\gamma, \ldots, g_{l-1}^\gamma, g_l^\gamma\big) \equiv 0 \tag{2.1}$$

for all paths γ, for which the associated analytic continuations exist.

Remark 2.9. At this point, we would like to emphasize the importance of identity (2.1), since it is the key to construct CDGSs and the reason, why we are studying the analytic continuations of the functions $c_{1-M}, \ldots, c_N$ derived in Lemma 2.4.

As we are interested in constructing CDGSs for an underlying algebraic construction $\mathscr{C}$, we are looking for instances that vary continuously along a family of h-movements. In order to derive an instance, we have to solve several multi-homogenous polynomial systems, namely one for each dependent object. In the proof of Lemma 2.4, we have reduced each of these systems to an adequate set of univariate polynomials. We have denoted these polynomials by $P_{1-M}, \ldots, P_N$. Instead of solving the multi-homogenous polynomials system, we can solve them to obtain an instance of the construction $\mathscr{C}$. By doing so, we have proved that there exist holomorphic functions on an appropriate open neighborhood V of $0 \in \mathbb{C}$ solving these polynomials. As a result, we can use them as coordinate functions of the objects to form instances of $\mathscr{C}$ for all $t \in V$.

[10] We also admit the case $U_{\mu_k} = \mathbb{C}$, if it is possible.

Identity (2.1) tells us that the analytic continuations of theses coordinate functions still satisfy the continued polynomials. By taking this into account, we are enabled to generate continuously varying instance of $\mathscr{C}$ by analytic continuation of these coordinate functions. This works as long as all analytic continuations along a path γ exist and the coordinate functions of one object do not vanish simultaneously. In these cases or when a singular point occurs, we do not get an instance by this concept. But these situations will be captured by Lemma 2.9 and Theorem 2.3, in which we prove that there also exist reasonable instances in this situations.

As already mentioned, there may also be points $z \in \mathbb{U}$, for which the analytic continuation of one of the germs $g_{1-M}, \ldots, g_N$ may not exist. They will play an important role for the ambiguity of a CDGS. However, we can characterize them easily.

Definition 2.8. *Let $P_l(g_{1-M}, \ldots, g_{l-1}, X) = \alpha_k X^k + \cdots + \alpha_0 \in \mathbb{O}_0[X]$ and γ be a path with starting point $0 \in \mathbb{U}$ and terminal point $z \in \mathbb{U}$. Furthermore, we define the sub-path*

$$\gamma_t : [0, t] \to \mathbb{U}, \; s \mapsto \gamma(s).$$

Then we call γ a critical path *of $P_l(g_{1-M}, \ldots, g_{l-1}, X)$, if one of the following conditions hold:*

(i) There exists $t \in]0, 1]$ so that $\gamma(t)$ is a singular point of g_j for some $j \in \{1-M, \ldots, l-1\}$,

(ii) there is $t \in]0, 1]$ so that the analytic continuation $\alpha_k^{\gamma_t}$ of the leading coefficient α_k of P_l along the sub-path γ_t vanishes at the terminal point t, i.e. $\widehat{\alpha_k^{\gamma_t}} = 0$,

(iii) there exists $t \in]0, 1]$ so that $P_l\Big(\widehat{g_{1-M}^{\gamma_t}}, \ldots, \widehat{g_{l-1}^{\gamma_t}}, X\Big) \in \mathbb{C}[X]$ has a zero of multiplicity at least 2.

We call such a point $\gamma(t) \in \mathbb{U}$ a critical point of $P_l(g_{1-M}, \ldots, g_{l-1}, X)$. Conversely, we call γ a non-critical path of $P_l(g_{1-M}, \ldots, g_{l-1}, X)$, if none of these three conditions hold.

Roughly speaking, the critical points are those, in which a polynomial P_l, or more precisely its continuations, is either not defined or has not its maximal number of simple zeros. Condition (i) covers the cases, where one of the coefficients of P_l is not defined; (ii) deals with the cases, where the degree of P_l degenerates; and the cases, where two zeros coincide, are captured by (iii).

The next goal is to show that the analytic continuation of a germ g_l along a non-critical path of $P_l(g_{1-M}, \ldots, g_{l-1}, X)$ exist. For this reason, we prove the following two lemmata. Let γ be a non-critical path of $P_l(g_{1-M}, \ldots, g_{l-1}, X) = \alpha_k X^k + \cdots + \alpha_0 \in \mathbb{O}_0[X]$. Then the polynomial

$$P_l\Big(\widehat{g_{1-M}^{\gamma}}, \ldots, \widehat{g_{l-1}^{\gamma}}, X\Big) \in \mathbb{C}[X]$$

has k distinct, single zeros $x_1, \ldots, x_k \in \mathbb{C}$. Under these assumptions, we can prove that all zeros of the underlying polynomial are described by holomorphic functions on an appropriate neighborhood of $\gamma(1) \in \mathbb{U}$.

Lemma 2.6. *There exists an open neighborhood $V \subseteq \mathbb{U}$ of the terminal point $\gamma(1)$ of γ and k germs $f_1, \ldots, f_k \in \mathbb{O}_{\gamma(1)}$ with the properties:*

(i) $P_l\big(g_{1-M}^{\gamma}, \ldots, g_{l-1}^{\gamma}, f_j\big) \equiv 0$ for all $j \in \{1, \ldots, k\}$,

(ii) $\widehat{f_j} = x_j$ *for all* $j \in \{1, \ldots, k\}$,

(iii) if $P_l\big(g^{\gamma}_{1-M}[z], \ldots, g^{\gamma}_{l-1}[z], x\big) = 0$ *for* $z \in V$*, then* $x = f_j[z]$ *for some* $j \in \{1, \ldots, k\}$.

Remark 2.10. First of all, the lemma ensures the existence of germs $f_1, \ldots, f_k$ that satisfy the continued polynomial $P_l\big(g^{\gamma}_{1-M}, \ldots, g^{\gamma}_{l-1}, X\big)$. Conversely, it states that all zeros are, in turn, described by $f_1, \ldots, f_k$.

Proof. Let us denote $\gamma(1)$ by z_0. Since $x_1, \ldots, x_k$ are simple zeros, there exist non overlapping discs $D_1, \ldots, D_k$, where D_j is centered at x_j ($j \in \{1, \ldots, k\}$). Furthermore, we denote the boundary of a disc D_j by δ_j and the partial derivative of $P_l\left(g^{\gamma}_{1-M}[z_0], \ldots, g^{\gamma}_{l-1}[z_0], X\right)$ w.r.t. X by

$$\partial P_l\left(g^{\gamma}_{1-M}[z_0], \ldots, g^{\gamma}_{l-1}[z_0], X\right).$$

Then $P_l\left(g^{\gamma}_{1-M}[z_0], \ldots, g^{\gamma}_{l-1}[z_0], x\right) \neq 0$ on δ_j and it holds

$$\frac{1}{2\pi i}\int_{\delta_j} \frac{\partial P_l\Big(g^{\gamma}_{1-M}[z_0], \ldots, g^{\gamma}_{l-1}[z_0], x\Big)}{P_l\Big(g^{\gamma}_{1-M}[z_0], \ldots, g^{\gamma}_{l-1}[z_0], x\Big)} dx = 1,$$

where this integral is the winding number of the image of δ_j under $P_l\big(g^{\gamma}_{1-M}[z_0], \ldots, g^{\gamma}_{l-1}[z_0], x\big)$ w.r.t. 0. Replacing z_0 by z yields a continuous function in a neighborhood of z_0. Since they can only take integer values, there exists a neighborhood $V \subseteq \mathbb{U}$ such that

$$\frac{1}{2\pi i}\int_{\delta_j} \frac{\partial P_l\big(g^{\gamma}_{1-M}[z], \ldots, g^{\gamma}_{l-1}[z], x\big)}{P_l\big(g^{\gamma}_{1-M}[z], \ldots, g^{\gamma}_{l-1}[z], x\big)} dx = 1$$

for all $z \in V$ implying that $P_l\big(g^{\gamma}_{1-M}[z], \ldots, g^{\gamma}_{l-1}[z], X\big)$ has exactly one zero in D_j. We denote this zero by $f_j(z)$. By the residue calculus, its value is given by

$$f_j(z) = \frac{1}{2\pi i}\int_{\delta_j} x \cdot \frac{\partial P_l\big(g^{\gamma}_{1-M}[z], \ldots, g^{\gamma}_{l-1}[z], x\big)}{P_l\big(g^{\gamma}_{1-M}[z], \ldots, g^{\gamma}_{l-1}[z], x\big)} dx.$$

Obviously, $f_j(z)$ is analytic on V, thus $f_j(z)$ induces a germ $f_j \in \mathbb{O}_{z_0}$ with $\widehat{f_j} = x_j$. The last thing is to conclude that there cannot be more zeros. But this follows from the fact that we exhibited k zeros and $P_l\big(g^{\gamma}_{1-M}[z], \ldots, g^{\gamma}_{l-1}[z], X\big)$ cannot have more than k zeros. ☐

With this preparatory work, we are able to derive a lemma stating that, along a non-critical path, a corresponding germ can always be continued analytically.

Lemma 2.7. *Let* $l \in \{1, \ldots, N\}$*. If* γ *is a non-critical path of* $P_l(g_{1-M}, \ldots, g_{l-1}, X)$*, then* g^{γ}_l *exists.*

Proof. We define the restricted path $\gamma_t : [0,1] \to \mathbb{U},\ s \mapsto \gamma(s \cdot t)$ for all $t \in [0,1]$. By the prerequisites, all analytic continuations $g^{\gamma}_{1-M}, \ldots, g^{\gamma}_{l-1}$ exist. As a result, all analytic continuations $g^{\gamma_t}_{1-M}, \ldots, g^{\gamma_t}_{l-1}$ exist as well. Since γ is a non-critical path of $P_l(g_{1-M}, \ldots, g_{l-1}, X)$, the polynomial

$$P_l\left(\widehat{g^{\gamma_t}_{1-M}}, \ldots, \widehat{g^{\gamma_t}_{l-1}}, X\right) \in \mathbb{C}[X]$$

only has simple zeros for all $t \in [0,1]$. Due to this fact, Lemma 2.6 states that for each $t \in [0,1]$, there exists an open neighborhood $D^t \subseteq \mathbb{U}$ of $\gamma(t)$ and germs $f^t_1, \ldots, f^t_{\deg(P_l)} \in \mathbb{O}_{\gamma(t)}$. These germs are roots of the polynomial P_l, this is,

$$P_l\left(g^{\gamma_t}_{1-M}, \ldots, g^{\gamma_t}_{l-1}, f^t_j\right) \equiv 0$$

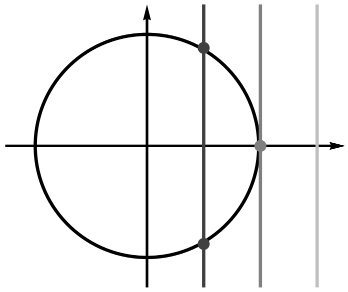

Figure 2.8: Illustrating Example 2.11 for $t = \frac{1}{4}$ (•), $t = \frac{1}{2}$ (•) and $t = \frac{3}{4}$ (•).

on D^t for all $j \in \{1, \ldots, \deg(P_l)\}$. The collection $\{D^t\}_{t\in[0,1]}$ is an open cover of $\gamma([0,1])$. As $\gamma([0,1])$ is compact, there is a finite subcover $\{D^{t_1}, \ldots, D^{t_n}\}$ still covering $\gamma([0,1])$. This cover induces a finite decomposition $t_1, \ldots, t_n$ of the interval $[0,1]$, where w.l.o.g. $t_1 = 0$, $t_n = 1$. By using Lemma 2.6, it follows that $g_l = f_{j_1}^{t_1}$ for some $j_1 \in \{1, \ldots, \deg(P_l)\}$. Again, Lemma 2.6 in combination with the identity theorem yields that $f_{j_1}^{t_1} = f_{j_2}^{t_2}$ on $D_{t_1} \cap D_{t_2}$ for some $j_2 \in \{1, \ldots, \deg(P_l)\}$ and so on. Inductively, it follows that the analytic continuation of g_l along γ exists and that $g_l^\gamma = f_{j_n}^{t_1}$ for some $j_n \in \{1, \ldots, \deg(P_l)\}$. □

So far, we have shown that we are done as long as the coordinate functions are continued analytically along a non-critical path, since their analytic continuations induce continuous movements of the underlying geometric objects.

However, what happens along a critical path? Along a critical path, there usually exists a pole or branch point of a germ. Fortunately, we can resolve the problem at these points by using so-called *local uniformizing coordinates* and corresponding LAURENT *series*, as we can use them to describe the behavior of the coordinate functions around these critical points. But before we prove that we can use this approach in general, we first have a look at a simple example.

Example 2.11. Similar to previous examples, the notation below will be informal to capture the essence. Then an easy example for our consideration is the intersection of the unit circle given by $x^2 + y^2 - z^2 = 0$ and a moving line that is, in turn, induced by $x = 2tz$, where $t \in [0,1]$. The line stays parallel to the y-axis during its motion and moves horizontally form the origin $[(0,0,1)^T]$ towards the point $[(2,0,1)^T]$ (see Fig. 2.8). Three polynomials P_1, P_2 and P_3 that determine the coordinates x, y and z of a point of intersection of the circle and the line, respectively, are

$$P_1(t,x) = x - 2t, \qquad P_2(t,y) = y^2 + 4t^2 - 1, \qquad P_3(t,z) = z - 1.$$

For a fixed $t \in [0,1]$, the coordinates are given by the zeros of these polynomials. P_2 already indicates that there basically are two points of intersection. Appropriate representatives for them are given by

$$\left(2t,\ \pm\sqrt{-4t^2+1},\ 1\right)^T.$$

As illustrated in Fig. 2.8, there are two distinct real points of intersection for $t \in [0, \frac{1}{2}[$ and two distinct complex ones for $t \in]\frac{1}{2}, 1]$, while these two points coincide and result in one real point for $t_0 := \frac{1}{2}$.

This ambiguity also shows up at the algebraic representation of these points. Let $g_2^+ \in \mathbb{O}_0$ and $g_2^- \in \mathbb{O}_0$ be the germs that are induced by $+\sqrt{-4t^2-1}$ and $-\sqrt{-4t^2-1}$, respectively.

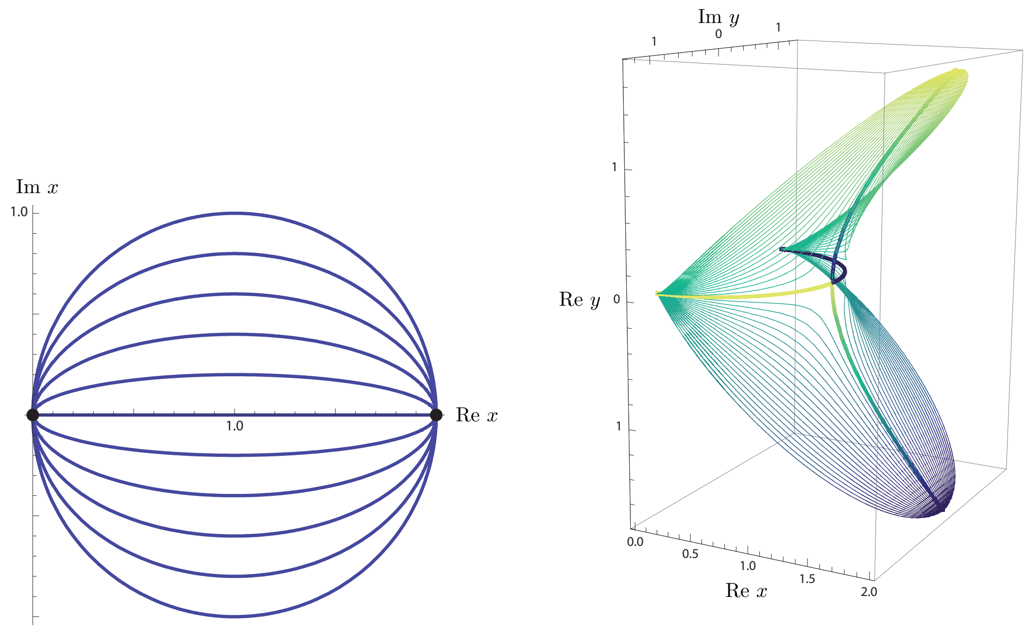

Figure 2.9: The left picture shows the path γ_b for various values of b, while the right picture illustrates the motion of the points of intersection under γ_b.

Their corresponding RIEMANN surface X_2 has two sheets and a branch point at t_0. To obtain a visualization of X_2, we fix the starting and terminal point of the x-coordinate of the points of intersection and additionally consider complex paths connecting them. In particular, they are given by

$$\gamma_b : [0,1] \to \mathbb{C};\ t \mapsto 1 + \cos\big((1-t)\pi\big) - ib\sin\big((1-t)\pi\big),$$

where $b \in [-1,1]$. Using the parameter b enables us to vary these paths in a continuous manner from a half circle above the real axis to its pendant below (compare Fig. 2.9 (left)). An illustration of X_2 is then obtained by plotting the real part of the x- and y-coordinate as well as the imaginary part of the y-coordinate as seen in Fig. 2.9 (right). Each path in this figure shows a path of one of the points of intersection under γ_b for a fixed $b \in [-1,1]$. The real situation, i.e. $b = 0$, corresponds to the thick lines in the picture. It can be seen that both points are real until their coinciding and complex afterwards. Furthermore, we observe that the structure of X_2 reflects accurately the geometric situation. Roughly speaking, the two sheets represent the two distinct points of intersection, while the branch point corresponds to the coinciding of the points.

We now come back to our original goal, this is, the construction of a continuous motion of the points of intersection under the movement of the vertical line. Therefore, let $g \in \{g_2^+, g_2^-\}$, denote the open disc with center t_0 and radius 1 by D_{t_0} and set $\dot{D}_{t_0} := D_{t_0} \setminus \{t_0\}$. Moreover, δ shall be a path within $\dot{D}_{t_0}$, i.e. $\delta([0,1]) \subset \dot{D}_{t_0}$, which starts at 0.

The point t_0 is a branch point, thus we cannot continue g analytically along a path induced by the segment $[0,1]$. However, the analytic continuation g^δ of g along δ exists, since t_0 is an isolated critical point with $t_0 \notin \delta([0,1])$.

Defining a continuous motion of the points of intersection is an easy task for each segment

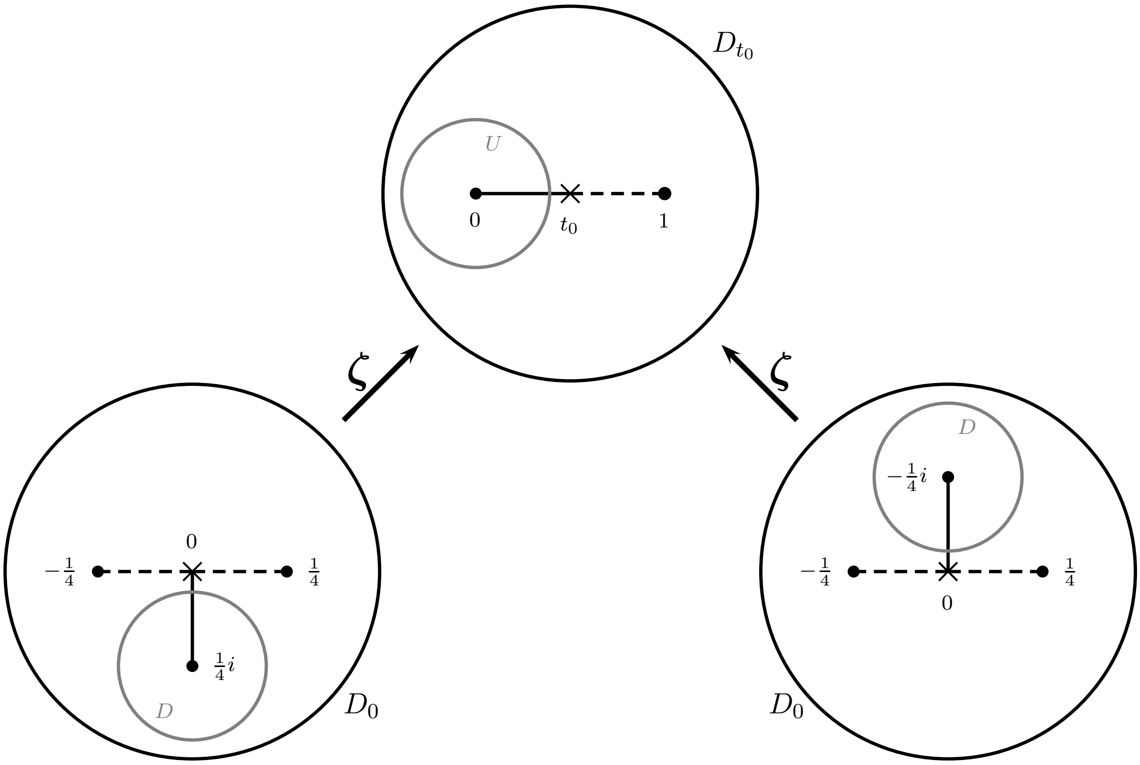

Figure 2.10: Illustration of local uniformizing variables.

$[0, t_0[$ and $]t_0, 1]$, since we can use analytic continuations of g_2^+ and g_2^- in these cases. The only problem left is to define a reasonable point of intersection at t_0 that merges the two separated motions for $[0, t_0[$ and $]t_0, 1]$ continuously. Fig. 2.8 as well as the algebraic representation already indicate an appropriate choice in our specific case, i.e. the point $[(1, 0, 1)^T]$.

However, to resolve this problem in general, this means to prove that critical points are removable singularities in the underlying object space[11]. For that, we rely on local uniformizing coordinates and Laurent series to prove that the coordinate functions do behave, in a certain sense, well[12] around a critical point. According to [1, sec. 8.1 & 8.2], there exist a Laurent series expansion

$$\sum_{k=-m}^{\infty} A_k t^k$$

with finite principal part[13] and a local uniformizing variable $\zeta(t) := t_0 + t^2$ that describe the analytic continuation g^δ. More precisely, it holds

$$\sum_{k=-m}^{\infty} A_k t^k = g^\delta \left[\zeta(t)\right]$$

[11] this is a suitable projective space.
[12] not like around an essential singularity, for instance.
[13] i.e. $m \in \mathbb{N}_0$.

on an appropriate (sub-)disc $D^\delta \subseteq \zeta^{-1}\big(\dot{D}_{t_0}\big)$. The ambiguity arising from the two distinct points of intersection is encapsulated in the map ζ, since ζ is a double cover of $\dot{D}_{t_0}$. Formally, there is an open disc $\dot{D}_0$ with its center 0 removed, which ζ maps surjectively onto D_{t_0} and the set $\zeta^{-1}(t)$ exactly consists of two distinct points for $t \in \dot{D}_0$. For instance, regardless whether $g = g_2^+$ or $g = g_2^-$, there exists an open neighborhood U of $0 \in \dot{D}_{t_0}$, on which g is well-defined. This neighborhood is illustrated by the gray disc around 0 in the upper circle D_{t_0} of Fig. 2.10. That there are two possible choices for g is reflected by the fact that there are two open neighborhoods D that are mapped onto U under ζ. They are visualized by the gray neighborhoods in the two lower circles of Fig. 2.10. Choosing one of the two points of intersection in a starting instance also fixes one of the two neighborhoods D. After that, the motion of this point of intersection is determined for $t \in [0, \frac{1}{2}[$. In Fig. 2.10, this corresponds to a unique solid line segment in each of the three circles. Similarly, there exist two possible continuous motions for $t \in]\frac{1}{2}, 1]$ independent of the choice in the starting instance. They are represented by the dashed line segments in Fig. 2.10.

The last step is to complete the construction of a continuous motion by proving that there always exists an appropriate object that removes the singularity in the object space and merges the different motions continuously. This will be done in Lemma 2.9 and in Theorem 2.3.

That we can generally use local uniformizing coordinates and corresponding Laurent series in the same way as presented in the previous example is the goal of the next lemma. But before we prove it, we introduce some further useful notations and recall the abbreviations used in the previous example. We denote an open disc with center $z \in \mathbb{C}$ by D_z, its punctured pendant $D_z \setminus \{z\}$ by $\dot{D}_z$ and the set of meromorphic function on an open set $U \subseteq \mathbb{C}$ is denoted by $\mathcal{M}(U)$.

Lemma 2.8. *X_l is a finitely sheeted* Riemann *surface and the set of all critical points of g_l is discrete for all $l \in \{1-M, \ldots, N\}$. Furthermore, for every $z_0 \in \mathbb{U}$, there exists a punctured disc $\dot{D}_{z_0} \subseteq \mathbb{U}$ centered at z_0 so that for every non-critical path γ of $P_l(g_{1-M}, \ldots, g_{l-1}, X)$ ($l \in \{1, \ldots, N\}$) that starts at $0 \in \mathbb{U}$ and ends in $\dot{D}_{z_0}$, there are a local uniformizing variable*

$$\zeta_\gamma(z) := z_0 + z^s,$$

where $s \in \mathbb{N}$ and N Laurent *series*

$$\sum_{k=-m_l}^{\infty} A_{l,k}^{\gamma} z^k$$

with finite principal part, i.e. $m_l \in \mathbb{N}_0$, and the following property: For every path δ that starts at the terminal point of γ and satisfies $\delta([0,1]) \subseteq \dot{D}_{z_0}$, it holds that

$$\sum_{k=-m_l}^{\infty} A_{l,k}^{\gamma} z^k = g_l^{\delta\circ\gamma}\left[\zeta_\gamma(z)\right], \tag{2.2}$$

for all z in an appropriate (sub-)disc $D^{\delta\circ\gamma} \subseteq \zeta_\gamma^{-1}\left(\dot{D}_{z_0}\right)$.

Remark 2.11. At this point, we shall point out a few things. The situation around a point z_0 may be path-dependent. For this reason, we have the partition into γ and δ. In addition, Laurent series and local uniformizing variables are actually only needed when a branch point or a pole occurs, since then they are capturing the ambiguity around a branch point or the behavior around a pole. However in the majority of cases, the LAURENT series becomes a usual power series and the local uniformizing variable a simple translation for all non-critical points $z_0 \in \mathbb{U}$. We will observe that there can only exist a finite number of critical points in the interval $[0, 1]$.

Proof. We prove the statement by induction over $N \in \mathbb{N}_0$, the number of dependent local coordinate functions. Let $z_0 \in \mathbb{U}$. For $N = 0$, the definition of the h-movements $\mu_{1-m}, \ldots, \mu_0$ induce[14] that there exists a punctured disc $\dot{D}_{z_0}$ and one local uniformizing variable

$$\zeta(z) = z_0 + z,$$

a disc D and a corresponding LAURENT series $\sum_{k=-m_l}^{\infty} A_{l,k} z^k$ independent of the detailed choice of γ connecting 0 and $\gamma(1) \in \dot{D}_{z_0}$ in $\mathbb{U}$. Moreover, the other parts of the claim are an immediate consequence of the definition of an h-movement.

Now, let $N > 0$. By induction hypothesis, there exists a punctured disc $\dot{D}_{z_0}$ so that property (2.2) holds for all $l \in \{1 - M, \ldots, N - 1\}$ and for every non-critical path γ of $P_N(g_{1-M}, \ldots, g_{N-1}, X)$. Therefore, let γ be such a path. As a result, we can represent the continuation of the polynomial $P_N(g_{1-M}, \ldots, g_{N-1}, X) \in \mathbb{O}_0[X]$ in terms of the LAURENT series given by the induction hypothesis. We just have to replace the coefficients in the continued polynomial by proper LAURENT series according to Equation (2.2). Let $\zeta_\gamma(z) = z_0 + z^s$ be the local uniformizing variable given by the induction hypothesis. Then there exists a polynomial $\tilde{P}_\gamma(z, X) \in \mathcal{M}\big(\dot{D}_0\big)[X]$ so that for every path δ with $\delta(0) = \gamma(1)$ and $\delta([0, 1]) \subseteq \dot{D}_{z_0}$, the identity

$$P_N\left(g_{1-M}^{\delta\circ\gamma}[\zeta_\gamma(z)], \ldots, g_{N-1}^{\delta\circ\gamma}[\zeta_\gamma(z)], X\right) = \tilde{P}_\gamma(z, X)$$

holds for all z in an appropriate disc $D^{\delta\circ\gamma} \subset \dot{D}_0$. The coefficients of $\tilde{P}_\gamma$ only have a finite principal part, since they are polynomials of LAURENT series with finite principal parts. Thus the zeros of each coefficient of $\tilde{P}_\gamma$ as well as of the discriminant of $\tilde{P}_\gamma$ are lying discrete in $\dot{D}_0$. By induction hypothesis, the set of critical points of $P_l(g_{1-M}, \ldots, g_{l-1}, X)$ for all $l \in \{1 - M, \ldots, N - 1\}$ is discrete. So there exists a punctured disc $\dot{B}_0^\gamma \subset \dot{D}_0$ centered at $0 \in \mathbb{C}$, on which all coefficients of $\tilde{P}_\gamma$ are well-defined and the discriminant as well as the leading coefficient of $\tilde{P}_\gamma$ do not vanish.

Since X_l is a finitely sheeted RIEMANN surface for all $l \in \{1 - M, \ldots, N - 1\}$, there exist only finitely many different of these polynomials $\tilde{P}_\gamma(z, X) \in \mathcal{M}\big(\dot{D}_0\big)[X]$ and discs $\dot{B}_0^\gamma$. For this reason, there exists a path-independent punctured disc $\dot{B}_0 \subset \dot{D}_0$ centered at $0 \in \mathbb{C}$, on which neither the leading coefficients of all $\tilde{P}_\gamma$ nor the discriminants of all $\tilde{P}_\gamma$ vanish and all coefficients of $\tilde{P}_\gamma$ are well-defined. This, in turn, induces a punctured disc $\dot{B}_{z_0}$ centered at $z_0 \in \mathbb{U}$ that does not contain any critical point of $P_N(g_{1-M}, \ldots, g_{l-1}, X)$. So it follows that: If γ is a non-critical path of $P_l(g_{1-M}, \ldots, g_{l-1}, X)$ for all $l \in \{1 - M, \ldots, N\}$, which starts at 0 and ends in $\dot{B}_{z_0}$, then the analytic continuations of $c_1, \ldots, c_N$ along $\delta \circ \gamma$ exist according to Lemma 2.7 for all paths δ with $\delta(0) = \gamma(1)$ and $\delta([0, 1]) \subseteq \dot{B}_{z_0}$.

To show that X_N is finitely sheeted, we study the set $\pi_N^{-1}(z_1)$ for every $z_1 \in \dot{B}_{z_0}$. Let $g \in \pi_N^{-1}(z_1)$. In the trivial case, g can be obtained by analytic continuation of g_N along a non-critical path γ of $P_N(g_{1-M}, \ldots, g_{l-1}, X)$. Due to this fact g has to be one of the finitely many satisfying germs of $P_N(g_{1-M}^\gamma, \ldots, g_{l-1}^\gamma, X)$. Since $X_{1-M}, \ldots, X_{N-1}$ is finitely sheeted, only a finite number of such polynomials $P_N(g_{1-M}^\gamma, \ldots, g_{l-1}^\gamma, X)$ exists. Apart from this

[14] Every h-movement is the restriction of a holomorphic function on $\mathbb{U}$.

case, we now assume that g can only be obtained by analytic continuation along a critical path γ of $P_N(g_{1-M},\dots,g_{l-1},X)$. But then one of the following three cases must be true: The leading coefficient of P_N vanishes on a non-discrete set, the discriminant vanishes on a non-discrete set or the set of critical points of $g_{1-M},\dots,g_{N-1}$ is not discrete. All of them result in contradiction. As a result, g can always be obtained by an analytic continuation along a non-critical path implying that X_N only has a finite number of sheets.

According to the construction in [1, sec. 8.1 & 8.2], there exists the claimed LAURENT series expansion corresponding to X_N and a local uniformizing variable $\tilde{\zeta}_\gamma(z) := z_0 + z^{\tilde{s}}$ so that property (2.2) holds for $g_N^{\delta\circ\gamma}$ on the appropriate (sub-)disc $\tilde{D}^{\delta\circ\gamma} \subseteq \dot{\tilde{D}}_0$. This also implies the set of singular points of g_N has to be discrete.

To complete the proof, we have to *merge* ζ_γ and $\tilde{\zeta}_\gamma$. Therefore, let K be the least common multiple of s and $\tilde{s}$ and $z' \in \mathbb{C}$ so that z' is an arbitrary K-th roots of $\delta(1)-z_0$. Furthermore, let ϑ,ι be the two K-th root of unity so that $(\vartheta z')^{\frac{K}{s}}$ is the center of the disc $D^{\delta\circ\gamma}$ and $(\iota z')^{\frac{K}{\tilde{s}}}$ is the center of the disc $\tilde{D}^{\delta\circ\gamma}$. By doing so, we get from equation (2.2)

$$\sum_{k=-m_l}^{\infty} A^\gamma_{l,k}\left((\vartheta z)^{\frac{K}{s}}\right)^k = g_l^{\delta\circ\gamma}\left[z_0+\left((\vartheta z)^{\frac{K}{s}}\right)^s\right] = g_l^{\delta\circ\gamma}\big[\underbrace{z_0+z^K}_{=:\zeta_\gamma(z)}\big], \tag{2.3}$$

for all z in an appropriate disc D centered at z' and $l \in \{1-M,\dots,N-1\}$. Analogously, we have

$$\sum_{k=-m_N}^{\infty} A^\gamma_{N,k}\left((\iota z)^{\frac{K}{\tilde{s}}}\right)^k = g_N^{\delta\circ\gamma}\left[z_0+\left((\iota z)^{\frac{K}{\tilde{s}}}\right)^{\tilde{s}}\right] = g_N^{\delta\circ\gamma}\big[\zeta_\gamma(z)\big], \tag{2.4}$$

for all z in D. By the use of the identity theorem, one gets that for every proper $\delta\circ\gamma$, there exists an appropriate subdisc $D^{\delta\circ\gamma}$, on which equation (2.3) respectively equation (2.4) holds. ☐

An immediate consequence of the previous lemma is the fact that the number of critical points in $[0,1]$ has to be finite.

Corollary 2.2. *The set of all critical points of $g_{1-M},\dots,g_N$ is finite in $[0,1]$.*

Proof. According to Lemma 2.8, there are punctured discs $\dot{D}_z$ for all $z \in [0,1]$. These discs including their centers form an open cover of $[0,1]$. Since this interval is compact, even a finite number of these discs cover already $[0,1]$. As a result, there is only a finite number of critical points. ☐

As already mentioned, we prove a continuation lemma. It provides two features: On the one hand, it enables us to remove singularities, which occur at branch points and poles, in the object space. On the other hand, it shows that the cases, where the representing vector becomes the all-zero vector, can be removed continuously in the object space as well.

Lemma 2.9 (continuation lemma)**.** *Let $U \subset \mathbb{C}$ be open and $c_1,\dots,c_{d+1} \in \mathcal{M}(U)$ be meromorphic functions without essential singularities and where at least one of them does not vanish identically. In addition, let S be the discrete set of all singularities and N the discrete set of all common zeros of $c_1,\dots,c_{d+1}$ in U. Then the induced object function*

$$\mathbf{o} : U\setminus(S\cup N) \to \mathbb{CP}^d,\ t \mapsto \left[\left(c_1(t),\dots,c_{d+1}(t)\right)^T\right],$$

is continuous and $\mathbf{o}$ can be extended continuously for all $t' \in S\cup N$.

Proof. First of all, we show that $\mathbf{o}$ is continuous. For $t' \in U \setminus (S \cup N)$, there exists a neighborhood W of t', on which at least one function, say c_1, does not vanish. In this situation,

$$\kappa : V \to \mathbb{C}^{d+1},\ \left[(z_1, \dots, z_{d+1})^T\right] \mapsto \left(1, \frac{z_2}{z_1}, \dots, \frac{z_{d+1}}{z_1}\right)^T,$$

where V is a neighborhood of $\mathbf{o}(t')$, is a valid chart of the complex manifold $\mathbb{CP}^d$. By definition, $\mathbf{o}$ is continuous at t', if $\kappa \circ \mathbf{o} : U \setminus (S \cup N) \to \mathbb{C}^{d+1}$ is continuous at t'. But it is obvious that

$$\kappa \circ \mathbf{o} : W \to \mathbb{C}^{d+1},\ t \mapsto \left(1, \frac{c_2(t)}{c_1(t)}, \dots, \frac{c_{d+1}(t)}{c_1(t)}\right)^T$$

is continuous at t'. Now let $t' \in N$ and $k, l \in \{1, \dots, d+1\}$. Then we define the relation

$$k \prec l \quad :\Longleftrightarrow \quad \lim_{t \to t'} \frac{c_l(t)}{c_k(t)} \notin \mathbb{C}.$$

Since there is at least one coordinate function, which does not vanish identically, there has to exist a $k' \in \{1, \dots, d+1\}$ with $k' \not\prec k$. As a result,

$$\lim_{t \to t'} \frac{c_k(t)}{c_{k'}(t)} \in \mathbb{C}$$

for $k \in \{1, \dots, d+1\} \setminus \{k'\}$ and

$$\kappa' : V' \to \mathbb{C}^{d+1},\ \left[(c_1(t), \dots, c_{d+1}(t))^T\right] \mapsto \left(\frac{c_1(t)}{c_{k'}(t)}, \dots, \frac{c_{d+1}(t)}{c_{k'}(t)}\right)^T,$$

where V' is a neighborhood of

$$\left[\lim_{t \to t'} \left(\frac{c_1(t)}{c_{k'}(t)}, \dots, \frac{c_{d+1}(t)}{c_{k'}(t)}\right)^T\right]$$

is a valid chart of $\mathbb{CP}^d$. In conclusion, $\kappa' \circ \mathbf{o} : W' \to \mathbb{C}^{d+1}$, where W' is an appropriate neighborhood of t', can be extended continuously at t', which also implies that there exists a continuous extension of $\mathbf{o}$ at t'. For $t' \in S$, one can apply an analogous argumentation. $\square$

2.4 Results Concerning Continuous Dynamic Geometry Systems

We will derive the main results of this chapter in the present section. As a first step, we are going to show constructively that there always exists a CDGS associated to an algebraic construction and a regular starting instance. Conversely, it will also turn out that every possible CDGS corresponding to an algebraic construction and a regular instance can be described by a CDGS that has already been derived in the constructive proof of the just mentioned existence theorem. The second subsection focuses on the limitations of the concept of algebraic constructions. For instance, we are going to observe that at least some transcendental functions cannot be rebuild by algebraic constructions. So far, all results rely essentially on the fact that h-movements are restrictions of holomorphic functions. However, we are going to prove that this assumption is necessary to obtain continuity in a third subsection. In other words, a CDGS cannot be based only on k-times differentiable functions describing the motion of the free objects, in general.

2.4.1 Possible Continuous Dynamic Geometry Systems and Their Existence

Now, we are able to prove constructively that there always exists a CDGS[15] in the sense of Definition 2.6 associated to a given algebraic construction and a regular starting instance. The proof bases on the idea presented in Example 2.11: Since the set of critical points is discrete in $[0,1]$, we can use analytic continuations of the coordinate functions to construct continuous motions along the segments that do not contain a critical point. Furthermore, we show that these motions can be merged continuously at the critical points yielding finally one continuous motion in the object space.

Theorem 2.3 (existence of CDGSs)**.** *Let $\mathscr{C}$ be an algebraic construction according to Definition 2.3 and $\vec{\mathbf{o}} \in \mathcal{I}_{\mathscr{C}}$ a regular instance. Then there exists a corresponding CDGS $\mathcal{D}$.*

Proof. Let Φ be the set of all functions $\mu : [0,1] \to O_{1-m} \times \ldots \times O_0$ with $\mu_j(0) = \mathbf{o}_j$ and μ_j is an h-movement for all $j \in \{1-m, \ldots, 0\}$. Furthermore, let Ψ be the set of all functions $\nu : [0,1] \to O_1 \times \ldots \times O_n$ with $\big(\mu(t)^T, \nu(t)^T\big) \in \mathcal{I}_{\mathscr{C}}$ for all $t \in [0,1]$. For $\mu \in \Phi$, we have to show that $\nu_j : [0,1] \to O_j$ may be chosen as a continuous mapping for all $j \in \{1, \ldots, n\}$. To simplify the notation, we fix $j \in \{1, \ldots, n\}$ and omit the index j below.

From Lemma 2.4, it follows that there exist germs $g_1, \ldots, g_{d+1} \in \mathbb{O}_0$ representing the initial behavior of the dependent object $\mathbf{o}$. According to Lemma 2.8, there exists a punctured disc $\dot{D}_{z_0}$ centered at z_0 associated to property (2.2) for every $z_0 \in \mathbb{U}$. The corresponding discs $D_{z_0} = \dot{D}_{z_0} \cup \{z_0\} \subset \mathbb{U}$ induce an open cover of $[0,1]$. Since $[0,1]$ is compact, we can find a finite subcover $\{D_{t_0}, \ldots, D_{t_r}\}$ of $[0,1]$ with $t_0 < \ldots < t_r \in [0,1]$. Thereby, we may assume w.l.o.g. that $t_0 = 0$, $t_r = 1$ and

$$\left\{ t \in [0,1] \;\middle|\; \exists\, \gamma \text{ such that } \left(\widehat{g_1^\gamma}, \ldots, \widehat{g_{d+1}^\gamma}\right)^T = (0, \ldots, 0) \in \mathbb{C}^{d+1} \right\} \subseteq \{t_0, \ldots, t_r\}.$$

Let the radius of D_{t_l} be denoted by ε_l. Below, we are going to define functions $f_l : [0, t_l + \varepsilon_l[\to \mathbb{CP}^d$ by an inductive construction over $l \in \{0, \ldots, r\}$ that are going to form a CDGS. Thereby, there should exist a path γ connecting 0 and t so that

$$f_l(t) = \left[\left(\widehat{g_1^\gamma}, \ldots, \widehat{g_{d+1}^\gamma}\right)^T \right] \tag{2.5}$$

for every $t \in [0, t_l + \varepsilon_l[\setminus \{t_0, \ldots, t_r\}$. In order to start the induction, we mention that the claim follows immediately by Lemma 2.4 for $l = 0$. Now, let $0 < l \leq r$. By induction hypothesis, we have a continuous mapping $f_{l-1} : [0, t_{l-1} + \varepsilon_{l-1}[\to \mathbb{CP}^d$ so that for all $t \in [0, t_{l-1} + \varepsilon_{l-1}) \setminus \{t_0, \ldots, t_r\}$, there exists a path γ connecting 0 and t and

$$f_{l-1}(t) = \left[\left(\widehat{g_1^\gamma}, \ldots, \widehat{g_{d+1}^\gamma}\right)^T \right].$$

Next, let $t \in \big([0, t_{l-1} + \varepsilon_{l-1}[\cap D_{t_l}\big) \setminus \{t_0, \ldots, t_r\}$ and γ be such a path. Moreover, let $\zeta_\gamma(z) := t_l + z^{s(t_l)}$ be the $s(t_l)$-fold cover of D_{t_l} according to Lemma 2.8 and $\sigma_1(t_l; z), \ldots, \sigma_{d+1}(t_l; z)$ the corresponding Laurent series. In addition, let D^γ be the disc of $\zeta_\gamma^{-1}(D_{t_l}) \setminus \{0\}$, on which $\sigma_j(t_l; z) = g_j^\gamma[\zeta_\gamma(z)]$ for all $j \in \{1, \ldots, d+1\}$. Since $[t, t_l + \varepsilon_l[\subset D_{t_l}$, there exists a path $\delta : [0,1[\to \zeta_\gamma^{-1}(D_{t_l})$ so that ζ_γ maps $\delta\big([0,1)\big)$ homeomorphically on $[t, t_l + \varepsilon_l[$. Besides, $\delta(0)$ is the center of D^γ. We denote this homeomorphism $\zeta_\gamma\big|_{\delta([0,1[)}$ by η.

[15] At that, we shall remark once more that the motion of the free objects has to be induced by holomorphic functions.

According to Lemma 2.9, there exists a continuous extension h of

$$z \mapsto \left[\left(\sigma_1(t_l;z),\ldots,\sigma_{d+1}(t_l;z)\right)^T\right] \in \mathbb{CP}^d$$

defined on $\zeta_\gamma^{-1}(D_{t_l})$. So we now can define a map

$$f_l : [0, t_l + \varepsilon_l[\to \mathbb{CP}^d,\ x \mapsto \begin{cases} f_{l-1}(x), & \text{if } x \in [0,t], \\ h(\eta^{-1}(x)), & \text{if } x \in]t, t_l + \varepsilon_l[\end{cases}$$

The next step is to show that f_l is continuous at t. Since $h(\eta^{-1}(x))$ is continuous on $[t, t_l + \varepsilon_l[$, this would follow from $f_{l-1}(t) = h(\eta^{-1}(t))$. But it holds that

$$h(\eta^{-1}(t)) = h(\delta(0)) = \left[\left(\sigma_1\big(t_l;\delta(0)\big),\ldots,\sigma_{d+1}\big(t_l;\delta(0)\big)\right)^T\right].$$

As a result, the definition of δ implies $f_{l-1}(t) = h(\delta(0))$. Finally, it is obvious that $f_l(t)$ satisfies the claimed continuation property given by (2.5). To complete the proof, we set $\nu_j = f_r\big|_{[0,1]}$. □

Remark 2.12. We observe that the CDGS $\mathcal{D}$ in the previous proof may not be unique. At each critical point $t \in \{t_0,\ldots,t_r\}$, we may have several opportunities to proceed continuously. The number of choice is encoded by the exponent of the local uniformizing variables.

This also matches our observations in Example 2.11, where we consider the intersection of a circle and a moving line. After choosing a starting instance, the motion of a corresponding point of intersection is uniquely determined until the tangent situation. However, there are two opportunities to move on continuously after the tangent situation.

Remark 2.13 (geometric theorems). At this point, we would like to address another worthwhile point that actually follows by the developed theory. Lots of geometric theorems base on a certain construction, which then admits a conclusion that is encoded by a homogenous polynomial[16] in the coordinates of the construction elements. Examples are three points that are collinear or three lines which are concurrent. The theorem holds, if a polynomial vanishes identically. Analogous to the preceding argumentation, the polynomial stays the zero polynomial along all possible analytic continuations implying that the geometric theorem remains valid under these continuations.

After we have shown that there exists a CDGS corresponding to a given algebraic construction and an associated regular starting instance, our next goal is to show that every possible CDGS has to be one of the CDGSs constructed in the previous proof. This also implies that there exists only a finite number of CDGSs. To do so, we first show the following lemma, in which we prove that whenever a CDGS outputs a regular instance of an algebraic construction, which can be described by a set of germs, then these germs describe also the CDGS on an entire segment.

Lemma 2.10. *Let $\mathcal{D}$ be a CDGS associated to a given algebraic construction $\mathscr{C}$ with $n \in \mathbb{N}_0$ dependent objects and a regular starting instance. Moreover, let $\vec{\mathbf{o}} \in \mathcal{I}_{\mathscr{C}}$ be a further regular instance, $\mu_{1-m},\ldots,\mu_0$ h-movements of $\mathscr{C}$ so that $\mu_j(t_0) = \mathbf{o}_j$ for some $t_0 \in [0,1]$ and for all $j \in \{1-m,\ldots,0\}$ and set $\mu := \left(\mu_{1-m},\ldots,\mu_0\right)^T$. If, for every $k \in \{1,\ldots,n\}$, there exist germs $g_{k,1},\ldots,g_{k,d_k+1} \in \mathbb{O}_{t_0}$ that represent the object $\mathbf{o}_k$ at t_0, i.e.*

$$\left[\left(\widehat{g_{k,1}},\ldots,\widehat{g_{k,d_k+1}}\right)^T\right] = \mathbf{o}_k,$$

[16]In general, an identity that is analytic in the coordinates of the construction elements would yield the same result.

then there is $\varepsilon > 0$ so that the motion of the object $\mathbf{o}_k$ under μ, i.e. the k-th component ν_k of $\mathcal{D}(\mu)$, is given by

$$\nu_k(t) = \left[\left(g_{k,1}[t], \ldots, g_{k,d_k+1}[t]\right)^T\right]$$

for $t \in \left]\max\{0, t_0 - \varepsilon\}, \min\{1, t_0 + \varepsilon\}\right[$ and $k \in \{1, \ldots, n\}$.

Proof. We show the claim by induction over $n \in \mathbb{N}_0$, the number of dependent objects of $\mathscr{C}$. For $n = 0$, there is nothing to show. Now, let $n > 0$. If one removes the last operation of $\mathscr{C}$, the induction hypothesis guarantees that $\nu_1, \ldots, \nu_{n-1}$ satisfy the claimed equation. Since

$$\nu_n(t_0) = \mathbf{o}_n = \left[\left(g_{n,1}[t_0], \ldots, g_{n,d_n+1}[t_0]\right)^T\right],$$

we can add to the homogeneous polynomial system P_{ω_n} an affine equation so that one of the possible solutions of this extended system representing $\nu_n(t_0)$ is

$$\left(g_{n,1}[t_0], \ldots, g_{n,d_n+1}[t_0]\right)^T.$$

Obviously, the coefficients as well as the solutions of the extended system depend continuously on t. As every possible solution is simple, the claim follows immediately by the continuity of ν_n. □

We are going to use the previous lemma to show the mentioned goal, namely that the construction of a CDGS in the proof of Theorem 2.3 captures all possible CDGS.

Theorem 2.4 (possible CDGSs). *Let $\mathscr{C}$ be an algebraic construction, $\vec{\mathbf{o}} \in \mathcal{I}_{\mathscr{C}}$ a regular starting instance, $\mu_{1-m}, \ldots, \mu_0$ h-movements of $\mathscr{C}$ and $\mathcal{D}$ an associated CDGS. For each $j \in \{1, \ldots, n\}$, denote the j-th component of $\mathcal{D}(\mu)$ by ν_j, where $\mu := (\mu_{1-m}, \ldots, \mu_0)$ and let $g_{j,1}, \ldots, g_{j,d_j+1} \in \mathbb{O}_0$ be germs with*

$$\left[(\widehat{g_{j,1}}, \ldots, \widehat{g_{j,d_j+1}})^T\right] = \nu_j(0).$$

Then, there exists a path γ connecting $0 \in \mathbb{C}$ and $t \in \mathbb{C}$ so that

$$\nu_j(t) = \left[\left(\widehat{g^\gamma_{j,1}}, \ldots, \widehat{g^\gamma_{j,d_j+1}}\right)^T\right]$$

for almost every $t \in [0, 1]$.

Remark 2.14. It is clear according to Lemma 2.4 that such germs always exist.

Proof. As usual, we are going to show the claim by induction over $n \in \mathbb{N}_0$, the number of dependent objects of $\mathscr{C}$. Since there is nothing to show for $n = 0$, let $n > 0$. The previous lemma guarantees that the claim holds for all $t \in [0, \varepsilon[$ for some $\varepsilon > 0$. Assuming that the analytic continuations of all germs exist at least along a segment $[0, t_1]$ for $t_1 > \varepsilon$, the claimed identity holds at least on $[0, t_1]$, since otherwise it would follow: There has to exist a $t' \in [0, t_1]$ so that $\left(\mu_{1-m}(t'), \ldots, \nu_n(t')\right)$ is a regular instance of $\mathscr{C}$, the continuations of all germs along $[0, t']$ exist and the desired identity does not hold. So we can generate the objects

$$\left[\left(g^{[0,t']}_{k,1}[t'], \ldots, g^{[0,t']}_{k,d_k+1}[t']\right)^T\right]$$

for all $k \in \{1, \ldots, n\}$. But they must be equal $\nu_k(t')$ according to continuity. Applying the previous Lemma again yields a contradiction to the existence of t'.

For this reason we assume that there is a minimal[17] and isolated critical point $t_1 \in]0,1]$ of one of the germs. Then, the claim holds for all $t \in [0, t_1[$. By using the induction hypothesis, it follows that there exists an open neighborhood U of t_1 so that for all $t \in U \cap [0,1]$ with $t > t_1$, there exists a path γ connecting 0 and t so that

$$\nu_j(t) = \left[\left(\widehat{g_{j,1}^{\gamma}}, \ldots, \widehat{g_{j,d_j+1}^{\gamma}}\right)^T\right]$$

for all $j \in \{1, \ldots, n-1\}$. This implies that coefficients of the polynomial system P_{ω_n} associated to ω_n are given by analytic continuations of the initial germs for all $t \in U \cap [0,1]$ with $t > t_1$. Since ν_n has to satisfy this system, it must hold

$$\nu_n(t) = \left[\left(\widehat{g_{n,1}^{\gamma}}, \ldots, \widehat{g_{n,d_n+1}^{\gamma}}\right)^T\right],$$

for $t \in U \cap [0,1]$ with $t > t_1$. The number of such critical points t_1 is finite in $[0,1]$, thus the claim follows inductively. □

2.4.2 The Power of Algebraic Constructions

Although, the concept of algebraic constructions captures lots of concrete geometric constructions, limitations in their power are given. A formal framework capturing this is the notion of constructible functions (compare [30]). Roughly speaking, we say a function is constructible, if we can construct a point with a coordinate that equals the function on an interval, while one free object moves straight.

Definition 2.9 (constructible function)**.** *Given* $\mu_0 : [0,1] \to \mathcal{P};\ t \mapsto \left[(t,0,1)^T\right]$. *Then we call a function* $f : [0,1] \to \mathbb{C}$ constructible, *if there are an algebraic construction* $\mathscr{C}$ *with one free point, i.e.* $\omega_0 = \mathcal{P}$, *a regular instance* $\vec{\mathbf{o}} \in \mathcal{I}_{\mathscr{C}}$ *with* $\vec{\mathbf{o}}_0 = \left[(0,0,1)^T\right]$ *and a CDGS* $\mathcal{D}$ *with* $\mathcal{D}_n(\mu_0) = \nu_n$ *so that*

$$\nu_n : [0,1] \to \mathcal{P};\ t \mapsto \left[\left(f(t), 0, 1\right)^T\right].$$

That this definition does not only admit trivial functions is shown by the following example. For the sake of brevity, we omit some of the building blocks like the so-called VON STAUDT-constructions providing a geometric addition and multiplication and a construction that extracts the coordinates of a point. But all of them can be built of the operations JOIN and MEET and one can find them in [48].

Example 2.12. In order to show that $f : [0,1] \to \mathbb{C};\ t \mapsto \sqrt{1-t}$ is a constructible function, we need to introduce a further geometric gadget. The algebraic operation

$$\begin{aligned}\text{SAMEDIST} := \Big\{ \big([a],[b],[l],[p]\big) \in \mathcal{P}^2 \times \mathcal{L} \times \mathcal{P} \;\big|\; l^T p = 0 \text{ and} \\ \det(b,a,\mathtt{I})\det(p,a,\mathtt{I})\det(b,p,\mathtt{J})^2 = \det(b,a,\mathtt{J})\det(p,a,\mathtt{J})\det(b,p,\mathtt{I})^2 \Big\},\end{aligned}$$

provides a point $[p]$ lying on $[l]$, whose euclidian distance to $[a]$ equals the euclidian distance from $[a]$ to $[b]$. Using the algebraic operations introduced on page 26, we state a subconstruction that provides the circle $\mathbf{c}$ to given center $\mathbf{M}$ and a further point $\mathbf{R}$ on the circle.

[17] This works, since the number of critical points is finite in $[0,1]$ (see Lemma 2.2).

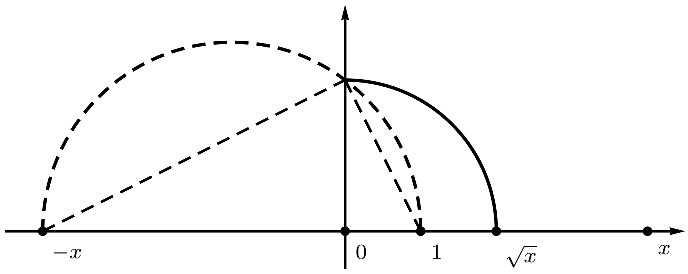

Figure 2.11: A construction for $\sqrt{x}$.

The construction steps are as follows:

$$\begin{aligned}
\mathbf{M} &\in \mathcal{P}\\
\mathbf{R} &\in \mathcal{P}\\
\mathbf{P}_1 &\in \mathrm{CONST}_{\mathbf{P}_1}\\
\mathbf{P}_2 &\in \mathrm{CONST}_{\mathbf{P}_2}\\
\mathbf{l}_1 &\in \mathrm{JOIN}(\mathbf{M}, \mathbf{P}_1)\\
\mathbf{l}_2 &\in \mathrm{JOIN}(\mathbf{M}, \mathbf{P}_2)\\
\mathbf{R}_1 &\in \mathrm{SAMEDIST}(\mathbf{M}, \mathbf{R}, \mathbf{l}_1)\\
\mathbf{R}_2 &\in \mathrm{SAMEDIST}(\mathbf{M}, \mathbf{R}, \mathbf{l}_2)\\
\mathbf{c} &\in \mathrm{CIRCLE}(\mathbf{R}, \mathbf{R}_1, \mathbf{R}_2)
\end{aligned}$$

This construction works, if $\mathbf{P}_1 \neq \mathbf{R}$ and $\mathbf{P}_2 \neq \mathbf{R}$. Additionally, it must hold that $\mathbf{M}$, $\mathbf{P}_1$ and $\mathbf{P}_2$ are not collinear. Such points, however, can always be found.

The next building block is a construction that yields a point $[(\sqrt{x}, 0, 1)^T]$ for a given point $[(x, 0, 1)^T]$ with $x \neq 0$. Figure 2.11 illustrates such a construction (compare [15, p. 376]). The correctness of it is easily verified, namely the product of the distance from 0 to 1 and $-x$ to 0 equals the square of the altitude seen in the triangle of this figure.

Finally, the term $\sqrt{1-t}$ induces a constructible function, since one can compose a desired algebraic construction by using the introduced algebraic operations together with the previous subconstruction.

After we have exemplified the notion of constructible functions, we confine our attention to the limitations of our formal model. Since we have started from an algebraic setting, it is not surprising that transcendental functions are essentially not constructible. This is shown in the next theorem and the directly following corollary.

Theorem 2.5. *The function* $\exp : [0,1] \to \mathbb{C};\ t \mapsto e^t$ *is not constructible.*

Before this theorem can be proved, we require that e^t is transcendental over $\mathbb{Q}$ for $t \in \mathbb{Q}$. This can be followed from the famous Theorem of LINDEMANN and WEIERSTRASS.

Theorem 2.6. *Given* $n \in \mathbb{N}$ *pairwise distinctive algebraic numbers* $a_1, \ldots, a_n$*. Furthermore, let* $b_1, \ldots, b_n$ *also be algebraic numbers not all of them zero. Then it holds that*

$$b_1 \cdot e^{a_1} + \cdots + b_n \cdot e^{a_n} \neq 0.$$

We refer to [35] for a proof of this theorem. We now use it to prove that e^t and $\log t$ are transcendental over $\mathbb{Q}$ for $t \in \mathbb{Q}$.

Lemma 2.11. *Let $t \in \mathbb{Q}$. Then e^t and $\log t$ are transcendental over $\mathbb{Q}$.*

Proof. For the first statement of lemma, we again refer to [35, pp. 223]. In order to show the second one, we assume that $\log t$ is an algebraic number. Obviously, t^{-1} is also algebraic. Applying Theorem 2.6 to $\log t$ and t^{-1} would yield

$$t^{-1} \cdot e^{\log t} - e^0 = 1 - 1 \neq 0.$$

As a result, $\log t$ must be transcendental. □

So we are able to prove Theorem 2.5, in which we have stated that the exponential function is not constructible.

Proof. Let us assume that exp is constructible. For this reason, there exist an algebraic construction $\mathscr{C}$, a CDGS $\mathcal{D}$ and a corresponding regular starting instance $\vec{\mathbf{o}}$ in the sense of definition 2.9. The number of points $t \in [0,1]$ so that

$$\Big[\big(\mu_0(t), \nu_1(t), \ldots, \nu_n(t)\big)\Big] \in \mathcal{I}_{\mathscr{C}}$$

forms a non-regular instance is finite in $[0,1]$. As a result, there exists a $t' \in [0,1] \cap \mathbb{Q}$ so that

$$\Big[\big(\mu_0(t'), \nu_1(t'), \ldots, \nu_n(t')\big)\Big] \in \mathcal{I}_{\mathscr{C}}$$

is a regular instance. Since we have assumed that exp is constructible, there must occur the point $\mathbf{p} := [(\exp(t'), 0, 1)^T]$ in the instance associated to $t' \in [0,1]$. When the representative of this point is the vector $\big(\exp(t'), 0, 1\big)^T$, it follows inductively that each component of this representative is algebraic over $\mathbb{Q}[t'] = \mathbb{Q}$ contradicting $\exp(t')$ transcendental over $\mathbb{Q}$. This argumentation applies to all representatives with a transcendental component. Due to this fact we assume that the point is given by a representative $a \cdot \big(\exp(t'), 0, 1\big)^T$ that only has algebraic entries. For this reason, $a \neq 0$ has to be algebraic as well as $a \cdot \exp(t')$. Now, we add the point $[(\exp(t'), 1, 1)^T]$ to the existing construction $\mathscr{C}$ and join it with $\mathbf{p}$. The resulting line is represented by $[(1, 0, -t')^T]$. Since $\mathbf{p}$ lies on this line, it follows that

$$a \cdot (\exp(t') - t') = 0 \iff \exp(t') = t'$$

contradicting again $\exp(t')$ transcendental over $\mathbb{Q}$. □

Corollary 2.3. *The functions $\sin(t), \cos(t), \log(1+t)$ are not constructible.*

Proof. Analogous proof as above. □

The next theorem may be a little bit more surprising. Informally, it states that no other functions than holomorphic ones can be constructed.

Theorem 2.7. *Let $f : [0,1] \to \mathbb{C}$ be a constructible function. Then, for almost every $t \in [0,1]$, there exists an open disc D_t such that $f\big|_{D_t \cap [0,1]}$ is the restriction of a holomorphic function in $\mathcal{O}(D_t)$*

Proof. Immediate consequence of Theorem 2.4. □

2.4.3 The Necessity of Holomorphic Movements

Maybe one would expect that a continuous behavior of the dependent elements of an algebraic construction is already achieved by continuous motions of the free elements. That this is not the case is proved by

Theorem 2.8. *There exists an algebraic construction $\mathscr{C}$, a regular instance $\vec{\mathbf{o}} \in \mathcal{I}_{\mathscr{C}}$ and continuous paths $\mu_k : [0,1] \to O_k$ for $k \in \{1-m, \ldots, 0\}$ with $\mu_k(0) = \mathbf{o}_k$ so that there is no sequence of continuous maps $\nu_l : [0,1] \to O_l$, $l \in \{1, \ldots, n\}$, with $\big(\mu_{1-m}(t), \ldots, \nu_n(t)\big)^T \in \mathcal{I}_{\mathscr{C}}$ for all $t \in [0,1]$.*

Proof. It is sufficient to give a counterexample. Therefore, we have a look at the construction $\mathscr{C}$ in example 2.9. It yields a point of intersection of two given conics and is defined by

$$\begin{aligned} \mathbf{o}_{-1} &\in \mathcal{C} \\ \mathbf{o}_0 &\in \mathcal{C} \\ \mathbf{o}_1 &\in \mathrm{MEETCC}(\mathbf{c}_{-1}, \mathbf{c}_0). \end{aligned}$$

Furthermore, we choose the regular instance $\vec{\mathbf{o}} = (\mathbf{o}_{-1}, \mathbf{o}_0, \mathbf{o}_1)^T$ given by

$$\mathbf{o}_{-1} := \left[(1,1,-1,0,0,0)^T\right], \quad \mathbf{o}_0 := \left[(1,1,0,0,-2,0)^T\right], \quad \mathbf{o}_1 := \left[\left(\frac{1}{2}, \frac{\sqrt{3}}{2}, 1\right)^T\right].$$

The free objects $\mathbf{o}_{-1}$ and $\mathbf{o}_0$ are two circles of radius 1 centered at $(0,0)^T$ and $(1,0)^T$, respectively. The last things that we have to specify are the paths. They are defined to be

$$\mu_{-1} : [0,1] \to \mathcal{C},\ t \mapsto \left[(1,1,-1,0,0,0)^T\right]$$

and

$$\mu_0 : [0,1] \to \mathcal{C},\ t \mapsto \begin{cases} \left[(1,1,4t(t-1),0,4t-2,0)^T\right], & \text{if } t \le \frac{1}{2}, \\ \left[(1,1,4t(t-1),0,0,2-4t)^T\right], & \text{otherwise.} \end{cases}$$

The path μ_0 is constant and therefore the circle $\mathbf{o}_{-1}$ remains at its position. The other path, however, causes a motion of $\mathbf{o}_0$, namely the circle does not change its form, but its position. It moves horizontally towards the origin, first. When the circle arrives the origin, it changes its direction and moves vertically from then on until its center has position $[(0,1,1)^T] \in \mathcal{P}$. The entire dynamic construction is illustrated in Figure 2.6.

To prove that there are no continuous movements for $\mathbf{o}_1$, we show that its behavior at $t = \frac{1}{2}$ has to be discontinuous. It is obvious that the point of intersection $\mathbf{o}_1$ can only move continuously, when it moves along the arc of $\mathbf{o}_{-1}$ towards the point $[(0,1,1)^T] \in \mathcal{P}$ during the first part of the movement of $\mathbf{o}_0$. However, if we consider the two possible points of intersection that can be obtained by running the second part of the movement of $\mathbf{o}_0$ with reverse orientation[18], we will get $[(1,0,1)^T]$ or $[(-1,0,1)^T]$ as possible and continuous alternatives for $\mathbf{o}_1$. This finally proves the claim. □

Remark 2.15. We observe that for $t = \frac{1}{2}$, the two circles $\mathbf{o}_{-1}$ and $\mathbf{o}_0$ coincide and thus each point on them would be a consistent choice for $\mathbf{o}_1$.

[18] Formally, we have to consider the limits for $t \overset{>}{\to} 0.5$, but this is not a problem at all, since we are interested in continuous movements.

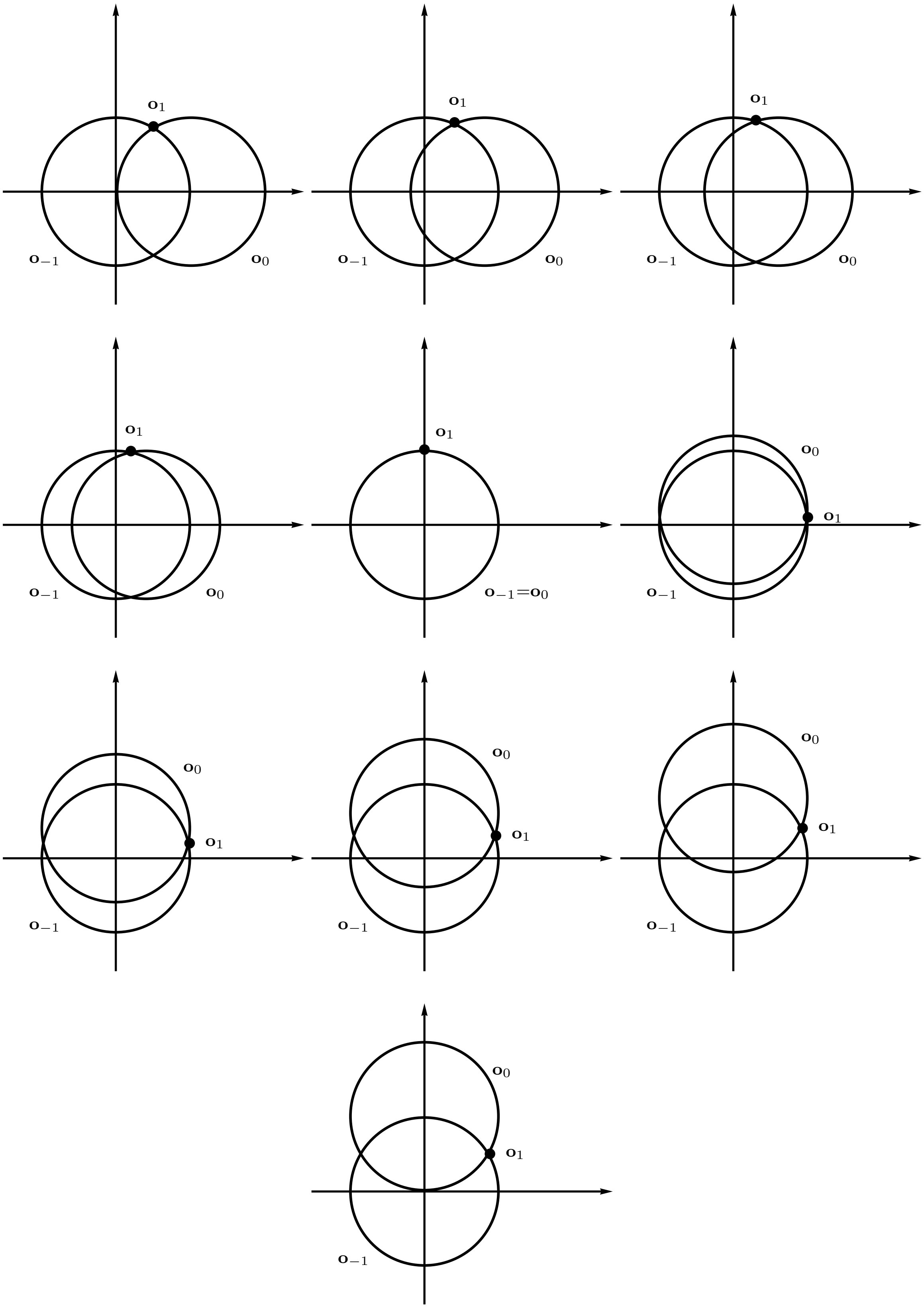

Figure 2.12: Illustration of the proof of Theorem 2.8.

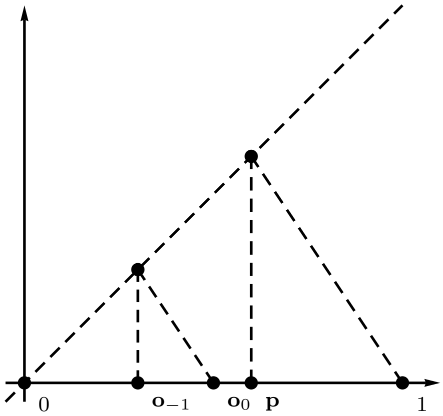

Figure 2.13: Geometric division.

Up to here, we have shown that only continuous input does not work, but it might be enough to rely on k-times differentiable movements in order to achieve continuity concerning the dependent objects. To realize that this does not work as well, we investigate another example.

Example 2.13. As a first step, we introduce three functions, which will be used to build an only k-times differentiable movement. For $k \in \mathbb{N}$, they are given by

$$\begin{aligned} f_1 &: [0,1] \to \mathbb{R}; \quad t \mapsto \left(t - \frac{1}{2}\right)^{k+1}, \\ f_2 &: [0,1] \to \mathbb{R}; \quad t \mapsto \left(t - \frac{1}{2}\right)^{k+2}, \\ f &: [0,1] \to \mathbb{R}; \quad t \mapsto \begin{cases} f_1(t), & \text{if } t \leq \frac{1}{2}, \\ f_2(t), & \text{otherwise.} \end{cases} \end{aligned}$$

It is not hard to see that the function f is only k-times differentiable at $t = \frac{1}{2}$. Using f, we can build an only k-times differentiable movement. The construction that we would like to study has two free points $\mathbf{o}_{-1} \in \mathcal{P}$ and $\mathbf{o}_0 \in \mathcal{P}$ with corresponding movements

$$\mu_{-1} : [0,1] \to \mathcal{P}, \quad t \mapsto \left[(f(t), 0, 1)^T\right]$$

and

$$\mu_0 : [0,1] \to \mathcal{P}, \quad t \mapsto \left[(f_1(t), 0, 1)^T\right].$$

Furthermore, we use the von Staudt-constructions[19] to get a dependent point $\mathbf{p}$ that first coordinate is the quotient of the first coordinates of $\mathbf{o}_{-1}$ and $\mathbf{o}_0$. Fig 2.13 shows a construction that provides this feature. As a result, the motion of $\mathbf{p}$ is described by the map

$$[0,1] \to \mathcal{P}; \quad t \mapsto \left[\left(\frac{f(t)}{f_1(t)}, 0, 1\right)^T\right].$$

[19]These constructions also provides a geometric division as shown in [12].

At this point, we remark that this construction and the chosen movements define implicitly a regular starting instance. The quotient

$$\frac{f(t)}{f_1(t)}$$

equals 1 for $t \in [0, \frac{1}{2}]$ and is equal $(t - \frac{1}{2})$ for $t \in]\frac{1}{2}, 1]$. Due to this fact this quotient cannot be continuous at $t = \frac{1}{2}$, since it holds

$$\lim_{t \to \frac{1}{2}} \left(t - \frac{1}{2} \right) = 0.$$

However, this in turn implies that the motion of the point $\mathbf{p}$ cannot be continuous, either.

According to the previous example, we can conclude that only k-times differentiable movements are not sufficient to achieve continuity concerning the dependent objects of an algebraic construction in general. This is formalized in

Theorem 2.9. *There exists an algebraic construction $\mathscr{C}$, a regular instance $\vec{\mathbf{o}} \in \mathcal{I}_{\mathscr{C}}$ and k-times differentiable paths $\mu_j : [0,1] \to O_j$ for $j \in \{1-m, \ldots, 0\}$ with $\mu_j(0) = \mathbf{o}_j$ so that there is no sequence of continuous maps $\nu_l : [0,1] \to O_l$, $l \in \{1, \ldots, n\}$, with $\left(\mu_{1-m}(t), \ldots, \nu_n(t)\right)^T \in \mathcal{I}_{\mathscr{C}}$ for all $t \in [0,1]$.*

Remark 2.16. This theorem states not only that a CDGS cannot be based on k-times differentiable movements, but also restricts the set of possible geometric operations. Basically, the output of a general operation under an h-movement must not be an only k-times differentiable map, since this would cause a discontinuous behavior according to Theorem 2.9.

3

Complexity and Dynamic Geometry

"Zu jeder ω-widerspruchsfreien rekursiven Klasse κ von Formeln *gibt es rekursive* Klassenzeichen r*, so daß weder* v Gen r *noch* Neg(v Gen r) *zu* Flg(κ) *gehört (wobei* v *die* freie Variable *aus* r *ist)."*

Found in Gödel's article [21]

The quote of Kurt Gödel is actually the 6-th Theorem from his famous article [21], in which he proved the existence of undecidable theorems in a self-consistent recursive axiomatic system that has got the power to describe the arithmetics of the natural numbers. So that, he opened the door to many undecidability results that have affected almost all fields in mathematics and computer sciences. A well-known example of such a result is the undecidability of Hilbert's 10th problem, which was proved by Davis, Putnam, Robinson and Matiyasevich [37, 45]. The problem is to decide, whether a Diophantine equation with integer coefficients is solvable over the integers or not. Apart from undecidability, decidable problems have taken increasingly center stage. Many of them have been classified according to their intrinsic complexity as well. Thereby, the complexity is measured by the effort depending on the input length one needs to solve the problem. The common complexity classes that we are going to use in this thesis are NP and coNP. NP is the class of decision problems, for which the instances that have the answer YES can be verified efficiently, this means by a deterministic Turing machine in polynomial time in terms of the input length. An example for such a problem is the Boolean Satisfiability Problem. The problem is to determine, whether a boolean formula is satisfiable. Whereas coNP is the class of decision problems, for which the instances that have the answer NO can be verified efficiently. Deciding, whether a boolean formula is unsatisfiable, is a problem from coNP, for example. For a detailed introduction and extensive discuss of this classical part of complexity theory, we refer to [20, 40]. Although, the classes NP and coNP cover many important problems, many other interesting problems have arisen in the last decades. For instance, the classical theory has been transfered to more general number fields as the reals or complex numbers to provide a more adequate framework for problems from numerical analysis (see [4]). Similarly, complexity theory has also found its way into geometry. An effective way to prove complexity theoretic results in geometry is to transfer the arithmetic operations of the reals or complex numbers to geometry by special constructions and use them to encode polynomials. This approach enabled Richter-Gebert and Kortenkamp to prove the first complexity results concerning continuous dynamic geometry in [48] .

3.1 Introduction

We now concentrate on a complexity theoretic question arising in dynamic geometry. Starting from a given geometric construction and a corresponding starting instance, one can ask, whether the free elements and the dependent elements can be moved continuously so that the given instance is deformed in another specific instance of the construction. Thereby, we require that only regular instances occur during this motion. This additional requirement is reasonable, since then the ambiguity that can occur during a motion is eliminated. Informally, we call this the *Reachability Problem in Dynamic Geometry.* A more formal definition of this problem is given below.

Reachability Problem in Dynamic Geometry. Given a construction $\mathscr{C}$ as in Definition 2.3 and a regular instance $\vec{\mathbf{o}} = (\mathbf{o}_{1-m}, \ldots, \mathbf{o}_n) \in \mathcal{I}_{\mathscr{C}}$ of $\mathscr{C}$. Additionally, there is a second given regular instance $\vec{\mathbf{q}} = (\mathbf{q}_{1-m}, \ldots, \mathbf{q}_n) \in \mathcal{I}_{\mathscr{C}}$ of $\mathscr{C}$. The problem of deciding, whether there exist a corresponding CDGS $\mathcal{D}$ and an associated function $\mu \in \Phi$ so that $\big(\mu(t)^T, \nu(t)^T\big)$ is a regular instance of $\mathscr{C}$ for all $t \in [0,1]$ and $\big(\mu(1)^T, \nu(1)^T\big) = \vec{\mathbf{q}}$ is called the *Reachability Problem in Dynamic Geometry*, short GEO-REACH.

In our studies concerning geometric complexity theory, we will only allow three quite simple sets of operations besides free ones in order to define constructions. These three sets are given by

$$\begin{aligned} QUAD &:= \{\mathrm{CONST}_{\mathbf{p}}, \mathrm{MEET}, \mathrm{JOIN}, \mathrm{MEETCL}\}, \\ KUBI &:= \{\mathrm{CONST}_{\mathbf{p}}, \mathrm{MEET}, \mathrm{JOIN}, \mathrm{MEETCL}, \mathrm{MEETKL}\}, \\ TRAN &:= \{\mathrm{CONST}_{\mathbf{p}}, \mathrm{MEET}, \mathrm{JOIN}, \mathrm{WHEEL}\}. \end{aligned}$$

When a construction $\mathscr{C}$ is only built of operations from a specific set S, we will emphasize that by adding the words *over* S. We have seen in the previous chapter that knowing the coordinates of the geometric objects under a family of h-movements is sufficient to determine the behavior of a CDGS. So providing operations that can be used to derive coordinates enables us to set up an equivalent model focused on coordinates.

In [12, 30, 48], it is discussed that points and lines together with MEET and JOIN can be used to imitate addition, subtraction, multiplication and division and vice versa. These results are based on the so-called VON-STAUDT constructions (see [59, 60]). Moreover, the transition from constructions consisting of MEET, JOIN and free objects to straight-line programs is stated explicitly in [30]. The operations MEETCL, however, enable us to rebuild the square root. A proof of this and more on the power of geometric constructions can be found in [36]. So, loosely speaking, the first set $QUAD$ is as powerful as solving quadratic polynomials. In contrast to $QUAD$, the set $KUBI$ contains a further operation MEETKL, which describes the intersection of an algebraic curve of degree 3 and a line. This operation can be used to construct 3-rd roots as shown in Fig. 3.14. Moreover, it is sufficient to use the 3-rd root in order to compute the corresponding points of intersection, since the greatest degree that occurs in the corresponding polynomial system is 3. As a result, this operation is as powerful as solving polynomials of degree 3 and therefore the set $KUBI$ is as powerful as solving polynomials of degree 3. The last set $TRAN$ has a transcendental aspect caused by the operation WHEEL. It is obvious that one needs the complex logarithm to determine the output set of the relation WHEEL. For this reason, the power of the set $TRAN$ is equivalent to the arithmetic operations and the complex logarithm.

Below we will, additionally, focus on the problem in $\mathbb{C}$, since non-trivial lower complexity bounds for GEO-REACH have already been established in $\mathbb{R}$. RICHTER-GEBERT and KORTENKAMP proved in their article [48] that GEO-REACH over a set similar to $QUAD$ is NP-hard in $\mathbb{R}$. However, there are just few results concerning $\mathbb{C}$. In [11], it was shown that the problem is basically decidable for constructions over $QUAD$. And in [31], it was proved that GEO-REACH over $QUAD$ is at least as hard as testing, whether a generalized

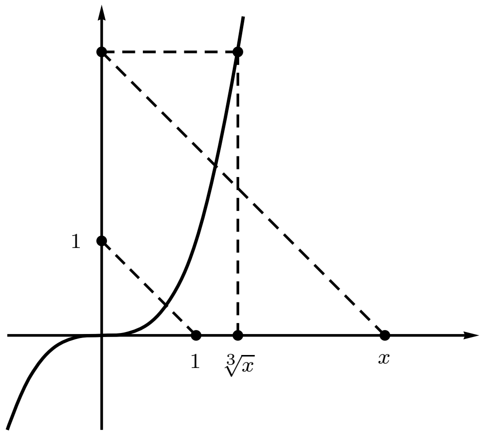

Figure 3.14: Constructing a 3-rd root of x can be achieved by JOIN, MEET and intersecting the curve $y = x^3$ with lines.

version of a straight-line program, more precisely, a straight-line program built of addition, subtraction, multiplication, division and square roots, vanishes identically.

As mentioned previously, our observations will focus on constructions that are built of operations from $QUAD$, $KUBI$ or $TRAN$. We have pointed out that constructions over $QUAD$, $KUBI$ or $TRAN$ are equivalent to algebraic pendants, thus we are going to deal with them instead of the geometric ones as done in [11, 48]. So our first step will be converting the problem GEO-REACH into a problem that is based on a generalization of straight-line programs. The idea of generalizing straight-line programs is not new in this context. Similar to several other works [11, 12, 31], the setup used in this chapter relies on the model presented in [48]. Following the approach in [48], the continuous dynamics of a construction are captured by so-called continuous evaluations of the just mentioned generalized straight-line programs. We will observe that continuous evaluations can be identified with analytic continuations of associated complex functions. Based on this setting, we are going to study the reachability concerning instances of straight-line programs. By doing so, several lower complexity bounds on weakened variants of this problem are going to be proved. Thereby, it turns out that the set of admitted operations of the underlying straight-line program is crucial for the complexity of a rechability problem where only parts of a terminal instance are specified, namely roots will enable us to prove results concerning NP and coNP, while logarithms will yield undecidability. The main idea in our reductions is to encode the problem, which shall be reduced, into a set of germs and associated RIEMANN surfaces. Then the sheets of these surfaces serve as counters and switches that are operated by analytic continuations.

In contrast to [48], we either restrict the movements of the free elements or do only prescribe parts of the terminal instance to prove complexity bounds. The second problem can be understood as a reduction of the information content of the terminal instance. Roughly speaking, we treat parts of straight-line programs as black boxes.

3.2 Straight-Line Programs, Continuous Evaluations and the Reachability Problem

The notion of a construction introduced in the previous chapter is closely related to the common concept of straight-line programs. The reason behind this is that they have an analogous structure. A construction is mainly defined by a sequence of construction steps, while a straight-line program is determined by a sequence of elementary computations.

Usually, straight-line programs are used to encode polynomials by stating a sequence of arithmetic operations on input variables and intermediate results. This encoding offers several useful features. For example, one does not have to refer to coefficients of a polynomial. Another amazing fact is that many polynomials of exponential degree can be described by a straight-line program of just linear length. In algebraic complexity theory, straight-line programs are one of the most fundamental and important models of evaluation. They have turned out to be useful as complexity measure for geometric questions related to polynomial systems (see [33]). Moreover, they play an essential role in SMALE's 4th problem in [56]. A good source for more results on straight-line programs is the book [6].

In the following, we will define our notion of straight-line programs (SLPs) over the complex numbers and over a fixed set of operations. In this way, we bridge the gap from geometric constructions using MEET, JOIN, $\mathrm{CONST}_{\mathbf{p}}$, MEETCL, MEETKL and WHEEL to a pure algebraic and equivalent model over the complex numbers.

As a first step, we have to transfer the idea of geometric objects and operations. Instead of several different types of geometric objects, we will only deal with one object type, namely the complex numbers $\mathbb{C}$. They play the role of the geometric objects introduced in Definition 2.1. Thereafter, we require operations that act on the complex numbers. A general *SLP-operation* is defined as follows:

Definition 3.10 (SLP-operation)**.** *Let $k \in \mathbb{N}$. Then we call a relation*

$$\omega \subseteq \underbrace{\mathbb{C} \times \ldots \times \mathbb{C}}_{k-times} \times \mathbb{C}$$

an SLP-operation*. Thereby, we set again* $\mathrm{in}(\omega) := k$*. Furthermore, we define the output set*

$$\omega(z_1, \ldots, z_k) := \big\{ w \in \mathbb{C} \bigm| (z_1, \ldots, z_k, w) \in \omega \big\}$$

of ω. As extension, we call $\omega = \mathbb{C}$ an input *or* free operation*. We set* $\mathrm{in}(\omega) := 0$ *in this case.*

Remark 3.17. Moreover, we assume that both SLP-operations and geometric operations are performed at unit costs.

Since we are basically interested in operations that are equivalent to MEET, JOIN, $\mathrm{CONST}_{\mathbf{p}}$, MEETCL, MEETKL and WHEEL, we restrict ourselves to the following SLP-operations: First of all, we need the basic arithmetic operations

$$\begin{aligned} + &:= \{(z_1, z_2, w) \in \mathbb{C}^3 \mid z_1 + z_2 = w\}, \\ - &:= \{(z_1, z_2, w) \in \mathbb{C}^3 \mid z_1 - z_2 = w\}, \\ * &:= \{(z_1, z_2, w) \in \mathbb{C}^3 \mid z_1 \cdot z_2 = w\}, \\ / &:= \{(z_1, z_2, w) \in \mathbb{C}^3 \mid z_1 = z_2 \cdot w \wedge z_2 \neq 0\}. \end{aligned}$$

As already mentioned, the arithmetic operations are enough to govern MEET and JOIN. Apart from them, we will need complex constants. A constant number $z \in \mathbb{C}$ is captured by

$$\mathrm{const}_z := \{z\}.$$

Last but not least, we introduce the relational pendants that are equivalent to MEETCL, MEETKL and WHEEL. They are the complex square root, 3rd root and the complex logarithm that are given by

$$\begin{aligned} \surd &:= \{(z,w) \in \mathbb{C}^2 \mid z = w^2\}, \\ \sqrt[3]{} &:= \{(z,w) \in \mathbb{C}^2 \mid z = w^3\} \end{aligned}$$

and

$$\mathscr{L} \;:=\; \{(z,w) \in \mathbb{C}^2 \mid z = e^w \wedge z \neq 0\}.$$

Remark 3.18. It is obvious that the ambiguity of the square root and the logarithm is encapsulated in the relation encoding them. For instance, it holds $\surd(1) = \{\pm 1\}$, $\sqrt[3]{}(1) = \{1, e^{\frac{2\pi i}{3}}, e^{\frac{4\pi i}{3}}\}$ as well as $\mathscr{L}(1) = \{2\pi k \in \mathbb{C} | k \in \mathbb{Z}\}$. The output set of the other operations, however, consists only of one element. Hence we call these operations *determined* and the square root, 3rd root and logarithm *not determined*. In addition, we shall remark that we only allow well-defined situations in the definitions of our SLP-operations. More precisely, divisions by zero and logarithms of zero are excluded in the definitions of / and $\mathscr{L}$.

Mainly, we will refer to different sets of operations throughout this chapter, thus we introduce besides the free ones the following sets of SLP-operations:

$$\begin{aligned} \mathit{quad} &:= \{\mathrm{const}_z \mid z \in \mathbb{C}\} \cup \{+, -, *, /, \surd\}, \\ \mathit{kubi} &:= \{\mathrm{const}_z \mid z \in \mathbb{C}\} \cup \{+, -, *, /, \surd, \sqrt[3]{}\}, \\ \mathit{tran} &:= \{\mathrm{const}_z \mid z \in \mathbb{C}\} \cup \{+, -, *, /, \mathscr{L}\}. \end{aligned}$$

Furthermore, we set $\aleph := \{\mathit{quad}, \mathit{kubi}, \mathit{tran}\}$ in order to simplify the notation and proceed with defining our notion of straight-line programs.

Definition 3.11 (straight-line programs). *Let $p \in \mathbb{N}$ and $q \in \mathbb{N}_0$. A* straight-line program, short SLP, $\mathscr{P}$ over $O \in \aleph$ *consists of*

(i) a sequence $\omega_{1-p}, \dots, \omega_q$ of SLP-operations with $\omega_k \in O$ for all $k \in \{1, \dots, q\}$ and $\omega_k = \mathbb{C}$ for all $k \in \{1-p, \dots, 0\}$ (free operations),

(ii) input assignments $\eta_k = (s_1^k, \dots, s_{\mathrm{in}(\omega_k)}^k) \in \mathbb{Z}^{\mathrm{in}(\omega_k)}$ *($k \in \{1, \dots, q\}$) such that $s_l^k < k$ for $l \in \{1, \dots, \mathrm{in}(\omega_k)\}$.*

The previous definition is analogous to the introduced definition of constructions in the previous chapter. It is obvious that the first p operations of an SLP are free and therefore they represent the input variables of it. For $k \in \{1, \dots, q\}$, each ω_k encodes a single computation, which input, in turn, is determined by the input assignments η_k. They serve as a pointer that refers to the appropriate outputs of other operations. The additional relation the input assignments are satisfying ensures that only values are used that are either input data or intermediates that have already been computed.

In order to become familiar with the definition above, we will have a look at a simple example of an SLP.

Example 3.14. To increase readability, we skip the formal notion and use instead an analogous notation to that introduced in Example 2.9.

$$\begin{aligned}
z_0 &\in \mathbb{C}\\
z_1 &\in \surd(z_0)\\
z_2 &\in \omega_1\\
z_3 &\in +(z_0, z_2)\\
z_4 &\in /(z_1, z_3)
\end{aligned}$$

The example illustrates the similarities and differences of original straight-line programs and SLPs in our sense. Just as original straight-line programs, the SLP in the example can be interpreted as an encoding of a special function. In this particular case, it is the function induced by

$$z \mapsto \frac{\sqrt{z}}{z+1}.$$

So we can always associate to each SLP a special function. Apart form this similarity, there are essential differences as well. Maybe the most important difference comes from the different sets of operations, which are used to build them. Originally, only the arithmetic operations have been permitted, while we also allow roots and logarithms. They bring up some non-determinism. For instance, if z_0 is set equal 1, one can derive $\frac{1}{2}$ as well as $-\frac{1}{2}$ for z_4. This depends on the output of the not determined operation $\surd$, which occurs in the numerator of the encoded function. Another point we should also mention is the fact that an SLP might not be defined for all $z \in \mathbb{C}$. There can be divisions by zero or the argument of a logarithm can be zero as well. For instance, if $z_0 = -1$ then a division by zero occur in the example. Last but not least, we address another similarity, this is, not all assignments of complex numbers satisfy the relations established by an SLP. For instance, if one sets $z_0 = 1$ then the only two choices for z_1 that satisfy the corresponding relation $z_0 = z_1^2$, are ± 1. Such assignments are of particular interest for us. An assignment of complex numbers so that all relations are satisfied is called an *SLP-instance*. More precisely, we get

Definition 3.12 (SLP-instances)**.** *Let $\mathscr{P}$ be an SLP over $O \in \aleph$ as in Definition 3.11. Then an assignment of complex numbers $Z = (Z_{1-p}, \dots, Z_q) \in \mathbb{C}^{p+q}$ with*

$$\left(Z_{s_1^k}, \dots, Z_{s^k_{\mathrm{in}(\omega_k)}}, Z_k\right) \in \omega_k$$

for all $k \in \{1, \dots, q\}$ is called an SLP-instance *of $\mathscr{P}$. In addition, we call Z* regular*, if it holds that $Z_{s_1^k} \neq 0$ for all $k \in \{1, \dots, q\}$ with $\omega_k \in \{\surd, \sqrt[3]{}\}$.*

Let us go a little bit more into detail and explain, what regularity achieves at SLP-instances. Basically, regularity ensures that the arguments of roots, i.e. $\surd$ and $\sqrt[3]{}$, are unequal zero. This is, in turn, crucial to eliminate the ambiguity that can occur during a movement of the input objects of an SLP. For this reason, we will claim later on that only regular instances are run through during such a movement. In a certain sense, this is analogous to the requirement that we have also claimed at Geo-Reach, where regular instances are also only admitted. Moreover, it is also worthwhile to mention that divisions by zero as well as logarithms of zero are excluded at SLP-instances. This follows directly from the definitions of an instance and the operations $/, \mathscr{L}$.

Our next goal is to model a continuous behavior of an SLP under a continuous movement of the input objects. Therefore, we introduce

Definition 3.13 (SLP-movements)**.** *Let $\mathscr{P}$ be an SLP over $O \in \aleph$ as in Definition 3.11. For $k \in \{1-p, \dots, 0\}$, let $\mu_k : [0,1] \to \mathbb{C}$ be continuous mappings. Then we call*

$$\mu : [0,1] \to \mathbb{C}^p;\ t \mapsto \Big(\mu_{1-p}(t), \dots, \mu_0(t)\Big)$$

an SLP-movement *of $\mathscr{P}$, if all SLP-instances that are induced by $\mu(t)$ for all $t \in [0,1]$ are regular.*

The last definition ensures two things that hold during an SLP-movement. Firstly, all computations are well-defined, and secondly, no branch point occurs in the computations. This is important to eliminate the choices, one has to leave a branch point continuously again. We have observed that such ambiguity occurs, when a CDGS passes through a degenerate situation. Based on this, we can define the concept of continuous evaluations of an SLP, which enables us to model continuous behavior of an SLP under a movement of the input elements.

Definition 3.14 (continuous evaluations of an SLP)**.** *Let $\mathscr{P}$ be an SLP over $O \in \aleph$ as in Definition 3.11 and μ an SLP-movement of $\mathscr{P}$. A* continuous evaluation *of $\mathscr{P}$* under *μ is an assignment of continuous mappings $c_1, \dots, c_q$, where $c_k : [0,1] \to \mathbb{C}$ for $k \in \{1, \dots, q\}$, so that*

$$\Big(\mu_{1-p}(t), \dots, \mu_0(t), c_1(t), \dots, c_q(t)\Big)$$

forms an SLP-instance of $\mathscr{P}$ for all $t \in [0,1]$.

Remark 3.19. We point out that only regular instance can occur during a continuous evaluation of an SLP.

Now, we have developed everything we need in order to state formally the Reachability Problem for SLPs.

Reachability Problem for SLPs. Let $\mathscr{P}$ be an SLP over $O \in \aleph$. Moreover, $Z, W \in \mathbb{C}^{p+q}$ are regular instances of $\mathscr{P}$. Then the problem of deciding, whether there exist a movement μ and an associated continuous evaluation of $\mathscr{P}$ with corresponding functions $c_1, \dots, c_q$ so that

$$\Big(\mu_{1-p}(0), \dots, \mu_0(0), c_1(0), \dots, c_q(0)\Big) = Z \text{ and } \Big(\mu_{1-p}(1), \dots, \mu_0(1), c_1(1), \dots, c_q(1)\Big) = W$$

hold, is called the *Reachability Problem for SLPs over O*, short SLP-Reach over O.

Remark 3.20. We often omit the prefix SLP at SLP-operations and SLP-instances in the following, since we only deal with SLP-operations and SLP-instances rather than geometric operations and geometric instances.

3.3 Continuous Evaluations, Holomorphic Functions and Their Analytic Continuations

Throughout the rest of this chapter, we introduce another restriction, namely we only study SLPs with one input variable. So let $\mathscr{P}$ be such an SLP over $O \in \aleph$ and $Z := \big(Z_0, \dots, Z_q\big) \in \mathbb{C}^{q+1}$ be an associated regular instance of $\mathscr{P}$. Furthermore, we denote the set of complex numbers, which induce a non-regular instance of $\mathscr{P}$ by $S \subset \mathbb{C}$. Since we will only be interested in SLPs, for which this set is discrete, we additionally assume that S is a discrete subset of $\mathbb{C}$. Starting from this setting, we can prove that an SLP can be described locally by holomorphic functions.

Lemma 3.12. *There is an open neighborhood $U \subset \big(\mathbb{C} \setminus S\big)$ of $Z_0 \in \mathbb{C}$ and holomorphic functions $f_0, \ldots, f_q$ on U that satisfy*

$$f_k : U \to \mathbb{C}; \quad z \mapsto w \in \omega_k\Big(z_{s_1^k}, \ldots, z_{s^k_{\mathrm{in}(\omega_k)}}\Big)$$

for $k \in \{0, \ldots, q\}$.

Proof. We prove the claim by induction over $k \in \{0, \ldots, q\}$. For $k = 0$, we set f_k equal the identity on $\mathbb{C} \setminus S$. Now, let $k > 0$. The easy case is $\omega_k = \mathrm{const}_w$. In this case, we set $f_k \equiv w$, which is an entire function. If ω_k is one of the arithmetic operations, then f_k is an addition, subtraction, multiplication or division of two preceding functions. By induction hypothesis, these two functions are holomorphic on an open neighborhood $\tilde{U}$ of Z_0. The regularity of the instance Z together with the discreteness of the set S ensures that there is an open neighborhood $U \subset \tilde{U}$, on which f_k is holomorphic. The last cases left are $\omega_k \in \{\sqrt{\ }, \sqrt[3]{\ }, \mathscr{L}\}$. In all three cases, we have a composition of a preceding function, which is holomorphic on an open neighborhood $\tilde{U}$ of $Z_0 \in \mathbb{C}$, and the square root, 3rd root or logarithm. The regularity of Z and the discreteness of S guarantees again that f_k is holomorphic on a proper open neighborhood $U \subset \tilde{U}$ of Z_0. Finally, the claim follows by the fact that the intersection of finitely many open sets is open again. $\square$

We now assume that there is a continuous evaluation under a movement $\mu : [0,1] \to \mathbb{C}$. By definition, there are associated continuous functions $c_1, \ldots, c_q : [0,1] \to \mathbb{C}$ completing μ to an instance of a corresponding SLP. In addition, there occur only regular instances along a movement. Moreover, we have shown in the last lemma that there are open neighborhoods U^z and holomorphic functions $f_0^z, \ldots, f_q^z$ for each $z \in \mu([0,1]) \subset \big(\mathbb{C} \setminus S\big)$ that described the associated instances locally, i.e.

$$c_k(t) = f_k^{\mu(t)}\big(\mu(t)\big)$$

for $k \in \{1, \ldots, q\}$. As $\mu([0,1])$ is compact, there is a finite subcover $\{U^{z_1}, \ldots, U^{z_n}\}$ of $\{U^z\}_{z \in \mu([0,1])}$ still covering $\mu([0,1])$. Due to this fact we can conclude that the intersections $U^{z_i} \cap U^{z_{i+1}}$ are not empty for $i \in \{1, \ldots, n-1\}$. For $k \in \{1, \ldots, q\}$, the continuity of c_k and the holomorphy of $f_k^{z_i}$ and $f_k^{z_{i+1}}$ imply $f_k^{z_i} \equiv f_k^{z_{i+1}}$ on $U^{z_i} \cap U^{z_{i+1}}$, which shows that $f_k^{\mu(1)}$ is the analytic continuation of $f_k^{\mu(0)}$ along μ.

Remark 3.21. Like at geometric constructions, we have discovered that continuous evaluations can be described by appropriate holomorphic functions and their analytic continuations. As a result, we can identify continuous evaluations with appropriate analytic continuations and vice versa. That we actually encounter the same tool to describe continuous evaluations shall not be surprising, since our aim was to develop an equivalent concept to geometric constructions and CDGSs.

3.4 Reducing the Information Content of the Terminal Instance

Our next goal is to prove several complexity theoretic results for the reachabilty problem, where only parts of the terminal instance are prescribed. In order to derive results concerning NP and coNP, we are going to transfer 3SAT formulas and truth assignments into functional pendants that can be encoded by SLPs over *quad*. In a second step, we will tackle some technical lemmata concerning the analytic continuations of these functions. Furthermore, a truth assignment is identified with a special analytic continuation along a closed path starting and ending at $0 \in \mathbb{C}$. In the end, we are able to show that a 3SAT

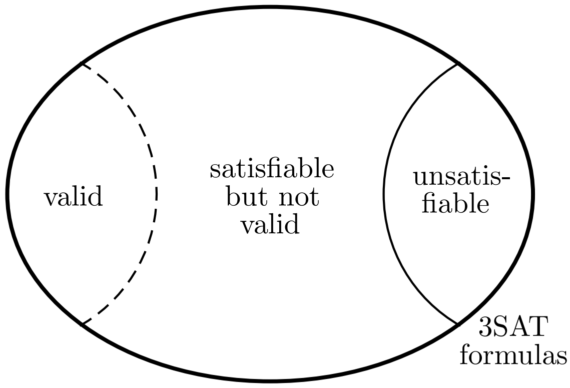

Figure 3.15: The geography of 3SAT formulas.

formula is satisfiable, if and only if the mentioned analytic continuation has a zero at $0 \in \mathbb{C}$. These are the parts of a terminal instance that we are going to prescribe in advance, since stating the rest of the instance would amount to solving the underlying 3SAT. To obtain undecidability results, we are going to perform an analogous construction using logarithms instead of square roots. Similarly, we ask here, whether there is a path so that the associated analytic continuation ends up with a zero at $0 \in \mathbb{C}$.

3.4.1 Results Concerning NP and coNP

As the title of this subsection already indicates, the first part of our lower bounds concerns NP and coNP. For this reason, we introduce the NP-complete problem 3SAT and its coNP-complete pendant co3SAT, before we develop the required tools to prove the bounds.

3SAT and co3SAT. Let $B = \{b_1, \ldots, b_n\}$ be a set of boolean variables and define the literals over B to be $\tilde{B} = \{b_1, \ldots, b_n, \neg b_1, \ldots, \neg b_n\}$. Now, let $C_1, \ldots, C_m$ be clauses formed by disjunction of three literals from $\tilde{B}$, that is,

$$C_j = l_{j,r} \vee l_{j,s} \vee l_{j,t},$$

for each $j = 1, \ldots, m$, where $l_{j,k} \in \{b_k, \neg b_k\}$. A 3SAT formula C is then the conjunction of clauses $C_1, \ldots, C_m$. In symbols, we get

$$C = C_1 \wedge \ldots \wedge C_m.$$

Moreover, a truth assignment χ for B is an n-tuple $(b_1, \ldots, b_n) \in \{\text{TRUE, FALSE}\}^n$. The problem of deciding, whether there is a truth assignment χ that satisfies C is called 3SAT. The complexity of 3SAT is proved in [20, 40] and stated below.

Theorem 3.10. *3SAT is NP-complete.*

Let C be again a 3SAT formula and χ a corresponding truth assignment. The problem of deciding, whether C is unsatisfiable is called co3SAT. To illustrate the connection between 3SAT and co3SAT, the set of 3SAT formulas is represented as shown in Fig. 3.15. Then in terms of this illustration, the problem 3SAT is to decide, whether a formula belongs to the areas labeled with *satisfiable but not valid* or *valid*. A formula that is unsatisfiable lies in the complement labeled with *unsatisfiable*. So the problem co3SAT is to decide, whether a formula belongs to this complement labeled with *unsatisfiable*. In [40], it is shown that the complexity of the complement of an NP-complete problem is, in turn, coNP-complete. As a result, we get

Theorem 3.11. *co3SAT is coNP-complete.*

From 3SAT Formulas to Germs

In the following, we present the construction of a function that encodes a 3SAT formula. Thereby, each symbol of a clause is converted to a multiplication or one of two terms involving square roots. The conversions are listed below explicitly:

$$\begin{aligned} \vee &\quad\to\quad \cdot \\ b_k &\quad\to\quad \frac{\sqrt{k}-\sqrt{k+z}}{2\sqrt{k}} \\ \neg b_k &\quad\to\quad \frac{\sqrt{k}+\sqrt{k+z}}{2\sqrt{k}} \end{aligned}$$

Obviously, there exists a sufficiently small open neighborhood of $0 \in \mathbb{C}$, on which these functions are holomorphic. Thus we can regard these terms as well-defined holomorphic functions around $0 \in \mathbb{C}$, i.e. they are germs with base $0 \in \mathbb{C}$. We should also point out that the definitions of the germs just introduced refer to the principal branch of a square root. Due to this fact a clause $C_j = l_{j,r} \vee l_{j,s} \vee l_{j,t}$ finally becomes a germ

$$X_j := \prod_{k\in\{r,s,t\}} x_{j,k} \in \mathbb{O}_0, \tag{3.6}$$

where $x_{j,k}$ is either induced by

$$\frac{\sqrt{k}-\sqrt{k+z}}{2\sqrt{k}} \quad \text{or} \quad \frac{\sqrt{k}+\sqrt{k+z}}{2\sqrt{k}}$$

depending on $l_{j,k} \in \{b_k, \neg b_k\}$. Consequently, a whole 3SAT formula C is represented by

$$X_C[z] := \big(X_1[z], \ldots, X_m[z]\big).$$

Remark 3.22. We emphasize that throughout this chapter, definitions involving roots and logarithms basically refer to principal branches.

Zeros and Branch Points

We have a closer look at the previously introduced germs in this subsection. It is not hard to see that the germ $x_{j,k} \in \mathbb{O}_0$ induces a germ for all $z \in \mathbb{C} \setminus \{-k\}$. As there is only one non-constant root in its definition, its associated RIEMANN surface has two sheets and its only branch point in $\mathbb{C}$ lies at $z = -k$.[20]

Apart from the global structure, the zeros of these functions are of particular interest for us, since we are aiming for them. Hence let $x_{j,k} \in \mathbb{O}_0$ be defined as previously, i.e. either induced by

$$\frac{\sqrt{k}-\sqrt{k+z}}{2\sqrt{k}} \quad \text{or} \quad \frac{\sqrt{k}+\sqrt{k+z}}{2\sqrt{k}}. \tag{3.7}$$

[20]A presentation of the theory, with which these observations can be verified, can be found in [34].

It is not hard to observe that both terms can only vanish, if two conditions hold simultaneously. On the one hand, z must equal $0 \in \mathbb{C}$ and on the other hand, both the sign in the numerator connecting the roots and the sign associated to the branch of the non-constant second square root in the numerator must differ, since we have chosen the principal branch in (3.7) by definition. As a result, it holds $x_{j,k}[0] = 0$ for the left definition and $x_{j,k}[0] = 1$ for the other one in (3.7). Next, we study the zeros that occur by continuing these two germs analytically in a second step. There might be other ones on their RIEMANN surfaces. According to the fact that the underlying RIEMANN surface has two sheets, there are only two essentially different groups of paths that we must take into account, this is, paths that have an odd winding number and paths that have an even winding number w.r.t. the branch point of $x_{j,k} \in \mathbb{O}_0$. An analytic continuation of $x_{j,k}$ along a path with an odd winding number changes the branch, while a path with an even winding number does not. Consequently, the analytic continuation $x_{j,k}^{\gamma} \in \mathbb{O}_0$ of $x_{j,k}$ along a closed path γ starting and ending at $z = 0$ with an odd winding number w.r.t. $-k$ yields $x_{j,k}^{\gamma}[0] = 1$ in the first case ((3.7) left) and $x_{j,k}^{\gamma}[0] = 0$ otherwise ((3.7) right). For this reason, we have encountered a zero for the germ induced by the right term in (3.7).

The last paragraphs have shown that the interplay of the sign in the numerator and the branch of the non-constant square root are crucial for zeros. In order to capture this, we introduce a new notation that reflects the branch a germ lies on. Therefore, we additionally use two upper indices. The first one denotes the sign between the roots in the numerator and a second index indicates the branch a germ lies on. For both indices, we use the symbols $+$ and $-$. Hereby, the first index is self-explanatory, while a $+$ in the second index stands for the principal branch of the square root. For instance, the germ $x_{j,k}^{-,+}$ refers to the germ that is induced by the left term in (3.7). By definition, it refers to the principal branch of the second square root in the numerator. We can summarize the previous insights in the following lemma.

Lemma 3.13. *Let $r, s \in \{+, -\}$ and $x_{j,k}^{r,s} \in \mathbb{O}_0$ be a corresponding germ in the previous sense. Then the top of the germ $x_{j,k}^{r,s}$ at $z = 0$ satisfies*

$$x_{j,k}^{r,s}[0] = \begin{cases} 0, & \text{if } r \neq s \\ 1, & \text{if } r = s \end{cases} .$$

An immediate consequence of the last lemma is

Corollary 3.4. *Given $j \in \{1, \ldots, m\}$ and a path γ starting at 0 and ending at z. Moreover, there exists the analytic continuation $X_j^{\gamma} \in \mathbb{O}_z$ of $X_j \in \mathbb{O}_0$ along γ. Then $X_j^{\gamma}[z] = 0$ implies $z = 0$.*

Proof. Since X_j^{γ} is the product $x_{j,r}^{\gamma} \cdot x_{j,s}^{\gamma} \cdot x_{j,t}^{\gamma}$, it follows that X_j^{γ} vanishes, if and only if one of its factors vanishes. According to Lemma 3.13, this can only be for $z = 0$. □

From Truth Assignments to Closed Paths and Analytic Continuations

As mentioned, we identify a truth assignment with analytic continuations of $X_1, \ldots, X_m \in \mathbb{O}_0$ along a closed path starting and ending at $z = 0$. Basically, the idea is to encode a value of a boolean variable b_k in the winding number of the path w.r.t. to the branch point $-k$. Therefore, let $\chi = (b_1, \ldots, b_n) \in \{\text{TRUE}, \text{FALSE}\}^n$ be a truth assignment for B. Starting from χ, we construct a closed path γ starting and ending at $z = 0$ that satisfies the following condition: The winding number $\eta(\gamma, -k)$ of γ w.r.t $-k$ should be even, if b_k is TRUE and odd, if b_k is FALSE. It is clear that such a path and the associated analytic continuations of $X_1, \ldots, X_m$ exist, since the set $\{-n, \ldots, -1\}$ of branch points is a finite subset of $\mathbb{C}$. It is the parity of the winding number $\eta(\gamma, -k)$ that is crucial for the determination of a branch an analytic continuation along γ ends up.

After we have transferred a truth assignment into a closed path γ, we look, whether the analytic continuations $X_1^\gamma, \ldots, X_m^\gamma$ of the germs $X_1, \ldots, X_m$ along γ vanish at $z = 0$ or not. For this reason, let $X_j = x_{j,r} \cdot x_{j,s} \cdot x_{j,t} \in \mathbb{O}_0$ be defined as in (3.6). According to the proof of Corollary 3.4, we can conclude that $X_j^\gamma[0] = 0$, if one of the analytic continuations of the factors of X_j vanishes at $z = 0$. Lemma 3.13 has shown that this is the case, if and only if

$$x_{j,k}^\gamma \in \left\{ x_{j,k}^{+,-}, x_{j,k}^{-,+} \right\}$$

holds for some $k \in \{r, s, t\}$. By taking this into account, we can prove the lemma below.

Lemma 3.14. *Let $j \in \{1, \ldots, m\}$ and $k \in \{r, s, t\}$. Then it holds that $x_{j,k}^\gamma[0] = 0$, if and only if one of the following two conditions hold:*

(i) $x_{j,k} = x_{j,k}^{+,+}$ *and* $\eta(\gamma, -k)$ *is odd or*

(ii) $x_{j,k} = x_{j,k}^{-,+}$ *and* $\eta(\gamma, -k)$ *is even.*

Proof. Let $x_{j,k} = x_{j,k}^{+,+} \in \mathbb{O}_0$ and denote the analytic continuation of $x_{j,k}$ along a closed path γ starting and ending at $z = 0$ by $x_{j,k}^\gamma$. Since the branch point $-k$ is simple, it follows that $x_{j,k}^\gamma = x_{j,k}^{+,-} \in \mathbb{O}_0$, if and only if $\eta(\gamma, -k)$ is odd. Then Lemma 3.13 shows that $x_{j,k}^\gamma[0] = 0$. For point (ii), we can apply an analogous argumentation. ☐

And Back to 3SAT Formulas

What we still have to do is to retransfer the findings of the last subsections into terms of boolean variables and 3SAT formulas. Retranslating Lemma 3.14 in terms of literals yields immediately

Lemma 3.15. *The analytic continuation of $x_{j,k}$ along γ vanishes at $z = 0$, in symbols $x_{j,k}^\gamma[0] = 0$, if and only if χ makes the corresponding literal $l_{j,k}$* TRUE*.*

A truth assignment χ makes a clause C_j of a 3SAT formula TRUE, when at least one of its literals is made TRUE by χ. In addition, χ makes an entire formula C TRUE, if all clause are satisfied.

Theorem 3.12. *The truth assignment χ makes C_j* TRUE*, if and only if $X_j^\gamma[0] = 0$. Moreover, χ makes C* TRUE*, if and only if*

$$X_C^\gamma[0] = \left(X_1^\gamma[0], \ldots, X_m^\gamma[0] \right) = (0, \ldots, 0) \in \mathbb{C}^m.$$

Proof. The first claim follows easily from Lemma 3.15 and the second claim is an immediate consequence of the first part of this theorem. ☐

A First Lower Bound

So far, we have shown that χ makes C TRUE, if and only if the constructed analytic continuations result in $X_C^\gamma[0] = (0, \ldots, 0) \in \mathbb{C}^m$. Now, we are going to reduce 3SAT to the reachability problem, where we only specify parts of a target instance.

Reachability for Partly Defined Target Instances. Let $\mathscr{P}$ be an SLP over $O \in \aleph$ and $Z \in \mathbb{C}^{p+q}$ a regular instance of $\mathscr{P}$. Moreover, let $S \in \mathbb{C}^p$, $I \subseteq \{1, \ldots, q\}$ and $w_j \in \mathbb{C}$ for $j \in I$. Then the problem of deciding, whether there exist a movement μ and an associated continuous evaluation of $\mathscr{P}$ with corresponding functions $c_1, \ldots, c_q$ so that

$$\begin{aligned} Z &= \Big(\mu_{1-p}(0), \ldots, \mu_0(0), c_1(0), \ldots, c_q(0)\Big), \\ S &= \Big(\mu_{1-p}(1), \ldots, \mu_0(1)\Big), \\ c_j(1) &= w_j \quad \text{for } j \in I \end{aligned}$$

hold, is called the *Reachability Problem for Partly Defined Target Instance over O* or short Partly-Reach over O.

Remark 3.23. The first condition ensures that the continuous evaluation matches a given starting instance Z, the vector S corresponds to predefined terminal points of the movement μ and the last condition guarantees that predefined values w_j ($j \in I$) are reached by a terminal instance.

In oder to complete the reduction, we have to ensure that X_C can be encoded by an SLP over *quad* polynomial in the coding length of C, this is, polynomial in mn.

Lemma 3.16. *Let C and X_C be defined as previously. Then there is an SLP $\mathscr{P}$ over quad that encodes X_C and the length $q+1$ of $\mathscr{P}$ is polynomial in the coding length of C.*

Proof. We have to show that the length of $\mathscr{P}$ is polynomial in mn. It is easy to observe that one can encode each factor $x_{j,k} \in \mathbb{O}_0$ by an SLP of length 9. Since each $X_j \in \mathbb{O}_0$ is a product of three of such factors, X_j can be realized by an SLP of length 29. Finally, X_C can be encoded by an SLP of length $29 \cdot m$. □

Let C be a 3SAT formula and $X_C \in \mathbb{O}_0$ the corresponding germ with base $0 \in \mathbb{C}$. According to Lemma 3.16, we can construct an SLP $\mathscr{P}$ that encodes X_C in an appropriate way. In Lemma 3.13, we have proved that $X_C[0] \in \{0,1\}^m$, so we can determine easily an associated regular starting instance Z of $\mathscr{P}$. For it, we use the principal branches of the involved roots, since our definitions refer to them. Now, we are asking, whether there exists a closed path γ starting and ending at $z = 0$ so that the analytic continuations of $X_1, \ldots, X_m$ along γ yields $X_C^\gamma[0] = (0, \ldots, 0) \in \mathbb{C}^m$.

If there exists such a path γ so that $X_C^\gamma[0] = (0, \ldots, 0) \in \mathbb{C}^m$, the previous subsections have shown that there is also a truth assignment χ that makes C TRUE. For each $k \in \{1, \ldots, n\}$, we set

$$b_k = \begin{cases} \text{TRUE}, & \text{if } \eta(\gamma, -k) \text{ even} \\ \text{FALSE}, & \text{if } \eta(\gamma, -k) \text{ odd} \end{cases} .$$

Applying Theorem 3.12 yields that $\chi = (b_1, \ldots, b_n) \in \{\text{TRUE, FALSE}\}^n$ is a truth assignment, which makes C TRUE. On the other hand, if $\chi = (b_1, \ldots, b_n) \in \{\text{TRUE, FALSE}\}^n$ is a truth assignment that makes C TRUE, we can find a closed path γ starting and ending at $z = 0$ with

$$\eta(\gamma, -k) = \begin{cases} 0, & \text{if } b_k = \text{TRUE} \\ 1, & \text{if } b_k = \text{FALSE} \end{cases}$$

so that the analytic continuations of $X_1, \ldots, X_m$ along γ exist. Again, Theorem 3.12 provides that $X_C^\gamma[0] = (0, \ldots, 0) \in \mathbb{C}^m$. This finally shows

Theorem 3.13. *There is an analytic continuation X_C^γ of X_C along a closed path γ starting and ending at $z = 0$ so that $X_C^\gamma[0] = (0, \dots, 0) \in \mathbb{C}^m$, if and only if C is satisfiable.*

Assuming that we have got an efficient algorithm that solves the decision problem Partly-Reach for SLPs over *quad* implies

Theorem 3.14. Partly-Reach *over quad is NP-hard.*

Remark 3.24. At this point, we shall go a little bit more into details, on why we only prescribed parts of the terminal instance. So what do we actually specify in advance in the terminal instance? We prescribed the value of the free element, this is, $z = 0$, and the outcome of the continued germs, this is, $X_C^\gamma[0] = 0$, but we do not prescribe the remaining intermediates. If we also want to prescribe the remaining intermediates, we must know explicitly a truth assignment that makes the corresponding 3SAT formula TRUE, since the branches of the germs $x_{j,k}^\gamma$ indicate exactly this assignment.

A Further Lower Bound

The equivalence in Theorem 3.12 enables us to show that the problem Partly-Reach over *quad* is also coNP-hard in this subsection. For this purpose, we reduce the coNP-complete problem co3SAT to it proving this result. The required lemmata have already been proved previously.

In order to solve co3SAT, one has to decide, whether a 3SAT formula C is unsatisfiable meaning all possible truth assignments make C FALSE. From Theorem 3.12, it follows that a formula C is satisfiable, if and only if there is a closed path γ starting at $z = 0$ so that the analytic continuations of $X_1, \dots, X_m$ yield $X_C^\gamma[0] = (0, \dots, 0) \in \mathbb{C}^m$. In other words, C is unsatisfiable, if and only if there do not exist analytic continuations of $X_1, \dots, X_m$ yielding $X_C^\gamma[0] = (0, \dots, 0) \in \mathbb{C}^m$. As a result, deciding the problem Partly-Reach over *quad* for a terminal instance that encodes $X_C^\gamma[0] = (0, \dots, 0) \in \mathbb{C}^m$ also solves co3SAT.

Since we have used the same germs as for 3SAT, it is clear that they can be encoded by an appropriate SLP. This proves

Theorem 3.15. Partly-Reach *over quad is coNP-hard.*

Remark 3.25. The Theorems 3.14 and 3.15 suggest that Partly-Reach over *quad* is neither in NP nor in coNP. Otherwise, it would follow that NP = coNP, which seems to be unlikely (see [40]).

Related Problems

We can derive a related complexity result easily by reinterpreting continuous evaluations in terms of analytic continuations and another one is received by reformulating analytic continuation in terms of complex integration. As a first step, we change our point of view in order to get a function which maps to $\mathbb{C}$. Let $X_1, \dots, X_m \in \mathbb{O}_0$ be the germs that have been constructed to encode a 3SAT formula. Then we introduce a germ Y_C, which is induced by the sum of $X_1, \dots, X_m$, i.e.

$$Y_C := \sum_{k=1}^{m} X_k \in \mathbb{O}_0.$$

We have proved in Lemma 3.13 that any analytic continuation Y_C^γ of Y_C satisfies $Y_C^\gamma[0] \in \{0, \dots, m\}$. Furthermore, a direct consequence of Theorem 3.12 is that there exists an analytic continuation Y_C^γ so that $Y_C^\gamma[0] = 0 \in \mathbb{C}$, if and only if the corresponding 3SAT formula is satisfiable. This motivates the following decision problem concerning analytic continuations.

Reachability Problem for Analytic Continuations. Given an SLP over $O \in \aleph$ that encodes a corresponding germ $g \in \mathbb{O}_{z_0}$ with $g[z_0] = w_0 \in \mathbb{C}$ and let $z_1, w_1 \in \mathbb{C}$ be two further complex numbers. Then the problem of deciding, whether there exist a path γ connecting z_0 and z_1 and an analytic continuation g^γ of g so that $g^\gamma[z_1] = w_1$ is called *Reachability Problem for Analytic Continuations over O*, short ANALYTIC-REACH over O.

It has been shown that continuous evaluations are associated with analytic continuations in Section 3.3. So ANALYTIC-REACH can be interpreted as a special case of the problem PARTLY-REACH. This implies that the Theorems 3.14 and 3.15 transfer directly to the problem ANALYTIC-REACH.

Corollary 3.5. *The problem* ANALYTIC-REACH *over quad is both NP-hard and coNP-hard.*

We can derive another result by simply reformulating analytic continuation in terms of integration. Let the analytic continuation of g along a certain path γ from z_0 to z_1 be denoted by g^γ, then the complex extension of the fundamental theorem of calculus provides

$$\int_\gamma g'(\zeta) d\zeta = g^\gamma(z_1) - g(z_0), \tag{3.8}$$

where g' denotes the derivative of g. Connecting the construction that encodes 3SAT formulas and the corresponding truth assignments with identity (3.8) leads to

$$\int_\gamma Y_C'(\zeta)\, d\zeta = Y_C^\gamma(0) - Y_C(0),$$

where γ is a closed path starting and ending at $z = 0$. By applying the same argumentation as previously, we observe

Lemma 3.17. *There is a path γ with*

$$\int_\gamma Y_C'(\zeta)\, d\zeta = -Y_C(0),$$

if and only if C is satisfiable.

Similarly, the just stated Lemma gives also rise to another decision problem concerning complex integration.

Reachability Problem for Contour Integrals. Given three complex numbers $z_0, z_1, w \in \mathbb{C}$ and an SLP over $O \in \aleph$ that encodes a corresponding germ $g \in \mathbb{O}_{z_0}$. Then the problem of deciding, whether there exists a path γ connecting z_0 and z_1 so that

$$\int_\gamma g(\zeta)\, d\zeta = w$$

is called *Reachability Problem for Contour Integrals over O*, short CONTOUR-REACH over O.

Remark 3.26. Another complexity result concerning contour integrals is presented in [42]. There it is shown that deciding, whether a contour integral of rational functions is not zero, is NP-hard. In [37], there can be found an undecidability result concerning integrals of functions of one real variable that can be constructed using the constant 1, addtion, subtraction, multiplication, division and the sine. Let f be such a function, then it is shown that deciding, whether definite integrals of the form $\int_{-\infty}^{\infty} f(t)\,dt$ converges, in undecidable.

There is only one step left that we have to prove in order to obtain a lower complexity bound concerning CONTOUR-REACH over *quad*. We must ensure that Y'_C can also be coded by an SLP over *quad* polynomial in the coding length of a corresponding 3SAT formula. Since

$$Y_C = \sum_{j=1}^{m} X_j \in \mathbb{O}_0$$

it follows that

$$Y'_C = \sum_{j=1}^{m} X'_j \in \mathbb{O}_0.$$

As a result, we have to study the germs X'_j of the sum. Since $X_j = x_{j,r} \cdot x_{j,s} \cdot x_{j,t} \in \mathbb{O}_0$ is induced by the product

$$\frac{(\sqrt{r} \oplus_r \sqrt{r+z})}{2\sqrt{r}} \cdot \frac{(\sqrt{s} \oplus_s \sqrt{s+z})}{2\sqrt{s}} \cdot \frac{(\sqrt{t} \oplus_t \sqrt{t+z})}{2\sqrt{t}},$$

where $\oplus_k \in \{+,-\}$ depending on $l_k^j \in \{x_k, \neg x_x\}$ for $k \in \{r,s,t\}$, a short computation shows that $X'_j \in \mathbb{O}_0$ is, in turn, induced by

$$\begin{array}{l} \oplus_t \frac{(\sqrt{r}\oplus_r\sqrt{r+z})\cdot(\sqrt{s}\oplus_s\sqrt{s+z})}{16\sqrt{rst}\cdot\sqrt{t+z}} \\ \oplus_s \frac{(\sqrt{r}\oplus_r\sqrt{r+z})\cdot(\sqrt{t}\oplus_t\sqrt{t+z})}{16\sqrt{rst}\cdot\sqrt{s+z}} \\ \oplus_r \frac{(\sqrt{s}\oplus_s\sqrt{s+z})\cdot(\sqrt{t}\oplus_t\sqrt{t+z})}{16\sqrt{rst}\cdot\sqrt{r+z}}. \end{array}$$

As previously, it is not hard to observe that Y'_C can be coded by an SLP polynomial in the length of the 3SAT formula as well. Consequently, the argumentation of the previous subsections yields

Corollary 3.6. *The decision problem* CONTOUR-REACH *over quad is NP-hard and coNP-hard.*

Remark 3.27. Actually, we can even strengthen the results in this subsection a bit by not using arbitrary constants in the set of admitted SLP-operations, but 0 and 1. This works, since all required constants can be generated by a sub-SLP of appropriate polynomial length that only uses the constants 0 and 1.

3.4.2 Results Concerning Undecidability

After we have proved several NP-hard and coNP-hard results, we focus on undecidability in the following. For this reason, we introduce an undecidable problem, which is used in the reductions. It is HILBERT's famous 10-th problem.

Hilbert's 10-th Problem. Let $p \in \mathbb{Z}[X_1, \ldots, X_n]$ be a polynomial with integer coefficients in the variables $X_1, \ldots, X_n$. Then the problem of deciding, whether there is an integer vector $y \in \mathbb{Z}^n$ with $p(y) = 0$ is called HILBERT's 10-th problem, short H10. The complexity of this problem is stated in the theorem below. We refer to [37] for a corresponding proof.

Theorem 3.16. *The problem H10 is undecidable.*

Remark 3.28. Actually, it is sufficient to set $n = 11$, since H10 is even undecidable, if polynomials in 11 variables are only considered (see [45]).

From Polynomials to Germs

Let $p \in \mathbb{Z}[X_1, \ldots, X_n]$ be a polynomial with integer coefficients in the unknowns $X_1, \ldots, X_n$. Similar to the approach at 3SAT formulas, we are going to construct germs that represent the variables $X_1, \ldots, X_n$. These germs, or more precisely, their analytic continuations, will have the feature that they will evaluate to an integer at $z = 0$. In fact, the specific integer will be dependent on a certain winding number of a path taken. For this reason, we can reproduce every vector $x \in \mathbb{Z}^n$ of integers by analytic continuations that are evaluated at $z = 0$. By doing so, we can imitate the evaluation of p for an arbitrary vector of integers by evaluating analytic continuations.

For each $k \in \{1, \ldots, n\}$, we define a germ $x_k \in \mathbb{O}_0$ by the use of the principal branch of the complex logarithm. The germ x_k is induced by

$$\frac{1}{2\pi i} \cdot \log\left(\frac{z}{k} + 1\right).$$

It is clear that the germs $x_1, \ldots, x_n \in \mathbb{O}_0$ are well-defined, since they are all holomorphic on an appropriate open neighborhood of $0 \in \mathbb{C}$. The next step is to encode the polynomial p. This is done by

$$x_p := p\left(x_1, \ldots, x_n\right) \in \mathbb{O}_0.$$

The germ $x_p \in \mathbb{O}_0$ is also well-defined, since x_p is holomorphic on an appropriate open neighborhood of $0 \in \mathbb{C}$. The set $\{-n, \ldots, -1\} \subset \mathbb{C}$ is the set of branch points of x_p and an element $-k \in \{-n, \ldots, -1\}$ is the only branch point of x_k. In addition, these branch points are all simple.

Analytic Continuations as Counter

Let $-k \in \{-n, \ldots, -1\}$ and $r \in \mathbb{Z}$. Then there are closed paths γ starting and ending at $z = 0$ with $\eta(\gamma, -k) = r$ and $\gamma([0,1]) \cap \{-n, \ldots, -1\} = \emptyset$, since $\{-n, \ldots, -1\}$ is a finite subset of $\mathbb{C}$. The winding number of $\frac{\gamma(t)}{k} + 1$ w.r.t. 0 is crucial for the branch the analytic continuation x_k^γ of x_k along γ ends up, finally. An easy consequence of the well-know argument principle gives us

$$\eta\Big(\frac{\gamma(t)}{k} + 1, 0\Big) = \sum_a \eta(\gamma, a),$$

where a are the zeros of $\frac{z}{k} + 1$ counted with their multiplicities. As a result, it follows that the analytic continuation x_k^γ of x_k along γ lies on the r-th branch of the corresponding logarithm, since $-k$ is a simple branch point of $x_k \in \mathbb{O}_0$. This implies $x_k[0] = r$. For this reason, we can use the germs $x_1, \ldots, x_n \in \mathbb{O}_0$ and their analytic continuations like counters that run through the integers. By definition, we observe directly

Lemma 3.18. *Given a closed path γ with $\gamma([0,1]) \cap \{-n,\dots,-1\} = \emptyset$ that starts and ends at $z = 0$. Then it holds $x_k^\gamma[0] = r \in \mathbb{Z}$, if and only if $\eta(\gamma, -k) = r$.*

So far, we have constructed counters for the variables $X_1, \dots, X_n$ that evaluate to an arbitrary integer at $z = 0$. This is all we need to prove the next crucial theorem, in which we connect analytic continuations of $x_1, \dots, x_n$ with the polynomial $p \in \mathbb{Z}[X_1, \dots, X_n]$.

Theorem 3.17. *There is a closed path γ starting and ending at $z = 0$ with $\gamma([0,1]) \cap \{-n,\dots,-1\} = \emptyset$ so that $x_p^\gamma[0] = 0$, if and only if there is a vector $y \in \mathbb{Z}^n$ so that $p(y) = 0$.*

Proof. At first, let γ be a closed path starting and ending at $z = 0$ so that $x_p^\gamma[0] = 0$. By definition, this is equivalent to $p(x_1^\gamma[0], \dots, x_n^\gamma[0]) = 0$. Since $x_k^\gamma[0] \in \mathbb{Z}$ for all $k \in \{1, \dots, n\}$ (see lemma 3.18), there is a $y \in \mathbb{Z}^n$ such that $p(y) = 0$.
Now we prove the converse. There is a $y = (y_1, \dots, y_n) \in \mathbb{Z}^n$ such that $p(y) = 0$. Then there exists a closed path γ starting and ending at $z = 0$ with $\eta(\gamma, -k) = y_k$ and $\gamma([0,1]) \cap \{-n,\dots,-1\} = \emptyset$. From Lemma 3.18, we get that the analytic continuation of x_k^γ of x_k along γ evaluates to the integer y_k at $z = 0$, i.e. $x_k^\gamma[0] = y_k$. As a result, it holds that $x_p^\gamma[0] = p(0) = 0$. □

Final Reduction and Related Problems

There is actually one thing left, namely we have to ensure that x_p can be encoded by an SLP with finite length. Therefore, we prove

Lemma 3.19. *There is an SLP of finite length that encodes $x_p \in \mathbb{O}_0$.*

Proof. We have to encode the germs $x_1, \dots, x_n \in \mathbb{O}_0$. Each of them can be encoded by an SLP of length bounded by 8 and thus all of them can be encoded by an SLP, which length is bounded by $8n$. The polynomial p has a total degree, say $d \in \mathbb{N}$. Due to this fact each monomial can be encoded by an SLP of length bounded by d. In addition, the number of monomials is bounded by $(d+1)^n$. So the length of a corresponding SLP is bounded by $8nd(d+1)^n$, which is finite. □

Now, denote the constant term of p by $c \in \mathbb{Z}$, i.e. $p(0) = x_p[0] = c$. Again, we ask, whether there exists a closed path γ starting and ending at $z = 0$ so that the associated analytic continuation of x_p along γ yields $x_p^\gamma[0] = 0$. From Theorem 3.17, we know that this is the case, if and only if p has an integral zero. In other words, we decide H10, if we can decide Partly-Reach over *tran*. So we have proved

Theorem 3.18. *The problem* Partly-Reach *over tran is undecidable.*

Just as previously, we can derive two direct corollaries concerning analytic continuation and complex integration. The first result is stated in the corollary below.

Corollary 3.7. Analytic-Reach *over tran is undecidable.*

Moreover, we apply again the fundamental theorem of calculus to connect analytic continuation and integration in the same way as before. This leads to

Corollary 3.8. Contour-Reach *over tran is undecidable.*

Proof. The only thing we have to prove is that the derivative of x_p w.r.t. z can be encoded by an SLP of finite length. It holds

$$x'_k(z) := \frac{d}{dz} x_k(z) \quad = \quad \frac{1}{2\pi i \cdot (z+k)}$$

and

$$\frac{d}{dz}\, x_p(z) \quad = \quad p\big(x_1(z), \ldots, x_n(z)\big) = \sum_{k=1}^{n} x'_k(z) \cdot \left. \frac{\partial\, p(X_1, \ldots, X_n)}{\partial X_k} \right|_{X_1 = x_1(z), \ldots, X_n = x_n(z)} .$$

The partial derivative $\frac{\partial\, p(X_1,\ldots,X_n)}{\partial X_k}$ is again a polynomial from $\mathbb{Z}[X_1, \ldots, X_n]$, which can be encoded by a sub-SLP of finite length. Moreover, it can be seen easily that there are sub-SLPs of finite length that encode $x'_1, \ldots, x'_n$. That $x_1, \ldots, x_n$ can be encoded by an appropriate SLP has already been proved in Lemma 3.19. Lastly, the claim follows from the fact that there are only n multiplications and n additions left in order to encode the entire derivative. □

3.5 Restricting Movements

Another way to derive lower complexity bounds is to restrict the movement of a free element. We are going to deal with two different types of such restrictions in the following. Firstly, we are only going to admit paths that are JORDAN curves. A JORDAN curve is a plane curve, which is topologically equivalent to the unit circle. In other words, it is closed and simple. The other restriction that we are going to use is to bound the length of a movement of a free element.

3.5.1 Only Jordan Curves

The crucial thing that we gain, when we only admit JORDAN curves, is that we can control winding numbers. This is due to the fact that the winding number $\eta(\gamma, z)$ w.r.t. a complex number $z \in \mathbb{C}$ of a JORDAN curve γ is either 0 or 1.[21] This enables us to construct germs that can be used to count the number of branch points that are circled around by a JORDAN curve. We are going to use the mentioned counting mechanism to reduce an NP-complete problem in coding theory to SLP-REACH over *quad*.

Subspace Weights Problem. Given a binary matrix $A \in \{0,1\}^{m \times n}$ and a nonnegative integer $\kappa \in \mathbb{N} \cup \{0\}$. Furthermore, we call the number of 1's in a binary vector $x \in \{0,1\}^m$ the HAMMING weight of x. Then the problem of deciding, whether there exists a binary vector $x \in \{0,1\}^n$ of HAMMING weight κ so that $A \cdot x$ modulo 2 equals the all-zero vector is called *Subspace Weights Problem*, short SUB-WEIGHTS. Its intrinsic complexity was proved in [3].

Theorem 3.19. *The decision problem* SUB-WEIGHTS *is NP-complete.*

[21] Actually, that is only half the story. Depending on the orientation of the JORDAN curve, there is also the case that a winding number is either 0 or -1. However, this is not a restriction, since we can simply replace 1 and -1 and get an analogous argumentation.

From Coding Theory to Germs

As a first step, we transfer the problem SUB-WEIGHTS to a function theoretic setting consisting of operations from the set *quad*. We begin with the binary matrix $A \in \{0,1\}^{m\times n}$. Thereby, we denote the entries of A by $a_{i,j} \in \{0,1\}$. For each row $(a_{k,1},\dots,a_{k,n}) \in \{0,1\}^n$ of A, we introduce a polynomial p_k, which is defined as follows:

$$\prod_{j=1}^{n} a_{k,j}(X-j) \in \mathbb{Z}[X] \qquad \big(k \in \{1,\dots,m\}\big). \tag{3.9}$$

Basically, the common zeros of the polynomials $p_1,\dots,p_m \in \mathbb{Z}[X]$ capture the combinatorial structure that is contained in the binary matrix A. Based on this, we use square roots to imitate the modular arithmetic. For this reason, we construct germs that are induced by the square root of each polynomial $p_k \in \mathbb{Z}[X]$. More precisely, we get a germ $g_k \in \mathbb{O}_0$ for each $k \in \{1,\dots,m\}$ that satisfies $g_k[z] = \sqrt{p_k(z)}$ on an appropriate open neighborhood of $0 \in \mathbb{C}$. Again, this definition refers to principal branches of the square roots.

Each germ $g_k \in \mathbb{O}_0$ induces a RIEMANN surface. This surface has two sheets, since only one square root is involved. The corresponding branch points are the zeros of the polynomial $p_k \in \mathbb{Z}[X]$. They are all contained in the finite set $\{1,\dots,n\} \subset \mathbb{C}$ and they are all simple. Starting from this setting, the germs $g_1,\dots,g_m \in \mathbb{O}_0$ can be used to imitate the matrix vector product $A \cdot x$ modulo 2 in the problem SUB-WEIGHTS as follows:

Let $g_k \in \mathbb{O}_0$ be one of the previously introduced germs and γ be a closed path starting and ending at $z = 0$ with $\gamma([0,1]) \cap \{1,\dots,n\} = \emptyset$. Then it is clear that the analytic continuation $g_k^\gamma \in \mathbb{O}_0$ of g_k along γ exists and the branch g_k^γ, on which g_k^γ lies, depends on the parity of the winding number $\eta(p \circ \gamma, 0)$ of the composition $p \circ \gamma$ w.r.t. 0. Using again the well-known argument principle gives us

$$\eta(p \circ \gamma, 0) = \sum_r \eta(\gamma, r), \tag{3.10}$$

where r are the zeros of p counted with their multiplicities. As a result, we obtain immediately the lemma below.

Lemma 3.20. *Let $k \in \{1,\dots,m\}$, $p_k \in \mathbb{Z}[X]$ defined as in (3.9) and γ be a* JORDAN *curve starting and ending at $z = 0$. Then $\eta(p \circ \gamma, 0)$ is odd, if and only if the cardinality of*

$$\big\{z \in \mathbb{C} \;\big|\; p_k(z) = 0 \wedge \eta(\gamma, z) = 1\big\}$$

is odd.

In a way, we can think of the zeros of a polynomial $p_k \in \mathbb{Z}[X]$ as switches for a light. Thereby, the switches are operated by the analytic continuation $g_k^\gamma \in \mathbb{O}_0$. The branch g_k^γ lies on tells us, whether the light is on or off. For this reason, we can prove the theorem below, in which we show that the product $A \cdot x$ modulo 2 form SUB-WEIGHTS can be identified with a constellation of branches the analytic continuations $g_1^\gamma,\dots,g_m^\gamma \in \mathbb{O}_0$ lie on.

Theorem 3.20. *Let $A \in \{0,1\}^{m\times n}$, $g_1,\dots,g_m \in \mathbb{O}_0$ be defined as previously. Then the following two statements are equivalent:*

(i) There exists a vector $x \in \{0,1\}^n$ so that $A \cdot x$ modulo 2 is the all-zero vector.

(ii) There is a JORDAN *curve γ starting and ending at $z = 0$ with $\gamma([0,1]) \cap \{1,\dots,n\}$ so that each of the germs $g_1^\gamma,\dots,g_m^\gamma$ lie on the principal branches of their corresponding square roots.*

Proof. First of all, we do some preparatory work. Let $(a_{k,1}, \dots, a_{k,n}) \in \{0,1\}^n$ be a row of A. Then the scalar product of this row and a vector $x = (x_1, \dots, x_n) \in \{0,1\}^n$ is given by the sum

$$\sum_{j=1}^{n} a_{k,j} \cdot x_j. \tag{3.11}$$

An addend $a_{k,j} \cdot x_j$ of this sum is 1, if and only if $a_{k,j} = 1$ and $x_j = 1$ hold. Now, we transfer this observation into terms of the germ $g_k \in \mathbb{O}_0$ and its analytic continuation. The condition $a_{k,j} = 1$ is encoded in the zeros of the polynomials, namely the polynomial $p_k \in \mathbb{Z}[X]$ has the simple zero j, if and only if the entry $a_{k,j}$ of A equals 1. Furthermore, $x_j = 1$ is reflected by an analytic continuation of the associated germ $g_k \in \mathbb{O}_0$ along a closed path γ. The continuation induces a change of the branch of the underlying square root in $g_k \in \mathbb{O}_0$, if and only if $\eta(\gamma, j)$ is odd. As a result, we have rebuilt the addend in terms of g_k and its analytic continuation. The next step is to consider the entire sum (3.11), but this works straight forward. We simply consider an analytic continuation of $g_k \in \mathbb{O}_0$ w.r.t. all zeros of p_k. Lemma 3.20 guarantees us that the modulo 2 arithmetic is reflected by the terminal branch $g_k^\gamma \in \mathbb{O}_0$ lies on. All in all, we have observed that the product $A \cdot x$ modulo 2 and the constellations of the terminal branches $g_1, \dots, g_m \in \mathbb{O}_0$ lie on are equivalent.

With this spadework in mind, it is not hard to prove the equivalence of the two statements in the theorem. We prove that the first statement implies the second one. Let $x \in \{0,1\}^n$ so that $A \cdot x$ modulo 2 is the all-zero vector. From x, we construct an appropriate Jordan curve γ that starts at $z = 0$. If the component x_k of x is 1, the winding number $\eta(\gamma, k)$ of γ w.r.t. k shall also be 1 and otherwise 0. It is clear that such a path exists, since the set $\{1, \dots, n\}$ is a finite subset of $\mathbb{C}$. The preliminary considerations complete the claim.

Now, we prove the reverse direction. Therefore, let γ be a Jordan curve with winding numbers $\eta(\gamma, 1), \dots, \eta(\gamma, n)$ and $\gamma([0,1]) \cap \{1, \dots, n\}$. In this case, we define a vector $x \in \{0,1\}^n$. The component x_k of x shall equal the winding number $\eta(\gamma, k)$. Again, the preliminary considerations of this proof complete the claim. □

A Counter

We require another ingredient before we can prove a lower complexity bound. As mentioned in the headline of this subsection, we are going to build a counter, which is used to count the number of points from the set $\{1, \dots, n\} \subset \mathbb{C}$ that are circled around by a Jordan curve. The idea is again to use the zeros of a polynomial in combination with square roots. In order to count more than just the parity as done in the last subsection, we use nested square roots to count the numbers between 0 and $2^n - 1$. We will observe that the germ $BC \in \mathbb{O}_0$, which is induced by

$$\sqrt[2^n]{\prod_{j=1}^{n} (z - j)} \tag{3.12}$$

has the desired feature. It is clear that $BC \in \mathbb{O}_0$ is well-defined and that it can be continued analytically along all paths γ with $\gamma([0,1]) \cap \{1, \dots, n\} = \emptyset$. In (3.12), it is again referred to the principal branch of the 2^n-th root. By doing so, $BC[0] = \sqrt[2^n]{n!}$ is a positive real number. Moreover, the associated Riemann surface has 2^n different sheets, since a 2^n-th root occurs in the definition. Now, let γ be Jorddan curve starting and ending at $z = 0$

with $\gamma([0,1]) \cap \{1,\ldots,n\} = \emptyset$. Moreover, we define ℓ to be the sum

$$\sum_{j=1}^{n} \eta(\gamma, j).$$

As $\eta(\gamma, j) \in \{0,1\}$ for $j \in \{1,\ldots,n\}$, it follows $\ell < 2^n$. Applying identity (3.10) together with the fact that all branch points $1,\ldots,n$ of BC are simple imply that the analytic continuation $BC^\gamma \in \mathbb{O}_0$ of BC along γ ends up on the ℓ-th branch of the 2^n-th root, i.e.

$$BC^\gamma[0] = BC[0] \cdot e^{\frac{2\pi i}{2^n}\cdot \ell}. \tag{3.13}$$

Since $\ell < 2^n$, it is impossible to get $BC[0] = BC^\gamma[0]$ for $\ell \neq 0$. So we have built a well-defined counter for the numbers $0,\ldots,2^n-1$ by simply reading off the number of the branch ℓ. This can be done by checking equation (3.13) for a fixed given ℓ.

Until now, we have just developed the basic idea. The next step is to show that this counter can be encoded by an SLP over *quad*. The trivial part is the encoding of the product under the root in BC. The encoding of the root, however, is done by n nested square roots. Basically, it is clear that n nested roots are actually building a 2^n-th root. So we only have to translate equation (3.13) into terms of nested square roots. At first, we will have a look at two nested square roots. The function

$$f : U \to \mathbb{C}; \quad z \mapsto \sqrt{\sqrt{z}}$$

is holomorphic on a proper open neighborhood U of $1 \in \mathbb{C}$. Furthermore, we can continue f analytically along a closed circular path γ that starts at $z = 1$ and circles around $0 \in \mathbb{C}$. Since the Riemann surface of the square root covers the complex plane twice, it follows that the image of the inner square root concerning the path γ has half the angular velocity of the path γ itself. This implies, in turn, that γ must have a winding number of 2 w.r.t. $0 \in \mathbb{C}$ in order to change the branch of the outer square root. Due to this fact the constellation of branches of the square roots reflects the winding number of γ w.r.t. $0 \in \mathbb{C}$. Thereby, we have to read this constellation as a binary representation of the winding number. If we assume that the principal branch is always initially chosen by the definition of f and if we identify the principal branch with 0 and the other one with 1, we will get the following branch constellations depending on the winding number of γ:

$\eta(\gamma, 0)$	outer root	inner root
0	0	0
1	0	1
2	1	0
3	1	1

It is obvious that the two columns on the right side are the binary representation of the corresponding winding numbers seen in the left column. It is easy to observe that this concept generalizes straight forward to n nested square roots, which provides the desired counter.

From Germs back to Coding Theory

It is still left to verify that the germs $g_1,\ldots,g_m, BC \in \mathbb{O}_0$ can be encoded by an SLP over *quad* that is polynomial in the coding length of Sub-Weights.

Lemma 3.21. *The germs $g_1, \dots, g_m, BC \in \mathbb{O}_0$ can be encoded by an SLP over quad that length is polynomial in mn.*

Proof. The counter BC can be realized by 1 free operation, n constants, n subtractions, $n-1$ multiplications and n square roots. Totally, an encoding SLP has a length that is bounded by $4n$. We just need one additional square root to encode $g_k \in \mathbb{O}_0$. Thus we can encode all germs $g_1, \dots, g_m \in \mathbb{O}_0$ by m further operations. Consequently, all germs can be encoded by an SLP, which length is bounded by $m + 4n$. □

Finally, we reduce SUB-WEIGHTS to the weaker version of SLP-REACH, where only JORDAN curves are admitted for the movement of a sole free object. Let $\mathscr{P}$ be a proper SLP that encodes the germs $g_1, \dots, g_m, BC \in \mathbb{O}_0$. Furthermore, we have to introduce a starting instance Z and a terminal instance W. For this reason, denote the constant terms of the polynomials $p_1, \dots, p_m$ by $c_1, \dots, c_m \in \mathbb{N}$. Now, we can define the instances Z and W implicitly by fixing the free elements and the corresponding outputs of the germs $g_1, \dots, g_m, BC \in \mathbb{O}_0$, since the remaining operations are uniquely determined by these settings. The starting instance Z is determined by

$$g_k[0] = \sqrt{c_k} \quad \big(k \in \{1, \dots, m\}\big) \quad \text{and} \quad BC[0] = \sqrt[2^n]{n!},$$

where it is referred to the principal branches of the underlying roots. Then the terminal instance W is determined by

$$g_k^\gamma[0] = g_k[0] \quad \big(k \in \{1, \dots, m\}\big) \quad \text{and} \quad BC^\gamma[0] = BC[0] \cdot e^{\frac{2\pi i}{2^n} \cdot \kappa},$$

where $\kappa \in \mathbb{N} \cup \{0\}$ is the nonnegative integer from SUB-WEIGHTS. The index γ shall visualize that these conditions refer to the terminal instance W. We have already shown in Theorem 3.20 that the part in W belonging to the conditions

$$g_k^\gamma[0] = g_k[0] \quad \big(k \in \{1, \dots, m\}\big)$$

can be reached by analytic continuations of $g_1, \dots, g_m \in \mathbb{O}_0$, if and only if there is a vector x so that the product $A \cdot x$ modulo 2 is the zero vector. Besides, we have discussed in the subsection about the counter that the parts of the instances concerning the counter BC make sure that the vector x additionally has HAMMING weight κ. So we have proved

Theorem 3.21. *The problem* SLP-REACH *over quad, where* JORDAN *curves are only admitted for the movements of the free objects, is NP-hard.*

3.5.2 Bounding the Length of Movements

In this part of the chapter, we concentrate on SLP-REACH over *kubi* with one free object. In contrast to the last subsection, we require that the length of the admitted movement is bounded. Similarly, the bounded length of a movement enables us to control the way and thus the winding numbers of the movement, again. In order to prove a lower complexity bound on this version of the problem, we are going to reduce the following variant of 3SAT to it.

Exact 3SAT. Given a 3SAT formula C. Then the problem of deciding, whether there exists a truth assignment χ of C so that χ makes exactly one literal TRUE in each clause of C is called *Exact 3SAT*, short X-3SAT. The complexity of this problem was proved in [53].

Theorem 3.22. *The problem* X-3SAT *is NP-complete.*

Outline of the Reduction

Before we start with the reduction itself, we point out briefly the four steps that are used in it. They are as follows:

1. We make the same transition as in Section 3.4.1. There a literal $l_{j,k}$ of a 3SAT formula C has been transferred into a germ $x_{j,k}$ involving square roots. Basically, $l_{j,k}$ can either be b_k or $\neg b_k$. These two states are represented by the fact that $x_{j,k}$ either vanishes at $z = 0$ or not. By doing so, a clause becomes a product X_j of three such germs in this setup. As presented in Section 3.4.1, a truth assignment χ of C is encoded by a closed path γ that is used for continuing $x_{j,k}$ analytically. Thereby, $x_{j,k}$ and γ are chosen so that the analytic continuation of $x_{j,k}$ along γ vanishes at $z = 0$, if and only if χ makes the literal $l_{j,k}$ true.
2. We must check the number of literals in a clause that have been made true by a truth assignment, since we reduce X-3SAT. This is achieved by 3-rd roots of X_j. In doing so, we can use the three branches of a 3-rd root as a counter for counting the number of vanishing factors in the analytic continuation of X_j along γ.
3. The bounded length enables us to control the way of γ. The idea is that if the length of γ is bounded by $2s$ and γ should pass two points that have a distance s, it follows that the way of γ has to be the segment joining these two points.
4. We assemble the previous points to prove the desired lower complexity bound.

Once More from 3SAT Formulas and Truth Assignments to Germs and Analytic Continuations

The entire reduction is based on the setup, which has been developed in Section 3.4.1. For this reason, we recall the construction briefly. Let

$$C = C_1 \wedge \ldots \wedge C_m$$

be 3SAT formula built of literals from $\{b_1, \ldots, b_n, \neg b_1, \ldots, \neg b_n\}$. For each $j \in \{1, \ldots, m\}$, a clause C_j equals $l_{j,r} \vee l_{j,s} \vee l_{j,t}$ where $l_{j,k} \in \{b_k, \neg b_k\}$. Each literal $l_{j,k}$ of C_j has been transferred to a germ $x_{j,k} \in \mathbb{O}_0$ induced by

$$\frac{\sqrt{k} \oplus_k \sqrt{k+z}}{2\sqrt{k}},$$

where $\oplus_k = -$, if $l_{j,k} = b_k$, and $+$ otherwise. Furthermore, a clause C_j has become a product $X_j = x_{j,r} \cdot x_{j,s} \cdot x_{j,t} \in \mathbb{O}_0$ and the entire formula has been represented by the vector of germs

$$X_C = \bigl(X_1, \ldots, X_m\bigr).$$

A truth assignment has been transferred to a closed path. Let $\chi = (b_1, \ldots, b_n)$ be a truth assignment for B. Corresponding to χ, we have construct a closed path γ starting and ending at $z = 0$ that satisfies the following condition: The winding number $\eta(\gamma, -k)$ of γ w.r.t $-k$ should be even, if b_k is TRUE and odd, if b_k is FALSE. It has been clear that such a path and the associated analytic continuations of $X_1, \ldots, X_m$ exist, since the set

$\{-n, \ldots, -1\}$ is a finite subset of $\mathbb{C}$. According to Lemma 3.15 and Theorem 3.12, this construction has yielded that a truth assignment χ makes TRUE:

$$\begin{aligned} l_{j,k} &\iff x_{j,k}^{\gamma}[0] = 0, \\ C_j &\iff X_j^{\gamma}[0] = 0, \\ C &\iff X_C^{\gamma}[0] = (0, \ldots, 0) \in \mathbb{C}^m. \end{aligned}$$

Counting Multiplicities

We must check the number of literals in a clause C_j that have been made TRUE by a truth assignment χ. If γ is a path corresponding to χ, the number of true literals in C_j equals the number of vanishing factors in $X_j^{\gamma}[0]$. This number of vanishing factors, however, can be checked by a 3-rd root. The idea is in essence the same as for the counter in Section 3.5.1. There, a 2^n-th root counted the number of points that had been circled around by a JORDAN curve.

As a first step, we choose another base for $X_1, \ldots, X_m \in \mathbb{O}_0$. It can be seen easily that their basic definitions also induce germs at $\hat{z} := -n - 1$. Due to this fact we can also think of $X_1, \ldots, X_m \in \mathbb{O}_{\hat{z}}$. Next, we introduce new germs $Y_1, \ldots, Y_m \in \mathbb{O}_{\hat{z}}$ by

$$Y_j := \sqrt[3]{X_j} \qquad \big(j \in \{1, \ldots, m\}\big),$$

whereby these definitions refer to the principal branches of the underlying roots. It is clear that the modified germs $Y_1, \ldots, Y_m \in \mathbb{O}_{\hat{z}}$ are well-defined, since they are all holomorphic on an open neighborhood of $\hat{z} \in \mathbb{C}$.

A germ $Y_j \in \mathbb{O}_{\hat{z}}$ has the same zero as $X_j \in \mathbb{O}_{\hat{z}}$, i.e. $z = 0$, since it is just the 3-rd root of X_j. So $z = 0$ is also a branch point of $Y_j \in \mathbb{O}_{\hat{z}}$. The crucial fact now is that Y_j can be used as a counter that counts modulo 3. To observe this, let γ be a closed path with $\eta(\gamma, 0) = 2$ and let us identify the three branches of a 3-rd root with 0,1 and 2, where 0 denotes the principal branch and the others are matched accordingly. Then the argument principle (3.10) gives us the number of vanishing factors in X_j^{γ}. More precisely, the analytic continuation Y_j^{γ} lies on branch 0, if the number of vanishing factors equals 0 or 3; it lies on branch 1, if it is 1; and it lies on branch 2, if this number is 2. In general, we assume that $\eta(\gamma, 0)$ is even. Then Y_j^{γ} lies on branch

$$M \cdot \frac{\eta(\gamma, 0)}{2} \pmod 3,$$

where M is the number of of vanishing factors in X_j^{γ}. In summary, the branches of $Y_1^{\gamma}, \ldots, Y_m^{\gamma}$ count the number of vanishing factors in $X_1^{\gamma}, \ldots, X_m^{\gamma}$, if the winding number $\eta(\gamma, 0)$ is even and known.

Remark 3.29. We will see that it is not necessary to know exactly the winding number $\eta(\gamma, 0)$, but that it is even. This, however, is not a further restriction on the path, since we can use an additional germ to guarantee this.

Using the Bounded Length

Apart from the encoding of a truth assignment, we extend the definition of γ a bit, since we would like to use the constraint on its length. We bound the length of γ by $2(n + 3) + \varepsilon$ with $0 < \varepsilon \leq 1$ and claim that γ should circle around the two further points $-n - 2$ and 1 in $\mathbb{C}$. Furthermore, it is necessary to require that γ is also rectifiable. However, this is not

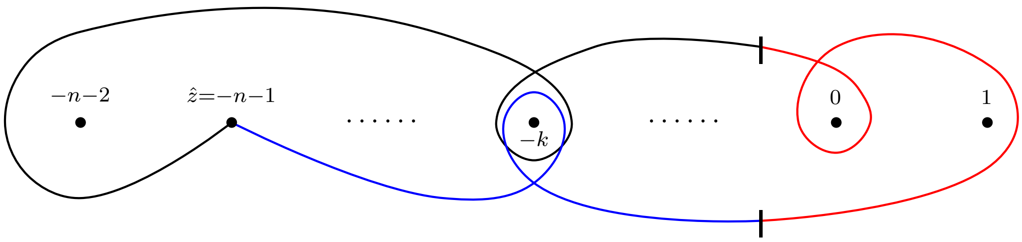

Figure 3.16: Illustration of $\gamma = \gamma_3 \circ \gamma_2 \circ \gamma_1$.

a restriction at all, since the set $\{-n-2, -n, \ldots, 1\}$ is a finite subset of $\mathbb{C}$ and γ can, for instance, be realized by a finite family of line segments. In doing so, we can observe three obvious facts concerning γ:

(i) In essence, the way of γ is the line segment joining $-n-2$ and 1.

(ii) The constraint on the length of γ admits arbitrary winding numbers $\eta(\gamma, k)$ for $k \in \{-n-2, -n, \ldots, 1\}$.

(iii) Since γ is closed and starts at $\hat{z}$, it passes $-n-2$ and 1 once and passes each point from the set $\{-n, \ldots, -1\}$ twice, namely once on the way to 1 and a second time on the way back to $\hat{z}$ or vice versa.

Taking these observations into account, we can assume w.l.o.g. that the way of γ is essentially as follows: It starts at $\hat{z}$, passes $-n-2$ first, then passes 1 and returns finally to $\hat{z}$. To simplify the notation throughout the rest of this chapter, we divide γ in three parts γ_1, γ_2 and γ_3 (see Fig. 3.16). γ_1 is the part of γ that starts from $\hat{z}$ and proceeds to 0. Thereby, we require that γ_1 should not circle around 0. γ_2 is the part of γ that starts at the terminal point of γ_1 and contains all circles around 0 and 1. γ_3 is the remaining part of γ.

Assembling the Parts

In this paragraph, we put things together and complete the reduction. Taking the partition $\gamma = \gamma_3 \circ \gamma_2 \circ \gamma_1$ into account, the first and most important goal is to prove the equivalence of the following two statements:

(i) There is a closed path γ with length smaller than $2(n+3)+\varepsilon$ $(0 < \varepsilon \leq 1)$ so that the analytic continuations $Y_1^\gamma, \ldots, Y_m^\gamma$ lie on branch 1 of the 3-rd roots.

(ii) C has a truth assignment that makes exactly one literal true in each clause.

As a first step, we focus on one clause. The crucial parts of γ are γ_1 and γ_2, since the continuation along γ_3 does not affect the branch of Y_j^γ at all. According to Lemma 3.15, it holds that there is a truth assignment that makes a literal $l_{j,k}$ TRUE in a clause C_j, if and only if there also exists a first part γ_1 of γ so that the corresponding continued factor $x_{j,k}^{\gamma_1}$ vanishes at $z = 0$ in $X_j^{\gamma_1}$. Furthermore, we have pointed out previously that the analytic continuation $Y_j^{\gamma_2 \circ \gamma_1}$ of Y_j along $\gamma_2 \circ \gamma_1$ lies on branch

$$M \cdot \frac{\eta(\gamma, 0)}{2} \pmod 3,$$

where M is the number of vanishing factors in $X_j^{\gamma_1}$. As a result, branch 1 can only be reached, if $M \in \{1,2\}$ and $\eta(\gamma,0)$ is appropriate.

Regarding now all clauses $C_1,\ldots,C_m$, we can assume that M is equal for all $X_1^{\gamma_1},\ldots,X_m^{\gamma_1}$, since

$$1 \cdot k \equiv 2 \cdot k \pmod 3 \iff k \equiv 0 \pmod 3.$$

In the case $M=1$, we are done, since this immediately implies the existence of a desired truth assignment. In the case $M=2$, there exists a truth assignment that makes exactly two literals TRUE in each clause. This, however, implies the existence of a truth assignment that makes exactly one literal TRUE in each clause as well. To obtain this assignment, one just has to switch TRUE to FALSE and vice versa. This shows the equivalence of statement (i) and (ii).

Remark 3.30. At this point, we should illuminate the connection between γ and truth assignments. According to the last observations, a truth assignment $\chi=(b_1,\ldots,b_n)$ is reflected by the first part γ_1, namely the number of circles around $-k$ of γ_1 should be even, if b_k is TRUE and odd, if b_k is FALSE. So it adjusts the number of vanishing factors in $X_1^{\gamma_1},\ldots,X_m^{\gamma_1}$ corresponding to χ. The part γ_2 is just used to operate the counters $Y_1,\ldots,Y_m$ to count the numbers of vanishing factors in $X_1^{\gamma_1},\ldots,X_m^{\gamma_1}$. The last part γ_3 has not played any role so far. Its only use is to reverse the adjustments of γ_1. This enables us to specify a simple terminal instance later on. For this reason, the number of circles around k of γ_3 should equal γ_1 for $k \in \{-n,\ldots,-1\}$.

An illustration of such a path γ is given in Fig. 3.16. There, the part γ_1 is drawn in black, γ_2 is the red part and the last part γ_3 is the blue one. Obviously, it holds $\eta(\gamma,-n-2)=\eta(\gamma,1)=1$. Moreover, it satisfies $\eta(\gamma,j)=1$ for $j\in\{-n,\ldots,-1\}\setminus\{-k\}$ and $\eta(\gamma,-k)=\eta(\gamma,0)=2$.

There are just few things left to do in order to prove the lower complexity bound. One the one hand, we must introduce three further functions that are required to control the path γ. They are induced by

$$Z_{-n-2}(z) := \sqrt{n+2+z}, \quad Z_0(z) := \sqrt{z} \quad \text{and} \quad Z_1(z) := \sqrt{-1+z}.$$

Obviously, they are all holomorphic on an open neighborhood of $\hat{z}\in\mathbb{C}$, so $Z_{-n-2}, Z_0, Z_1 \in \mathbb{O}_{\hat{z}}$. Z_{-n-2} and Z_1 are needed to check, whether γ circles around $-n-2$ and 1, while Z_0 is used to ensure that $\eta(\gamma,0)$ is even.

On the other hand, we must guarantee that all involved functions can be encoded by an SLP with a length polynomial in the coding length of a 3SAT formula C, this is, mn. However, this can be verified easily, since we have already proved that $X_1,\ldots,X_m$ can be encoded by an adequate SLP.

Last but not least, we specify regular instances Z and W of an encoding SLP that can only be reached by a continuous evaluation, if and only if C has a truth assignment that makes exactly one literal true in each clause. For this reason, we set $X_j[\hat{z}]=a_j\in\mathbb{R}$ and $Y_j[\hat{z}]=b_j\in\mathbb{R}$ for $j\in\{1,\ldots,m\}$. Then the starting instance Z is implicitly defined by

$$\begin{aligned}
X_j[\hat{z}] &= a_j\\
Y_j[\hat{z}] &= b_j\\
Z_0[\hat{z}] &= \sqrt{\hat{z}}\\
Z_{-n-2}[\hat{z}] &= \sqrt{n+2+\hat{z}}\\
Z_1[\hat{z}] &= \sqrt{1+\hat{z}}
\end{aligned}$$

and the terminal instance W by

$$\begin{aligned}
X_j^\gamma[\hat{z}] &= X_j[\hat{z}] \\
Y_j^\gamma[\hat{z}] &= Y_j[\hat{z}] \cdot e^{\frac{2\pi i}{3}} \\
Z_0^\gamma[\hat{z}] &= Z_0[\hat{z}] \\
Z_{-n-2}^\gamma[\hat{z}] &= -Z_{-n-2}[\hat{z}] \\
Z_1^\gamma[\hat{z}] &= -Z_1[\hat{z}]
\end{aligned}$$

Remark 3.31. Let us again go a little bit more into detail on the relevance of the conditions imposed by these two instances. In the previous reduction, where 3SAT was reduced to PARTLY-REACH over *quad*, we did not specify the entire terminal instance. This was due to the fact that a solution of an underlying 3SAT problem can be read off the branches of the terminal instance. The difference here is that we avoid this problem, since the values of the germs $X_1, \ldots, X_m$ at $\hat{z}$ are the same in both instances Z and W. As a result, we need not to know an appropriate truth assignment that makes the underlying 3SAT formula C TRUE in advance. Instead, we have proved that the conditions for $Y_1, \ldots, Y_m$ tell us, whether there is an appropriate truth assignment that makes exactly one literal TRUE in each clause of C or not. The rest of the instances concerns the path and its constraint. The same values at the parts concerning Z_0 ensure the even winding number w.r.t. $0 \in \mathbb{C}$, while the different values for the parts concerning Z_{-n-2} and Z_1 force the corresponding path to circle around $-n-2$ and 1.

Now, assuming that we have got an efficient algorithm that solves the Reachability Problem, where only movements of bounded length are allowed, implies

Theorem 3.23. *The problem* SLP-REACH *over kubi, where only movements of a bounded length are allowed, is NP-hard.*

Remark 3.32. Similarly, we can strengthen this result in this subsection by not using arbitrary constants in the set of admitted SLP-operations, but 0 and 1. This works, since all required constants in the encoding SLP can be generated by a sub-SLP of appropriate polynomial length that only uses the constants 0 and 1.

4

Geometry and Numerical Integration

"$\frac{1}{r^2}$ has a nasty singularity at $r = 0$, but it did not bother Newton – the moon is far enough."

Edward Witten (AMS Gibbs Lecture, Baltimore, Jan. 7, 1998)

"The shortest route between two truths in the real domain passes through the complex domain."

Jacques Salomon Hadamard (1865-1963)

At first sight, one might ask what dynamic geometry has to do with numerical integration. A possible answer is that we can use such methods to compute the motion of an algebraic construction under a set of movements of the free objects. One just has to formulate an initial value problem (IVP), which describes this motion:

Let $\mathscr{C}$ be an algebraic construction according to Definition 2.3, $\vec{\mathbf{o}} \in \mathcal{I}_{\mathscr{C}}$ a regular starting instance and $\mu_{1-m}, \ldots, \mu_0$ corresponding explicitly given movements. Since the movements are restrictions of holomorphic functions, there are germs $c_{1-M}, \ldots, c_0 \in \mathbb{O}_0$ describing their coordinates on an appropriate neighborhood of $0 \in \mathbb{C}$. From Subsection 2.3, it can be concluded that the coordinates of the dependent objects are determined by polynomials

$$P_l \in \mathbb{C}[C_{1-M}, \ldots, C_{l-1}, X],$$

where $l \in \{1, \ldots, N\}$. Starting from this setting, an IVP can be inductively formulated, which governs the evolution of each coordinate during the movement and thus the motion of the construction $\mathscr{C}$ is described by it.

The ordinary differential equations (ODEs) capturing the coordinates $c_{1-M}, \ldots, c_0 \in \mathbb{O}_0$ of the movements are simply given by their derivatives. We now concentrate on P_1 to constitute an ODE for c_1. The coordinate function c_1 satisfies

$$P_1\big(c_{1-M}(t), \ldots, c_0(t), c_1(t)\big) \equiv 0$$

on an appropriate neighborhood of $0 \in \mathbb{C}$. As a result, deriving this equation yields an identity for the derivative $\dot{c}_1$. By an abuse of notation, we are going to omit the argument t to increase the readability. Then one obtains

$$\frac{d}{dt} P_1(c_{1-M}, \ldots, c_0, c_1) = \sum_{j=1-M}^{1} \dot{c}_j \cdot \underbrace{\left.\frac{\partial P_1(C_{1-M}, \ldots, C_1)}{\partial C_j}\right|_{C_{1-M}=c_{1-M}, \ldots, C_1=c_1}}_{=:P_{1,c_j}} \equiv 0.$$

We now observe that the derivative $P_{1,c_1} \not\equiv 0$. Otherwise this would mean that $P_1 \in \mathbb{C}[C_{1-M}, \ldots, C_0, X]$ is constant w.r.t. X implying the non-existence of c_1. As a result, we can rearrange the equation and get

$$\dot{c_1} = \frac{1}{P_{1,c_1}} \cdot \sum_{j=1-M}^{0} \dot{c}_j \cdot P_{1,c_j}.$$

The last step to obtain a first-order ODE for c_1 is to express the occurring derivatives $\dot{c}_{1-M}, \ldots, \dot{c}_0$ in terms of the variable t and $c_{1-M}, \ldots, c_0$. This is possible due to the fact that the coordinates of the movements $\mu_{1-m}, \ldots, \mu_0$ are given explicitly.

Inductively, this procedure applies straight-forward for each $l \in \{1, \ldots, N\}$ yielding a system of first-order ODEs that describes the dynamics of the construction $\mathscr{C}$ under the movements $\mu_{1-m}, \ldots, \mu_0$. If we additionally claim that the coordinate functions $c_{1-M}, \ldots, c_N$ form the instance $\vec{\mathbf{o}}$ for $t = 0$, an initial value problem will be defined.

Simulating them has to be done by an adequate numerical solver that usually works over the real numbers. Against the background of the successful complex strategy achieving continuity in dynamic geometry, it arises the question, whether a complex approach is also useful in the context of integrating ODEs. In order to answer this question, we imagine a restricted two-body problem, where a planet moves around a fixed star under NEWTON's law of gravity. If the initial velocity of the planet points directly towards the sun, the system will necessarily run into a singular situation. In addition, scenarios nearby this situation will cause numerically difficulties. A setup that also allows complex time, leaves the possibility to circumvent the singular situation by introducing a time flow that makes a detour through complex values. An example, where this strategy has been applied successfully is studied in [5]. Resolving singularities by applying complex detours at the numerical integration of the two-body problem mentioned was the initial motivation in [38].

The study of complex detours for ODEs entails several interesting issues. Assuming that a holomorphic function solves locally an ODE, the principle of continuity states that each analytic continuation of this function will solve the also analytically continued[22] ODE. For this reason, monodromic behavior can be caused by the right side of the ODE. If this is the case and one would like to integrate this ODE numerically, it must be ensured that the solver mimics the analyticity of the underlying equation. This requires the implementation of so-called tracing strategies for the right side of an ODE (see again [38]). However, there can be another source of monodromy, namely the solution itself. It even may induce monodromic behavior, though the right side does not. Although, such problems may provide an interesting mathematical structure over the complex numbers, they have been rarely studied in literature (for instance see [7]).

Besides the structural and implementation-related issues already mentioned, there is another worthwhile point to address. The complex approach also allows one to gain computational benefits. There are at least two different effects.

At first, recall the constellation, which causes an almost singular situation at the two-body problem. A numerical solver with adaptive step size control requires a quite high resolution, when the planet swings closely around the star caused by a nearby lying singularity. If we are just interested in a position of the planet, which lies afterwards the star is passed, we can reach this point in time by a complex detour that stays far away from this singularity. As a result, the step size control does not have to use such a high resolution (see [9]).

[22] This continuation concerns the ODE's right side.

A completely different effect that we are going to study throughout the rest of this chapter concerns superconvergence in the presence of certain complex time grids. We are going to exemplify this effect at the probably simplest possible scenario, this is, the ODE $\dot{x} = x$. Hereby, we just have one branch and can neglect the problems concerning monodromy. In this case, carefully chosen complex time grids can help to reduce the global integration error by at least one order. Motivated by this, we are going to develop a theory that explains these superconvergence effects. Based on the derived theory, a generalization to a non-linear setup is provided thereafter. By doing so, it will turn out that the complex geometric approach enables us to interpret the class of composition methods with complex coefficients as the application of basic methods along complex time grid. This is a different derivation of these methods apart from the usual way of solving some suitable set of order conditions over the reals or the complex numbers such as done in [23, 24, 58].

4.1 An Introductory Example

In this section, we would like to make the reader familiar with the subject of numerical integration along complex paths (and its benefits) in an informal, but hopefully self-explanatory way. In order to concentrate on the fundamental ideas, this and the following section is written in a very easy and self-containing manner.

We choose the simplest possible initial value problem

$$\dot{x}(t) = f\big(t, x(t)\big), \quad x(t_0) = x_0, \tag{4.14}$$

where $f : \mathbb{R}^2 \to \mathbb{R},\ (x, t) \mapsto x$ and $(t_0, x_0) := (0, 1)$. The corresponding analytic solution curve $\varphi : \mathbb{R} \to \mathbb{R},\ t \mapsto \varphi(t)$ is given by the well-known exponential function $\varphi(t) := e^t$ for all $t \in \mathbb{R}$.

In order to compute $\varphi(1)$, a standard approach is given by using an explicit Runge-Kutta method (eRKM) along an equidistant decomposition of the interval $[0, 1]$. Here, we refer to [13] for an extensive introduction to the topic of numerical integration of ODEs.

In contrast to the real interval $[0, 1]$, we now study the use of complex paths connecting 0 and 1. As (4.14) is a well-defined[23] initial value problem given by an autonomous linear first order differential equation with constant coefficients, $t \in \mathbb{R}$ seems to be an unnecessary restriction. To be more precisely, we observe that our IVP (4.14) can be written equivalently as the integral equation

$$x(t) = x_0 + \int_{t_0}^{t} x(s)\, ds = x_0 + \int_{\gamma} x(\zeta)\, d\zeta,$$

where $\gamma : [t_0, t] \to \mathbb{C},\ s \mapsto s$ is a path. As the solution φ has to be an entire function, Cauchy's well-known integral formula states that $\varphi(t)$ is independent of the detailed choice of γ, i.e. every path γ with starting point t_0 and terminal point t yields the same result for $\varphi(t)$.

With this "analytic" picture in mind, we investigate the behavior of the explicit Euler method[24] along the equidistant complex time grid

$$\tilde{t}_j := \frac{1}{2}\left(e^{i\pi\left(1-\frac{j}{n}\right)} + 1\right),$$

[23] For existence and uniqueness of a complex solution curve have a look at [54].

[24] It can be found in [13, p. 135].

where $j \in \{0, \ldots, n\}$. This is a discretization of the upper complex half circle from 0 to 1. Starting at the initial value $x_0 = 1$, the explicit EULER method generates n, not necessary real-valued, approximations

$$\tilde{x}_j \approx \varphi(\tilde{t}_j) = e^{\tilde{t}_j},$$

where again $j \in \{0, \ldots, n\}$. In the following, we compare the explicit EULER method along the mentioned complex time grid and the equidistant composition of the interval $[0, 1]$ given by

$$t_j := \frac{j}{n},$$

where $j \in \{0, \ldots, n\}$. Figure 4.17 illustrates the results of the explicit EULER along both the real and the complex time grid. On the one hand, one can see the exact values at the different grid points. The values corresponding to the complex and the real gird points are denoted by $+$ and $\times$, respectively, while their approximations are given by $\square$ and $\diamond$. The figure already indicates the observations that are finally proved by the numerical results show in Table 4.1. We point out the following observations:

1. As expected, both terminal points seem to be an approximation of $e \approx 2.718281828$.
2. The imaginary part of the complex terminal point seems to be zero.
3. The complex construction yields more correct digits. To be more precise, the complex terminal point $\tilde{x}_{10}$ is more than 16 times closer to e than the real terminal point.

Particularly, the last observation seems a worthwhile phenomenon. Below, we are going to develop a theory that explains this effect in the more general context of numerical integration of linear IVPs with constant coefficients by the use of eRKMs along complex paths of integration.

4.2 Complex Flows, Time Grids and Errors

Before we can concentrate on the discovered observations, we have to introduce the required notation in this section. As a first step, we define complex initial value problems.

Definition 4.15 (complex initial value problems)**.** *Let $d \in \mathbb{N}$, $\Omega_f \subseteq \mathbb{C} \times \mathbb{C}^d$ and $f : \Omega_f \to \mathbb{C}^d;\ (t, x) \mapsto f(t, x)$ be a continuous function. Then*

$$\dot{x}(t) = f\big(t, x(t)\big) \tag{4.15}$$

is called an explicit first order differential equation*. A* solution curve *of (4.15) is a complex differentiable function $\varphi : U \to \mathbb{C}^d;\ t \mapsto \varphi(t)$ with*

1. *$\Gamma_\varphi := \big\{\big(t, \varphi(t)\big)\big| t \in U\big\} \subseteq \Omega_f$ and*
2. *$\dot{\varphi}(t) = f\big(t, \varphi(t)\big)$ for all $t \in U$,*

where $U \subseteq \mathbb{C}$ is an open set. (4.15) together with a point $(t_0, x_0) \in \Omega_f$, the initial conditions*, is called a* complex initial value problem (CIVP)*. A* solution *to a CIVP is a solution curve $\varphi : U \to \mathbb{C}^d$ of (4.15), where $t_0 \in U$ and*

$$\varphi(t_0) = x_0.$$

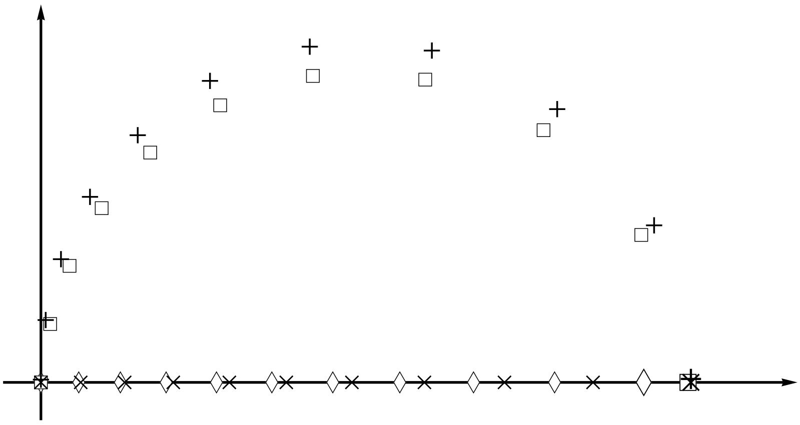

Figure 4.17: Results of the explicit EULER method applied to $\dot{x} = x$ with the initial value $x(0) = 1$. Thereby, the two previously introduced time grids $\left[\tilde{t}_0, \ldots, \tilde{t}_{10}\right]$ and $\left[t_0, \ldots, t_{10}\right]$ are used. The symbols $\square$ and $\diamond$ represent the computed approximations of the solution $\varphi(t) = e^t$ along the upper complex half circle and the real interval from 0 to 1, respectively. The symbols $+$ and $\times$ mark the corresponding exact values of the solution.

step j	real grid ($\diamond$)	complex grid ($\square$)
0	1	1
1	1.100000000	$1.024471742 + 0.1545084969 \cdot i$
2	1.210000000	$1.075693448 + 0.3082767551 \cdot i$
3	1.331000000	$1.160581914 + 0.4613658247 \cdot i$
4	1.464100000	$1.289582523 + 0.6080971485 \cdot i$
5	1.610510000	$1.473952784 + 0.7336116556 \cdot i$
6	1.771561000	$1.719643769 + 0.8108906981 \cdot i$
7	1.948717100	$2.016924082 + 0.8017873023 \cdot i$
8	2.143588810	$2.328718297 + 0.6673738888 \cdot i$
9	2.357947691	$2.587124642 + 0.3901842509 \cdot i$
10	2.593742460	$2.710722870 - 0.0000000006 \cdot i$

Table 4.1: The corresponding numeric results to computations illustrated in Fig. 4.17.

Remark 4.33. When the right side f is complex differentiable on an appropriate closed subset of $\mathbb{C} \times \mathbb{C}^d$ and it is additionally LIPSCHITZ continuous and bounded on this set, then it can be shown that there exists a local and unique solution to an associated CIVP. The corresponding theory can be found in [25].

Now, let γ be a given path with $\gamma(0) = t_0 \in \mathbb{C}$ and $\gamma(1) = t_1 \in \mathbb{C}$. Furthermore, we assume that for every $y \in N$, where N is a neighborhood of $x_0 \in \mathbb{C}^d$, there exists a local solution $\varphi_{(t_0,y_0)}$ to the CIVP

$$\dot{x}(t) = f\big(t, x(t)\big), \quad x(t_0) = y_0 \in \mathbb{C}^d.$$

For every such solution $\varphi_{(t_0,y_0)}$, let us additionally assume that the analytic continuation of $\varphi_{(t_0,y_0)}$ along γ exists and denote it by $\varphi^{\gamma}_{(t_0,y_0)}$. Due to this fact we have a map

$$\Phi^{\gamma}_f : N \to \mathbb{C}^d, \quad y_0 \mapsto \Phi^{\gamma}_f y_0 := \varphi^{\gamma}_{(t_0,y_0)}(t_1).$$

Remark 4.34. Since Φ^{γ}_f maps the initial value y_0 at t_0 to the corresponding state $\Phi^{\gamma}_f y_0$ at t_1, we call this map the *complex flow* (or the *evolution induced by f*). The complex flow can be interpreted as the path-dependent function mapping an initial value y_0 to the value of the analytic continuation $\varphi^{\gamma}_{(t_0,y_0)}$ evaluated at t_1.

In the introductory example, both of our computed approximations made use of a certain discretization of a given complex path $\gamma : [0,1] \to U \subseteq \mathbb{C}$. Since $\mathbb{C}$ is not equipped with an order relation, we adapt the real-valued concept of consecutive points in time $t_0 \leq \ldots \leq t_n \in \mathbb{R}$ by the use of indexed sets

$$\Delta := \big[t^{\Delta}_0, \ldots, t^{\Delta}_{n_\Delta}\big] \subset \mathbb{C}.$$

We call them *time grids* with $n_\Delta \in \mathbb{N}$ *time steps*

$$\tau^{\Delta}_j := t^{\Delta}_{j+1} - t^{\Delta}_j,$$

where $j \in \{0, \ldots, n_\Delta - 1\}$. In addition, we denote the *maximum step size* of a time grid Δ by

$$\tau_\Delta := \max_{j \in \{0,\ldots,n_\Delta - 1\}} \big|\tau^{\Delta}_j\big|.$$

Whenever it helps, we will omit the symbol Δ to increase readability throughout the rest of this chapter. In the example of the previous Section, our attempt was to construct a corresponding *grid function*

$$x_\Delta : \Delta \to \mathbb{C}^d; \quad t \mapsto x_\Delta(t),$$

which approximates a solution of a CIVP, i.e.

$$x_\Delta(t) \approx \varphi(t)$$

for all $t \in \Delta \subseteq U$, where $\varphi : U \to \mathbb{C}^d$ is a solution of (4.14). At first, we set $x_\Delta(t_0) := x_0$ and then we used EULER's idea of small "tangential" update steps to compute the missing $x_\Delta(t_j)$'s. To generalize this idea, we just have to replace the method for computing a new value $x_\Delta(t_{j+1})$ from a point $x_\Delta(t_j)$ already derived. For this purpose, we introduce a function

$$\Psi : V \to \mathbb{C}^d; \quad (t, s, x) \mapsto \Psi^{t,s} x,$$

where $V \subseteq \mathbb{C} \times \mathbb{C} \times \mathbb{C}^d$ is an adequate set. The variables s, t and x play the role of *current point in time*, *next point in time* and *current state or value*, respectively. At this point, one naturally assumes that $\Psi^{t,s}x$ is defined for every choice of $(t, s, x) \in \mathbb{C} \times \mathbb{C} \times \mathbb{C}^d$ with $(s, x) \in \Omega_f$ and $|t - s|$ being sufficiently small. Under these assumptions, Ψ is often denoted as the *discrete evolution* in analogy to the complex flow. It enables us to define

$$x_\Delta(t_{j+1}) := \Psi^{t_{j+1},t_j} x_\Delta(t_j)$$

for all $j \in \{0, \dots, n_\Delta - 1\}$. Whenever a grid function x_Δ is given by a such a process, we are going to say that x_Δ is generated by a *one-step method.* Moreover, we introduce the notation

$$\Psi^\Delta x_0 := x_\Delta(t_n)$$

in analogy to $\Phi^\gamma x_0$. This is the terminal point, which is reached, when the discrete evolution Ψ is developed along the time grid Δ.

The last issue we address in this section concerns approximations errors. Given a time grid $\Delta := [t_0, \dots, t_{n_\Delta}] \subset \mathbb{C}$ and let x_Δ be a grid function generated by a one-step method applied to (4.15). As we have already seen by the example in the previous section, the approximation process is expected to cause *grid errors* $\varepsilon_j := \varepsilon_\Delta(t_j)$, where $j \in \{0, \dots, n_\Delta\}$ and

$$\varepsilon_\Delta : \Delta \to \mathbb{C}^d; \quad t_j \mapsto \Phi^{t_j,t_0} x_0 - x_\Delta(t_j).$$

At this point, we define for all $0 \le i \le j \le n_\Delta$,

$$\Phi^{t_j,t_i} := \Phi^{\gamma_\Delta^{j,i}},$$

where $\gamma_\Delta^{j,i}$ represents the traverse sequentially visiting $t_i, \dots, t_j$. The grid errors are generated by the local inaccuracy of the discrete evolution Ψ – the so-called *consistence error.* It is denoted by[25]

$$\varepsilon(t, x, \tau) := \Phi^{t+\tau,t} x - \Psi^{t+\tau,t} x.$$

These local errors interact with the sensitivity of Φ^{t_j,t_0} to perturbations of x. In general, they can be damped, amplified or fortunately be extinguished throughout the one-step recursion. A special variant of the latter case will be analyzed in what follows.

4.3 Theory Concerning Linear ODEs

In this section, we are going to disclose the secrets of the introductory example in the more general context that is given by a CIVP of the form

$$\dot{x}(t) = Ax, \quad x(t_0) = x_0, \tag{4.16}$$

where $A \in \mathbb{C}^{d \times d}$, $d \in \mathbb{N}$ and $(t_0, x_0) \in \mathbb{C} \times \mathbb{C}^d$. Although, there is not a dependence of the complex flow $\Phi^{t,t_0} x_0$ on the path of integration from the analytic point of view, several numerical case studies have indicated that the order of convergence of a numerical approximation obtained by a Runge-Kutta method (RKM) seems to depend heavily on the detailed choice of the path of integration.

Let us denote the path of integration by γ. It shall connect a starting point t_0 and a terminal point t. For every $n \in \mathbb{N}$, we define the corresponding time grid

$$\Delta_n^\gamma := \left[\gamma(0), \gamma\left(\frac{1}{n}\right), \dots, \gamma\left(\frac{n-1}{n}\right), \gamma(1)\right].$$

[25] Since Ψ is a discrete evolution, $\varepsilon(t, x, \tau)$ is defined for all $(t, x) \in \Omega_f$, if τ becomes sufficiently small.

Now, an RKM of order $p \in \mathbb{N}$ can be applied along this grid Δ_n^γ inducing a grid function $x_{\Delta_n^\gamma}$. In the context of (4.16), the convergence theory of one-step methods assures that the family of grid functions

$$\left(x_{\Delta_n^\gamma}\right)_{n\in\mathbb{N}}$$

obtained by the used RKM converges to the solution $\varphi(t) := \Phi^{t,t_0}x_0$ with order p for an arbitrary choice of γ. Surprisingly, we are going to show in the following that for every RKM of order p, there exists a path of integration γ^* so that the order of convergence at the terminal point is at least $p+1$, i.e.

$$\varepsilon_n^* := \varepsilon_{\Delta_n^{\gamma^*}}(t) = \mathcal{O}\left(\frac{1}{n^{p+1}}\right).$$

This is a superconvergence effect, which only affects the terminal point. In contrast to the terminal point, the maximum grid error has not to be of order $p+1$. Furthermore, γ^* can be chosen so that the length $|\gamma^*|$ of γ^* satisfies

$$|\gamma^*| \le \frac{|t-t_0|\cdot\pi}{2}$$

implying its practical interest.

4.3.1 Effects of Complex Conjugation

However, before the mentioned superconvergence is dealt with, we study, when the computed terminal point is indeed real. This seemed to be case at the introductory example.

Let us therefore have a look at the detailed structure of the discrete evolution $\Psi^{t+\tau,t}x$ of an arbitrary s-stage eRKM applied to problem (4.16). For $i,j \in \{1,\dots,s\}$, the coefficients of the eRKM are denoted by $\mathfrak{A}_{i,j}, \mathfrak{b}_i \in \mathbb{C}$. Then the discrete evolution is given by

$$\Psi^{t+\tau,t}x = x + \sum_{i=1}^{s} \mathfrak{b}_i\, k_i,$$

where

$$k_i := A\,x + \tau \sum_{j=1}^{i-1} \mathfrak{A}_{i,j}\, A\, k_j$$

is the *i-th stage* of the used eRKM. Our first lemma that we are going to show is a structural result concerning the discrete evolution. More precisely, applying an eRKM to (4.16) yields that the evolution of one time step $\tau \in \mathbb{C}$ from an initial point $x \in \mathbb{C}^d$ can be expressed as a simple matrix multiplication Mx. In addition, the matrix M can be expressed as a matrix polynomial in τA. The formal statement is captured by the lemma below.

Lemma 4.22. *Let $s \in \mathbb{N}$ and $\Psi^{t+\tau,t}x$ be the discrete evolution induced by an arbitrary s-stage eRKM with coefficients $\mathfrak{A} \in \mathbb{C}^{s\times s}$ and $\mathfrak{b},\mathfrak{c} \in \mathbb{C}^s$. If this method is applied to problem (4.16), then for all $x \in \mathbb{C}^d$, it holds*

$$\Psi^{t+\tau,t}x = P(\tau A)\,x,$$

where P is a polynomial of degree s with coefficients from

$$\mathbb{Q}[\mathfrak{A}_{2,1},\mathfrak{A}_{3,1},\mathfrak{A}_{3,2},\dots,\mathfrak{A}_{s,1},\dots,\mathfrak{A}_{s,s-1},\mathfrak{b}_1,\dots,\mathfrak{b}_s].$$

Proof. Let us first have a look at the stages k_i. By induction over $i \in \{1, \ldots, s\}$, we are going to show that

$$k_i = P_i(\tau A)\, A\, x,$$

where P_i is a polynomial of degree $i-1$ with coefficients in $\mathbb{Q}[\mathfrak{A}_{2,1}, \ldots, \mathfrak{A}_{i,1}, \ldots, \mathfrak{A}_{i,i-1}]$. Considering the first stage of the eRKM yields

$$k_1 = Ax = P_1(\tau A)\, A\, x,$$

where $P_1 := 1$. Let now $i \in \{2, \ldots, s\}$. It follows from the induction hypothesis that

$$\begin{aligned} k_i &= A\, x + \tau \sum_{j=1}^{i-1} \mathfrak{A}_{i,j}\, A\, k_j \\ &= A\, x + \tau \sum_{j=1}^{i-1} \mathfrak{A}_{i,j}\, A\, P_j(\tau A)\, A\, x \\ &= \left(I + \sum_{j=1}^{i-1} \mathfrak{A}_{i,j}\, \tau A\, P_j(\tau A) \right) A\, x. \end{aligned}$$

As one can easily see, the expression in the bracket is induced by a polynomial P_i satisfying the desired conditions. In conclusion,

$$\Psi^{t+\tau,t} x = x + \tau \sum_{i=1}^{s} \mathfrak{b}_i\, k_i = \left(I + \sum_{i=1}^{s} \mathfrak{b}_i\, P_i(\tau A)\, \tau A \right) x$$

yields the desired claim. At this point, we observe that there is not a problem of commutativity in the matrix polynomials, because of $A^n A^m = A^m A^n$ for all $n, m \in \mathbb{N}_0$. ☐

Remark 4.35. The polynomial P of the previous lemma is also known as the *stability function* of the underlying eRKM.

A further ingredient that was crucial in the introductory example concerns the used complex time grid. It was of a special form, which is formalized in the next lemma.

Definition 4.16 (symmetric time grids)**.** *Let* $\Delta := \left[t_0^\Delta, \ldots, t_{n_\Delta}^\Delta\right]$ *be a time grid. We call* Δ *a* symmetric *time grid, if there exists an involutory permutation* $\pi \in \mathcal{S}_{n_\Delta}$*, i.e.* $\pi^2 = \mathrm{id}$*, so that*

$$\tau_j^\Delta = \overline{\tau_{\pi(j+1)-1}^\Delta}$$

for all $j \in \{0, \ldots, n_\Delta - 1\}$*.*

Based on the structural result concerning eRKMs and symmetric time grids, we are able to prove the desired theorem, in which we link terminal points and coefficient of eRKMs.

Theorem 4.24. *Given a linear CIVP (4.16) with real coefficients, i.e.* $A \in \mathbb{R}^{d \times d}$ *and real initial values, i.e.* $t_0 \in \mathbb{R}$ *and* $x_0 \in \mathbb{R}^d$*. Furthermore, let* $n := n_\Delta \in \mathbb{N}$*,* $\Delta := \Delta_n^\gamma$ *be a symmetric time grid with* $t_0^\Delta = t_0 \in \mathbb{R}$ *and* $t_n^\Delta \in \mathbb{R}$ *and* x_Δ *be a corresponding grid function constructed by an* s*-stage eRKM with real coefficients, i.e.* $\mathfrak{A} \in \mathbb{R}^{s \times s}$ *and* $\mathfrak{b}, \mathfrak{c} \in \mathbb{R}^s$*. Then the terminal point of the grid function* x_Δ *is real as well, i.e.*

$$x_\Delta\left(t_n^\Delta\right) \in \mathbb{R}^d.$$

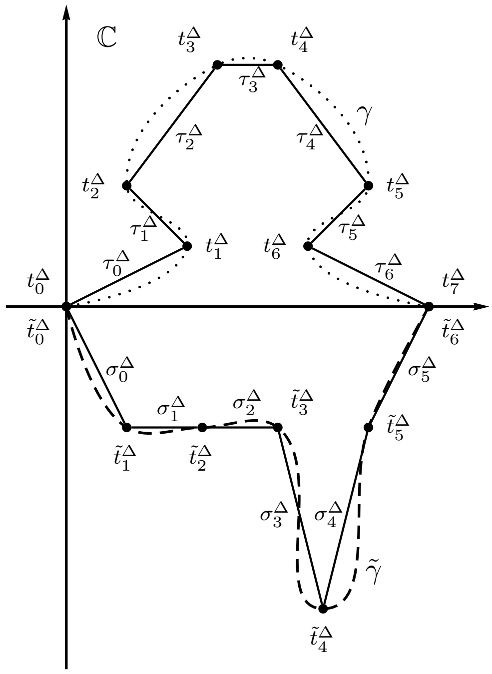

Figure 4.18: Two symmetric time grids $\Delta_7^\gamma = \left[t_0^\Delta, \ldots, t_7^\Delta\right]$ and $\Delta_6^{\tilde{\gamma}} = \left[\tilde{t}_0^\Delta, \ldots, \tilde{t}_6^\Delta\right]$ along γ (dotted) and $\tilde{\gamma}$ (dashed), respectively.

Proof. We denote $\tau_j := \tau_j^\Delta$ for $j \in \{0, \ldots, n-1\}$. From Lemma 4.22, it can be concluded that

$$x_n = \left(\prod_{j=0}^{n-1} P\left(\tau_j A\right)\right) \cdot x_0,$$

where P is a polynomial of degree s with real-valued coefficients. Due to this fact it is sufficient to show that the product of polynomials is in $\mathbb{R}^{s\times s}$. As a first step, we observe that

$$P(\overline{\tau} A) = \overline{P(\tau A)} \quad \text{and} \quad P(\tau A)P(\sigma A) = P(\sigma A)P(\tau A)$$

hold for all $\sigma, \tau \in \mathbb{C}$. Taking these identities into account, we separate the product as follows and use the symmetry of the time grid:

$$\begin{aligned}
\prod_{j=0}^{n-1} P\left(\tau_j A\right) &= \prod_{\substack{j\in\{0,\ldots,n-1\}\\ j+1=\pi(j+1)}} \overbrace{P\left(\tau_j A\right)}^{\in\mathbb{R}^{s\times s}} \cdot \prod_{\substack{j\in\{0,\ldots,n-1\}\\ j+1<\pi(j+1)}} P\left(\tau_j A\right) P\left(\tau_{\pi(j+1)-1} A\right) \\
&= \prod_{\substack{j\in\{0,\ldots,n-1\}\\ j+1=\pi(j+1)}} \underbrace{P\left(\tau_j A\right)}_{\in\mathbb{R}^{s\times s}} \cdot \prod_{\substack{j\in\{0,\ldots,n-1\}\\ j+1<\pi(j+1)}} \underbrace{P\left(\tau_j A\right) \overline{P\left(\tau_j A\right)}}_{\in\mathbb{R}^{s\times s}},
\end{aligned}$$

where $\pi \in \mathcal{S}_n$ is the permutation corresponding to the symmetric time grid Δ. Pay attention to the fact that for every $M \in \mathbb{C}^{s\times s}$, it holds that $M\overline{M} = \overline{M}M \Rightarrow M\overline{M} \in \mathbb{R}^{s\times s}$. □

Theorem 4.24 justifies the second observation of the introductory example in the more general context of a CIVP given by an explicit linear autonomous system of first order differential equations. There the used complex time grid was chosen to be symmetric along a half circle in the complex plane.

4.3.2 The Terminal Error

This section is dedicated to the third observation of the introductory example, this is, accuracy gain at the terminal point. The main idea is to represent the leading term of the error at the terminal point in terms of the step sizes. Then an appropriate choice of the step sizes can cancel the leading error term and therefore the order of convergence increases. In order to derive the desired representation, we are going to expand the error at the terminal point, this is, the difference of the complex flow and the discrete evolution.

Theorem 4.25 (terminal error)**.** *Let γ be a rectifiable curve with starting point $\gamma(0) = t_0 \in \mathbb{C}$ and terminal point $\gamma(1) = t \in \mathbb{C}$. For every $n \in \mathbb{N}$, we denote the time grid Δ_n^γ by Δ, abbreviate the step sizes by*

$$\tau_j(n) := t_{j+1}^\Delta - t_j^\Delta,$$

where $j \in \{0, \dots, n-1\}$ and define δ_n to be the maximum step size, i.e. $\delta_n := \tau_\Delta$. Furthermore, let $L(\gamma)$ be the length of γ, i.e.

$$L(\gamma) := \int_0^1 |\dot{\gamma}(t)| \, dt$$

and $C > 0$ a positive constant so that

$$\delta_n \leq C \cdot \frac{L(\gamma)}{n}.$$

Then the grid error ε_n at the ending point $t \in \mathbb{C}$ corresponding to a p-stage eRKM of order $p \in \mathbb{N}$ applied to problem (4.16) along Δ satisfies

$$\lim_{n\to\infty} \frac{\varepsilon_n}{\delta_n^p} = \lim_{n\to\infty} \frac{1}{\delta_n^p} \sum_{j=0}^{n-1} \big(\tau_j(n)\big)^{p+1} \cdot \frac{A^{p+1} e^{(t-t_0)A}}{(p+1)!} x_0. \tag{4.17}$$

Proof. To increase readability, we divide the exponential series for all $n \in \mathbb{N}$ into the two terms

$$\psi_j(n) := \sum_{k=0}^{p} \frac{\big(\tau_j(n)A\big)^k}{k!} \quad \text{and} \quad \omega_j(n) := \sum_{k=p+1}^{\infty} \frac{\big(\tau_j(n)A\big)^k}{k!},$$

where $j \in \{0, \dots, n-1\}$. Moreover, we set $E := e^{CL(\gamma)\|A\|}$ and denote the terminal approximation of the discrete evolution by $x_n := \Psi^\Delta x_0$. As a first step, we have a look at the quotient

$$\begin{aligned}
\frac{\varepsilon_n}{\delta_n^p} &= \frac{\Phi^{t,t_0} x_0 - x_n}{\delta_n^p} = \frac{1}{\delta_n^p} \left[e^{(t-t_0)A} x_0 - x_n \right] \\
&= \frac{1}{\delta_n^p} \left[\prod_{j=0}^{n-1} e^{\tau_j(n)A} - \prod_{j=0}^{n-1} \psi_j(n) \right] x_0 \\
&= \frac{1}{\delta_n^p} \left[\prod_{j=0}^{n-1} \big(\psi_j(n) + \omega_j(n)\big) - \prod_{j=0}^{n-1} \psi_j(n) \right] x_0.
\end{aligned}$$

Expanding the first product yields

$$\frac{\varepsilon_n}{\delta_n^p} = \frac{1}{\delta_n^p}\left[\sum_{j=0}^{n-1}\omega_j(n)\prod_{\substack{l=0\\l\neq j}}^{n-1}\psi_l(n) + \rho(n)\right]x_0. \tag{4.18}$$

Hereby, the remaining term $\rho(n)$ satisfies

$$\begin{aligned}
\frac{|\rho(n)|}{\delta_n^p} &\leq \frac{E}{\delta_n^p}\sum_{r=2}^{n}\binom{n}{r}\left(\sum_{k=p+1}^{\infty}\frac{\left(CL(\gamma)\,\|A\|\,\right)^k}{k!\cdot n^k}\right)^r \\
&\leq \frac{E}{\delta_n^p}\sum_{r=2}^{n}\underbrace{\frac{n(n-1)\dots(n-r+1)}{r!\cdot n^r}}_{\leq 1}\left(n\sum_{k=p+1}^{\infty}\frac{\left(CL(\gamma)\,\|A\|\,\right)^k}{k!\cdot n^k}\right)^r \\
&\leq \frac{E\cdot n^p}{|t-t_0|^p}\sum_{r=2}^{n}\left(\frac{E}{n^p}\right)^r \leq \frac{E\cdot n^p}{|t-t_0|^p}\left(\frac{1}{1-\frac{E}{n^p}} - 1 - \frac{E}{n^p}\right) = \frac{E}{|t-t_0|^p}\cdot\frac{E^2}{n^p-E},
\end{aligned}$$

which, in turn, implies

$$\lim_{n\to\infty}\frac{\rho(n)}{\delta_n^p}x_0 = 0.$$

As a result, we can neglect $\rho(n)$ in our further considerations. In the next steps, we are going to simplify (4.18), namely we show that the missing $\psi_j(n)$ (products) and the higher order terms of the $\omega_j(n)$'s in (4.18) can be omitted as n tends to infinity. Using

$$\begin{aligned}
&\frac{1}{\delta_n^p}\left\|\sum_{j=0}^{n-1}\omega_j(n)\cdot x_n - \sum_{j=0}^{n-1}\omega_j(n)\prod_{\substack{l=0\\l\neq j}}^{n-1}\psi_l(n)\cdot x_0\right\| \\
&\leq \sum_{j=0}^{n-1}\left[\left\|\frac{\omega_j(n)}{\delta_n^p}\right\|\cdot\left\|x_n - \prod_{\substack{l=0\\l\neq j}}^{n-1}\psi_l(n)\cdot x_0\right\|\right] \leq \sum_{j=0}^{n-1}\left[\left\|\frac{\omega_j(n)}{\delta_n^p}\right\|\cdot\left\|\prod_{l=0}^{n-1}\psi_l(n) - \prod_{\substack{l=0\\l\neq j}}^{n-1}\psi_l(n)\right\|\right]\cdot\|x_0\| \\
&\leq \sum_{j=0}^{n-1}\left[\frac{\sum_{k=p+1}^{\infty}\frac{(CL(\gamma)\|A\|)^k}{k!\cdot n^{k-p}}}{|t-t_0|^p}\cdot\left\|\sum_{k=0}^{p}\frac{\left(\tau_j(n)A\right)^k}{k!} - I\right\|\cdot\left\|\prod_{\substack{l=0\\l\neq j}}^{n-1}\psi_l(n)\right\|\right]\cdot\|x_0\| \\
&\leq \sum_{j=0}^{n-1}\frac{E}{n\,|t-t_0|^p}\cdot\frac{E}{n}\cdot E\,\|x_0\| = \frac{E^3\,\|x_0\|}{|t-t_0|^p}\cdot\frac{1}{n},
\end{aligned}$$

yields that

$$\begin{aligned}
\lim_{n\to\infty}\frac{\varepsilon_n}{\delta_n^p} &= \lim_{n\to\infty}\frac{1}{\delta_n^p}\sum_{j=0}^{n-1}\omega_j(n)\cdot x_n \\
&= \lim_{n\to\infty}\left[\frac{1}{\delta_n^p}\sum_{j=0}^{n-1}\frac{\left(\tau_j(n)A\right)^{p+1}}{(p+1)!} + \frac{1}{\delta_n^p}\sum_{j=0}^{n-1}\sum_{k=p+2}^{\infty}\frac{\left(\tau_j(n)A\right)^k}{k!}\right]x_n.
\end{aligned}$$

Finally, the estimate

$$\left| \frac{1}{\delta_n^p} \sum_{j=0}^{n-1} \sum_{k=p+2}^{\infty} \frac{\big(\tau_j(n)A\big)^k}{k!} \right| \leq \frac{1}{|t-t_0|^p} \sum_{j=0}^{n-1} \sum_{k=p+2}^{\infty} \frac{n^p \big(CL(\gamma)\,\|A\|\,\big)^k}{k! \cdot n^k}$$

$$\leq \frac{1}{|t-t_0|^p \cdot n^2} \sum_{j=0}^{n-1} E = \frac{E}{|t-t_0|^p} \cdot \frac{1}{n}$$

and

$$\lim_{n\to\infty} x_n = \Phi^{t,t_0} x_0 = e^{(t-t_0)A} x_0$$

show the desired claim. □

Remark 4.36 (s-stage eRKMs & implicit RKMs). The proof of the previous theorem bases on an eRKM of order $p \in \mathbb{N}$ with p stages. However, what does happen, if the number of stages of the eRKM is larger than its order of convergence or the underlying method is an implicit RKM?

In the case of an s-stage eRKM of order $p \in \mathbb{N}$ with $p < s$, one can derive an analogous statement to Theorem 4.25. We have omitted this case for a better readability of the preceding proof. The only difference to the results above is given by a constant $C \in \mathbb{C}^{s\times s}$, the $p+1$ coefficient of the polynomial $P(\tau A))$ associated to the RKM so that

$$\lim_{n\to\infty} \frac{\varepsilon_n}{\delta_n^p} = \lim_{n\to\infty} \frac{1}{\delta_n^p} \sum_{j=0}^{n-1} \big(\tau_j(n)\big)^{p+1} \cdot e^{(t-t_0)A} \left[\frac{A^{p+1}}{(p+1)!} - C \right] x_0.$$

The corresponding proof is achieved by some additional estimations of higher order terms of $\frac{\varepsilon_n}{\delta_n^p}$ analogously to the ones seen previously.

Similarly, one can show that the structure of the error expansion (4.17) for an s-stage implicit RKM is also of the form

$$\lim_{n\to\infty} \frac{1}{\delta_n^p} \sum_{j=0}^{n-1} \big(\tau_j(n)\big)^{p+1} \cdot C(t, t_0, A) x_0,$$

where $C(t, t_0, A) \in \mathbb{C}^{s\times s}$.

The structure of the error expansion in Theorem 4.25 provides us a condition to increase the order $p \in \mathbb{N}$ of an RKM applied to (4.16). If the used step sizes are chosen so that

$$\sum_{j=0}^{n-1} \big(\tau_j(n)\big)^{p+1} = 0$$

for all $n \geq N \in \mathbb{N}_0$, then the order of convergence increases from p to $p+1$ at the terminal point, i.e.

$$\lim_{n\to\infty} \frac{\varepsilon_n}{\delta_n^p} = \lim_{n\to\infty} \frac{\Phi^{t,t_0} x_0 - \Psi^{\Delta} x_0}{\delta_n^p} = 0.$$

4.3.3 Superconvergent Time Grids

Motivated by the previous observations, our next goal is obvious: We would like to construct complex time grids that provide us superconvergence for linear ODEs. We consider an s-stage RKM with convergence order $p \in \mathbb{N}$ and use it to compute an approximation of the complex flow $\Phi^{t,t_0} x_0$ corresponding to problem (4.16). Thereby, we admit $t \in \mathbb{C}$. Basically, we are interested in time grid families

$$\left(\Delta = \left[t_0^\Delta, \dots, t_{n_\Delta}^\Delta\right]\right)_{n \in \mathbb{N}}$$

with three features:

1. First and foremost, the time grid shall connect the desired starting and terminal point, i.e. $t_0 = t_0^\Delta$ and $t = t_{n_\Delta}^\Delta$. In other words, it shall hold
$$\sum_{j=0}^{n_\Delta - 1} \tau_j^\Delta = t - t_0.$$
2. According to the prerequisite of Theorem 4.25, the maximal step size τ_Δ of the time grid must decrease at least like n_Δ^{-1} as $n_\Delta \to \infty$, i.e.
$$\tau_\Delta \leq \frac{D}{n_\Delta}$$
for some $D > 0$.
3. Last but not least, the time grid shall provide superconvergence. Due to this fact the step sizes shall satisfy
$$\sum_{j=0}^{n_\Delta - 1} \left(\tau_j^\Delta\right)^{p+1} = 0.$$

Fortunately, the superconvergence criterion can be interpreted geometrically which enables us to derive a canonical family of time grids for every convergence order $p \in \mathbb{N}$ of an RKM. We use the fact that all n-th roots of unity sum up to zero. More precisely, we set $\zeta_n(j) := e^{2\pi i \frac{j}{n}}$ for $j \in \mathbb{Z}$. Then it holds for every $\alpha \in \mathbb{C}$ that

$$\sum_{j=0}^{n-1} \alpha \cdot \zeta_n(j) = 0.$$

Now, we choose the time grid Δ so that the corresponding exponentiated step sizes $(\tau_j^\Delta)^{p+1}$ $(j \in \{0, \dots, n_\Delta - 1\})$ become such a collection of roots of unity. As illustrated by Figure 4.19, n consecutive $n(p+1)$-th roots of unity $\zeta_{n(p+1)}(k), \dots, \zeta_{n(p+1)}(k + (n-1))$, are transformed by complex exponentiation $z \mapsto z^{p+1}$ to the set of n-th roots of unity. Formally, one obtains the following lemma.

Lemma 4.23. *Let $n, p \in \mathbb{N}$, $k \in \mathbb{Z}$ and $\alpha \in \mathbb{C}$. Then it holds*

$$\sum_{j=k}^{k+(n-1)} \left(\alpha \cdot \zeta_{n(p+1)}(j)\right)^{p+1} = 0.$$

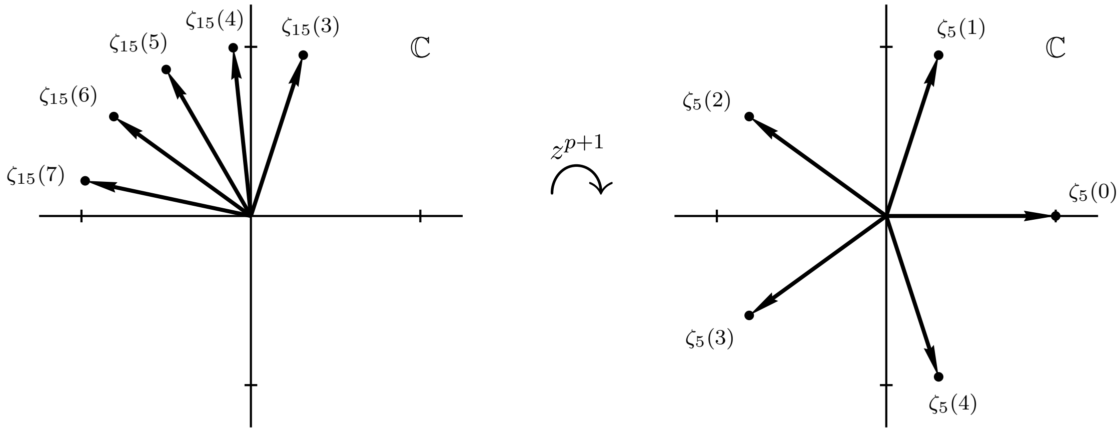

Figure 4.19: Transformation for $n = 5$, $p = 2$ and $k = 3$.

Proof. The claim is proved by the equations below:

$$\begin{aligned}
\sum_{j=k}^{k+(n-1)} \left(\alpha \cdot \zeta_{n(p+1)}(j)\right)^{p+1} &= \alpha^{p+1} \sum_{j=k}^{k+(n-1)} \left(e^{2\pi i \frac{j}{n(p+1)}}\right)^{p+1} = \alpha^{p+1} \sum_{j=k}^{k+(n-1)} e^{2\pi i \frac{j}{n}} \\
&= \alpha^{p+1} \sum_{j=k}^{n-1} e^{2\pi i \frac{j}{n}} + \alpha^{p+1} \sum_{j=n}^{k+(n-1)} e^{2\pi i \frac{j}{n}} \\
&= \alpha^{p+1} \sum_{j=k}^{n-1} e^{2\pi i \frac{j}{n}} + \alpha^{p+1} \sum_{l=0}^{k-1} e^{2\pi i \frac{l+n}{n}} \\
&= \alpha^{p+1} \sum_{j=0}^{n-1} e^{2\pi i \frac{j}{n}} = 0.
\end{aligned}$$

□

In conclusion, one way to build a superconvergent time grid $\Delta = \left[t_0^\Delta, \dots, t_{n_\Delta}^\Delta\right]$ is to take time steps $\tau_1^\Delta, \dots, \tau_{n_\Delta}^\Delta$ represented by n consecutive $n(p+1)$-th roots of unity which have been suitable scaled and rotated by a factor $\alpha \in \mathbb{C}$ so that they connect the starting and the terminal point, i.e.

$$\sum_{j=0}^{n_\Delta - 1} \tau_j^\Delta = t - t_0$$

Figure 4.20 illustrates this situation for $n = 5$, $p = 2$ and $k = 3$. In general, an appropriate choice for this factor is

$$\alpha = \frac{t - t_0}{\sum_{j=k}^{k+(n-1)} \zeta_{n(p+1)}(j)}.$$

Up to here, it has been easier to focus on the time steps $\tau_1^\Delta, \dots, \tau_{n_\Delta}^\Delta$ in order to derive a superconvergent family of time grids, but now we change our point of view from the time steps to the time grid elements $t_1^\Delta, \dots, t_{n_\Delta}^\Delta$. As exemplified in Figure 4.20, a superconvergent choice of the time steps $\tau_1^\Delta, \dots \tau_{n_\Delta}^\Delta$ induces special positions of the associated time grid

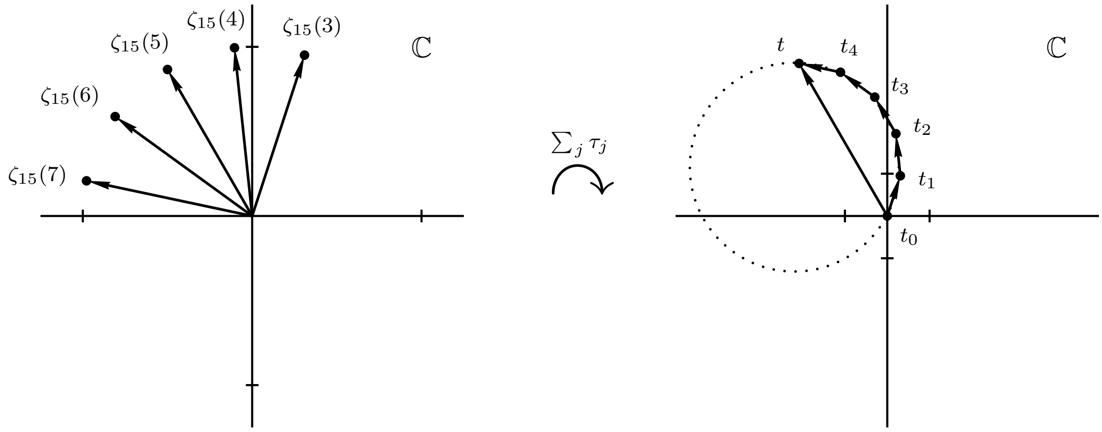

Figure 4.20: Counterparts for $n = 5$, $p = 2$ and $k = 3$ (no global scaling and rotation).

elements $t_1^\Delta, \dots, t_{n_\Delta}^\Delta$. They are located on a circle segment that depends on the order $p \in \mathbb{N}$ of the underlying method and the starting and terminal point $t_0, t \in \mathbb{C}$. Figure 4.20 shows the situation for $p = 2$. There, the corresponding time grid elements lie on a third segment of a circle. In the introductory example, the explicit EULER method has been used, i.e. $p = 1$, and the time grid elements have been lain on a half circle segment. This indicates that if the order of the underlying method is $p \in \mathbb{N}$, locating the time grid elements t_j^Δ equidistantly on a $1/(p+1)$-th circle segment connecting $t_0, t \in \mathbb{C}$ provides the desired superconvergence effect. The theorem below formalizes this idea.

Theorem 4.26. *Let $\gamma^* := \gamma_{t_0,t}^p$, where*

$$\gamma_{t_0,t}^p : [0,1] \to \mathbb{C};\; x \mapsto \frac{t_0 - t}{2 \cdot i \cdot \sin\left(\frac{\pi}{p+1}\right)} \left[e^{i\pi\left(\frac{1-2x}{p+1}\right)} - \cos\left(\frac{\pi}{p+1}\right)\right] + \frac{t_0 + t}{2}. \tag{4.19}$$

Furthermore, let $n \in \mathbb{N}$ and $x_{\Delta_n^{\gamma^*}}$ be the grid function generated by an s-stage RKM of order $p \in \mathbb{N}$ applied to (4.16). Then

$$t_0 = \gamma^*(0) \quad \textit{and} \quad t = \gamma^*(1).$$

and

$$\lim_{n\to\infty} \frac{\Phi^{t,t_0} x_0 - x_{\Delta_n^{\gamma^*}}(t)}{\delta_n^p} = 0.$$

Proof. We just must show that $\Delta_n^{\gamma^*}$ satisfies the three criterions on page 100. By the use of EULER's formula, a simple calculation shows that $t_0 = \gamma^*(0)$ and $t = \gamma^*(1)$. Moreover,

$$\tau_j^{\Delta_n^{\gamma^*}} = \frac{t_0 - t}{2 \cdot i \cdot \sin \frac{\pi}{p+1}} \left(e^{i\pi \frac{1 - 2\frac{j+1}{n}}{p+1}} - e^{i\pi \frac{1-2\frac{j}{n}}{p+1}} \right)$$

for all $j \in \{0, \dots, n-1\}$, so it is not hard to see that

$$\left| \tau_j^{\Delta_n^{\gamma^*}} \right| = \mathrm{const}_{t_0,t,p} \cdot \left| e^{-\frac{2\pi i}{n(p+1)}} - 1 \right|$$

for all $j \in \{0, \ldots, n-1\}$, which implies $\delta_n \in \Theta(n^{-1})$ and additionally the equidistance of the time steps

$$\tau_1^{\Delta_n^{\gamma^*}}, \ldots, \tau_n^{\Delta_n^{\gamma^*}}.$$

Finally, it is sufficient to show that

$$\sum_{j=0}^{n-1} \left(\tau_j^{\Delta_n^{\gamma^*}}\right)^{p+1} = \sum_{j=0}^{n-1} \left(t_{j+1}^{\Delta_n^{\gamma^*}} - t_j^{\Delta_n^{\gamma^*}}\right)^{p+1} = 0.$$

It holds that

$$\begin{aligned}
\sum_{j=0}^{n-1} \left(\tau_j^{\Delta_n^{\gamma^*}}\right)^{p+1} = 0 \quad &\Longleftarrow \quad \sum_{j=0}^{n-1} \left(e^{-\frac{2\pi i}{p+1}\cdot\frac{j+1}{n}} - e^{-\frac{2\pi i}{p+1}\cdot\frac{j}{n}}\right)^{p+1} = 0 \\
&\Longleftrightarrow \quad \sum_{j=0}^{n-1} \left(e^{-\frac{2\pi i}{p+1}\cdot\frac{j}{n} - \frac{2\pi i}{p+1}\cdot\frac{1}{n}} - e^{-\frac{2\pi i}{p+1}\cdot\frac{j}{n}}\right)^{p+1} = 0 \\
&\Longleftrightarrow \quad \left(e^{-\frac{2\pi i}{p+1}\cdot\frac{1}{n}} - 1\right)^{p+1} \underbrace{\sum_{j=0}^{n-1} e^{-2\pi i\frac{j}{n}}}_{=:(*)} = 0.
\end{aligned}$$

Since $(*)$ is the sum of the n-th roots of unity, the desired claim follows from Theorem 4.25. □

The last theorem yields immediately the following corollary, in which the superconvergence effect is stated.

Corollary 4.9 (superconvergence of RKMs)**.** *Given a* Runge-Kutta *method of order $p \in \mathbb{N}$ applied to (4.16) along γ^*. The approximation at the terminal point is of order $p+1$.*

Remark 4.37. As complex conjugation does not affect the argumentation so far, the previous corollary holds also for $\gamma^* := \overline{\gamma^*}$, where $\overline{\gamma^*}(t) := \overline{\gamma^*(t)}$ for all $t \in [0, 1]$.

4.4 Nonlinear Case

As we have seen in the last section, RKMs are superconvergent along certain complex time grids, if the right-hand side of the corresponding CIVP is linear. Consequently, it is natural to ask, whether this feature transfers to the case of an arbitrary right-hand side. Our study was again triggered by several numerical experiments. These experiments suggested an adapted version of the previous superconvergence statement, namely it still holds locally in the nonlinear case.

The main idea is to study the order of convergence not for infinitely many time steps, i.e. $n_\Delta \to \infty$, as in the linear case, but for an infinitely small macro step size, i.e. $t \to t_0$. Therefore, let now $n_\Delta = k \in \mathbb{N}$ be a fixed number of time steps and set $h := t - t_0$. Roughly speaking, the superconvergence effect is independent of $|t - t_0|$ in the linear case, this is,

$$\varepsilon_n \in O\left(\frac{1}{n_\Delta^{p+1}}\right).$$

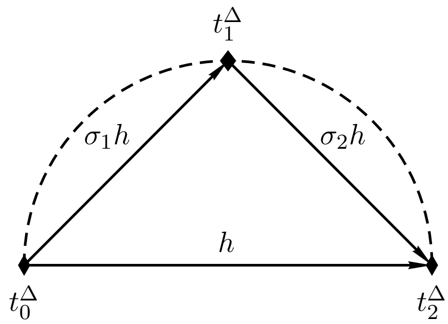

Figure 4.21: Scaled path segment from t_0^Δ to t_2^Δ for $p = 1$ and $k = 2$.

In contrast, we will observe in the nonlinear case that

$$\varepsilon_n \in O\left(|h|^{p+1}\right)$$

holds, if the integration is performed along an equidistant time grid induced by an appropriate chosen path $\gamma^p_{t_0,t}$ according to (4.19).

But to start with, we should exemplify the idea at the probably simplest case, this is, two micro time steps ($k = 2$) along a half circle ($p = 1$). Figure 4.21 illustrates this situation, in which a numerical integration along a discretization Δ is done by two Runge-Kutta iterations. If we regard this path as a scaled and rotated copy of a normalized[26] superconvergent circle segment, we can express the two time steps of Δ as

$$t_1^\Delta - t_0^\Delta = \sigma_1 h \quad \text{and} \quad t_2^\Delta - t_1^\Delta = \sigma_2 h,$$

where $\sigma_1 = \frac{1}{2} + \frac{1}{2} \cdot i$ and $\sigma_2 = \frac{1}{2} - \frac{1}{2} \cdot i$ are the two time steps of the grid

$$\Delta_2^{\gamma^1_{0,1}} = \left[0, \frac{1}{2} + \frac{1}{2} \cdot i, 1\right] \subset \mathbb{C}.$$

The general case $k > 2$ generalizes straight forward. For $l \in \{1, \dots, k\}$, let σ_l be the l-th time step of the normalized superconvergent time grid

$$\Delta_k^{\gamma^p_{0,1}}.$$

Due to this fact the corresponding scaled time steps can be expressed by $\sigma_l h$, where $h := t - t_0$.

4.4.1 Composition Methods

We now are able to draw a connection between composition methods and our idea of numerical integration along complex paths. Since it is not a restriction at all, we assume an autonomous problem to simplify the notation. Let $\Psi^\tau := \Psi^{0+\tau,0}$ be the underlying discrete evolution of a basic RKM of order $p \in \mathbb{N}$ and set

$$\Delta := \Delta_k^{\gamma^p_{0,h}}.$$

[26]This means a $1/(p+1)$-th circle segment $\gamma^p_{0,1}$ connecting 0 and 1, $p \in \mathbb{N}$.

If we introduce the k-term composition method

$$\Upsilon^h := \Psi^{\sigma_k h} \circ \cdots \circ \Psi^{\sigma_1 h}, \tag{4.20}$$

the process of constructing $x_k := x_\Delta\left(t_k^\Delta\right)$ from $x_0 := x_\Delta\left(t_0^\Delta\right)$ is given by

$$x_k = \Upsilon^h x_0.$$

According to the conditions of Section 4.3.3 achieving superconvergence, our $\sigma_1, \ldots, \sigma_k$ satisfy the two equations

$$\sigma_1 + \ldots + \sigma_k = 1 \quad \text{and} \quad \sigma_1^{p+1} + \ldots + \sigma_k^{p+1} = 0. \tag{4.21}$$

However, these are exactly the criteria for a classical theorem on composition methods. From [23, p. 39, Theorem 4.1], it follows that Υ^h is again a one-step method of convergence order at least $p+1$.

Remark 4.38. The order conditions (4.21) can be found in [23]. There, some specific composition methods have been introduced by explicitly solving these order conditions over the reals. [24] and [58] also stated methods, but they solved these equations over the complex numbers. In contrast to them, our starting point has not been the equations (4.21). The adaption of superconvergent paths $\gamma_{t_0,t}^p$ developed for linear ODEs solves implicitly the sufficient order conditions (4.21) and thus it provides an entire class of composition methods.

4.4.2 Iterations

The method given by (4.20) is called a k-term composition method. If the underlying discrete evolution Ψ has order $p \in \mathbb{N}$, it follows that Υ is a one-step method of order at least $p+1$. Since concatenation of RKMs is again an RKM, we can regard Υ as a new basic RKM. Hence Υ can again be used in the same manner to construct a composition method, but this time, the obtained order is at least $p+2$. Applying this process iteratively, it is possible to generate methods of arbitrary order of convergence.

We start with a basic RKM Ψ of order p. Iterating Ψ leads to methods Υ_r^h that are recursively defined by

$$\begin{aligned}\Upsilon_0^h &:= \Psi^h,\\ \Upsilon_{r+1}^h &:= \Upsilon_r^{\Delta_r},\end{aligned}$$

where $\Delta_r := \Delta_k^{\gamma_{0,h}^{p(r)}}$ and $r \geq 0$. Then Υ_r^h is a method of order $p(r)$. Hereby,

$$\begin{aligned}p(0) &:= p,\\ p(r+1) &:= p(r) + g,\end{aligned}$$

where $g \in \mathbb{N}$ is the gain of order of convergence depending on Ψ. For instance, if Ψ is the discrete evolution corresponding to an eRKM, then $g = 1$. If Ψ is, however, the discrete evolution of a symmetric method, then it follows $g = 2$.

Our geometric approach enables us now to interpret the method Υ_r with respect to a basic method Ψ. The previous recursion for Υ_r implies that $\Upsilon_r^h x_0 = \Psi^\Gamma x_0$, where Γ is a recursively defined complex time grid depending on the parameters r (depth of recursion), k (number of time steps), p (order of Ψ) and h (global step size). To get an impression of

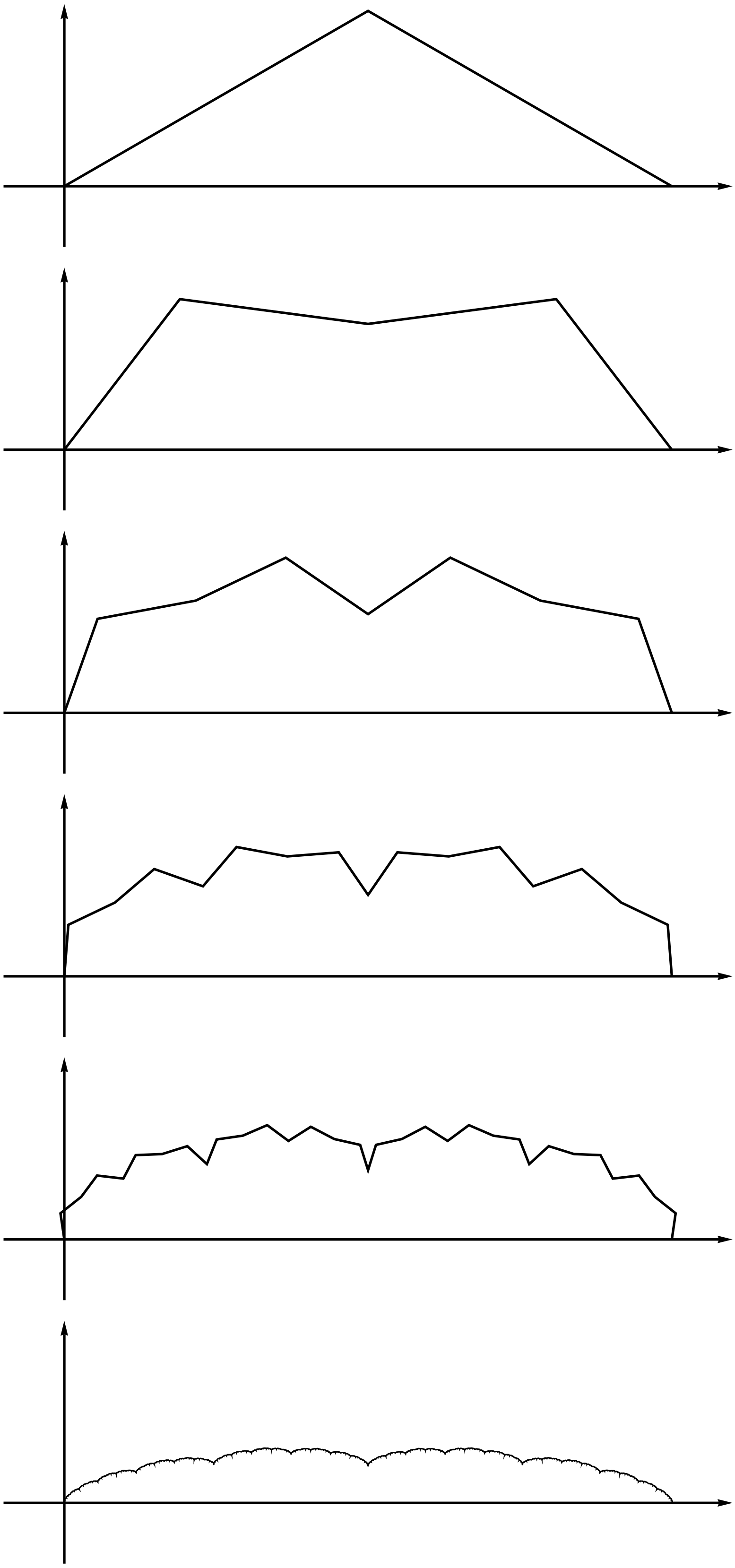

Figure 4.22: Time grid Γ for $r = 1, 2, 3, 4, 5$ and 11 ($g = h = 1$ and $p = k = 2$).

the recursive time grids, we exemplify Γ for several choices of r in Figure 4.22. As one can see, Γ becomes more and more a fractal-like structure. Furthermore, we remark that the time steps of Γ comply with the coefficients $\sigma_{r,j}$ ($j \in \{1,2\}$) of the composition methods derived in [24, p. 5, "two term composition"] ($k > 2$ analogously).

Remark 4.39. As mentioned previously, our approach leads to the same sets of coefficients, which have already been discovered in [24] and [58]. In contrast to the construction of k-term composition methods by solving the system of order conditions (4.21), we explicitly get these methods by applying a basic RKM along special complex time grids. From our geometric point of view, the iteration done in [24] and [58] to construct higher order methods based on a certain RKM transfers to an application of this RKM along the recursively defined time grid Γ.

4.4.3 A Complex Orbit

As a benchmark and illustration of the ideas that we have just introduced, a classical example of celestial mechanics, this is, the restricted three body problem, is considered. The corresponding system of differential equations

$$\begin{aligned}
\ddot{x}_1 &= x_1 + 2\dot{x}_2 - \hat{\mu}\frac{x_1+\mu}{\sqrt{\left((x_1+\mu)^2+x_2^2\right)^3}} - \mu\frac{x_1-\hat{\mu}}{\sqrt{\left((x_1-\hat{\mu})^2+x_2^2\right)^3}},\\
\ddot{x}_2 &= x_2 - 2\dot{x}_1 - \hat{\mu}\frac{x_2}{\sqrt{\left((x_1+\mu)^2+x_2^2\right)^3}} - \mu\frac{x_2}{\sqrt{\left((x_1-\hat{\mu})^2+x_2^2\right)^3}},
\end{aligned} \tag{4.22}$$

is motivated by the motion of a satellite with respect to the gravitational potential induced by the moon and the earth. Thereby, it holds $\hat{\mu} := 1 - \mu$, while $\mu := 0.012277471$ is the ratio of the mass of the moon to the mass of the total system. Furthermore, the mass of the satellite can be neglected. Since the motion stays in a plane, $\big(x_1(t), x_2(t)\big)^T$ represents the vector of the satellite's coordinates with respect to a "rotating" coordinate system whose origin represents the gravitational center of the moon and the earth and in which both celestial bodies stay on predefined points on the x_1-axis.

Due to the american mathematician R. Arenstorf[27], we know that there exists a periodic solution to the corresponding initial value problem given by the system (4.22) and a suitable chosen initial value. As illustrated by Figure 4.23, the error at the terminal point[28] by the use of the explicit Euler method (discrete evolution denoted by Ψ) is more than 55 times larger than of the associated composition method Υ_1 with $k = 2$. Hereby, both approaches have been calculated by using equidistant time grids with the same number of time steps.

[27] Compare [2]. At this point, we want to mention that Arenstorf showed the existence of such orbits by analytic continuation of parameterized two-body problems for which the exact solutions (conic sections) are explicitly known.

[28] Integration from $t_0 = 0$ to $T := 17.065216560157960$, where T is the period of the corresponding exact orbit.

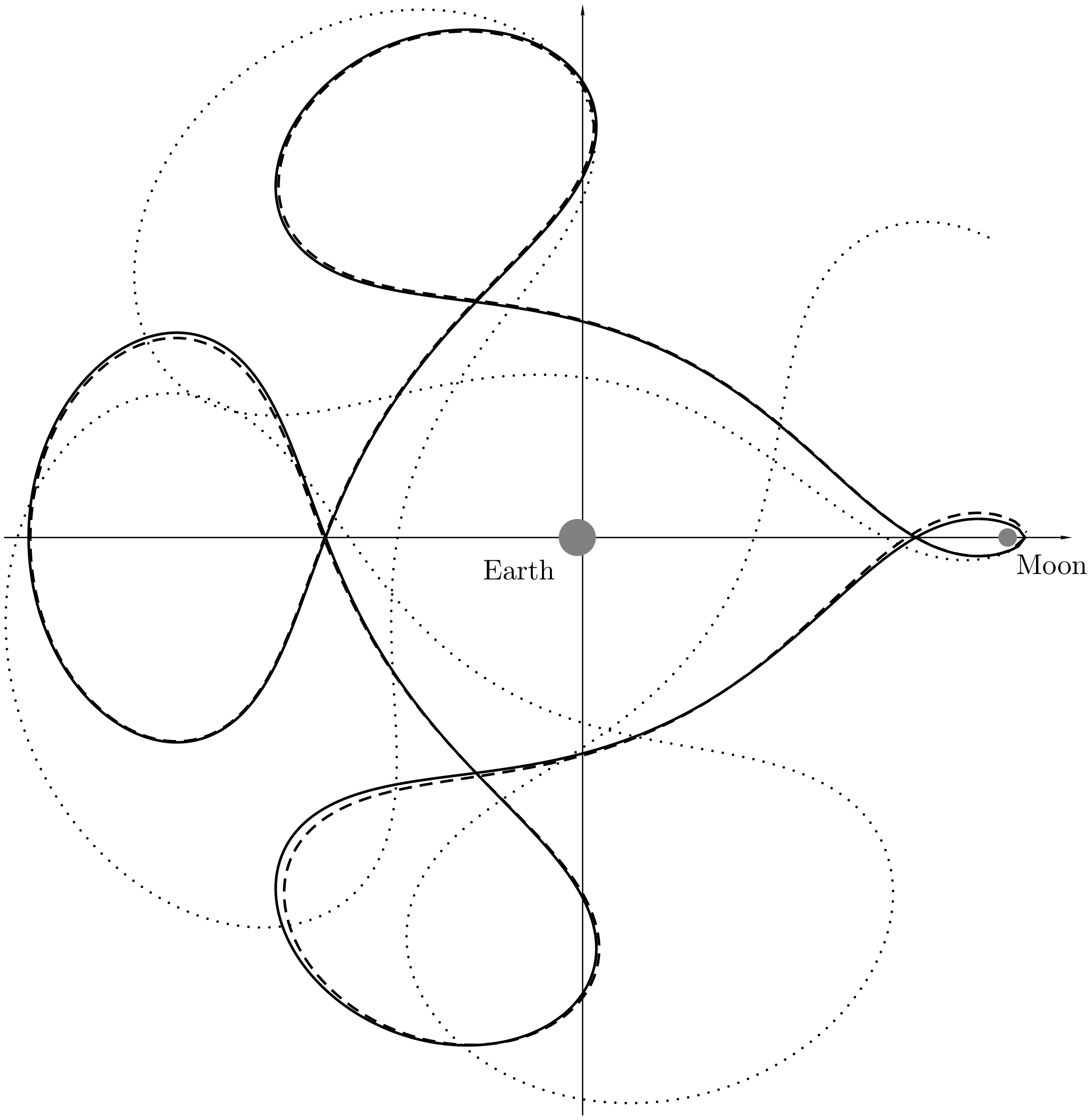

Figure 4.23: The "exact" Arenstorf orbit (corresponding initial value $\big(x_1(0), x_2(0), \dot{x}_1(0), \dot{x}_2(0)\big)^T = \big(0.9940, 0.0, 0.0, -2.001585106379080\big)^T$, period $T = 17.065216560157960$) given by 100000 real-valued equidistant steps of Dopri5 (5-th order eRKM, solid), approximation by 100000 equidistant real-valued steps of the explicit Euler method Ψ (dotted) and real part of the approximation by 50000 (2 "micro" steps for each "macro" time step, this is, $k = 2$) time steps of the complex explicit Euler method (dashed) induced by Υ_1.

5

Outlook

"Thinking outside the box."

English idiom

In the present chapter, we would like to conclude briefly the various ideas and results derived in the previous parts, point out some possible next steps to extend them and mention some further worthwhile topics we have not dealt with.

Continuity and Dynamic Geometry

After we modeled continuous dynamic geometry, we proved the existence as well as some kind of uniqueness of CDGSs corresponding to an algebraic construction and an associated given regular starting instance. By doing so, it has become apparent that continuous dynamic geometry and complex analysis are closely related. For instance, the entire argumentation relied on the fact that the movements of the free elements of an underlying construction are restrictions of appropriate holomorphic functions. In addition, we showed that only finitely many differentiable maps are not sufficient to achieve continuity in general.

Up to here, we have not studied movements that are induced by meromorphic functions. What is about them? In Lemma 2.9, we showed that even coordinates from a certain class of meromorphic functions, this is, functions without essential singularities, induce a continuous behavior in a suitable projective space. This observation gives rise to the assumption that also movements that are induced by this class could admit the construction of CDGSs.

Another interesting question concerns geometric operations. We concentrated on algebraic operations and proved that the output of a geometric operation under an h-movement must not be finite differentiable. So, there might be a larger class of such operations allowing the construction of CDGSs.

Apart from this, a topic that we have not yet mentioned, but is worthwhile to mention, is the measurement of distances and angles. On the first sight, it may seem that metric properties and projective geometry are two completely incompatible concepts. Although, appearances are often deceiving and first impressions are anything, but accurate, many mathematicians are still convinced by this deception, which is best described by a popular quote from [27, p. 138]:

> "Es [die Formel von LAGUERRE] blieb aber lange unbeachtet, vermutlich weil sich die Geometer an den Gedanken gewöhnt hatten, daß Metrik und projektive Geometrie in keiner Beziehung zueinander ständen."

The stated quote traces back to FELIX KLEIN, who is one of the persons CAYLEY-KLEIN geometries[29] are named after. These geometries are the mathematical theory that embeds among others Euclidian, hyperbolic or elliptic metric in projective geometry in a natural way.

[29] The interested reader is referred to [46].

Complexity and Dynamic Geometry

As a first step, we transferred continuous dynamic geometry, and thus also the Reachabilty Problem in Dynamic Geometry, to a setup that was based on a generalization of straight-line programs. Consequently, we did not only admit the arithmetic operations in SLPs, but also roots and logarithms. It turned out that analytic continuations along a fixed path of holomorphic functions that described the behavior of an SLP around a regular starting instance induce a continuous evaluation of this SLP along the path used for the analytic continuation. Throughout the rest of the chapter, we proved several lower complexity bounds on weakened variants of the Reachabilty Problem concerning SLPs. Thereby, we either bounded the length of the underlying path or specified its topological shape.

Due to this fact the Reachabilty Problem over the complex numbers $\mathbb{C}$ in Dynamic Geometry as well as for SLPs is still open. This is even true, if one only allows operations from the sets $QUAD$ or *quad*, respectively.

Apart from lower complexity bounds, upper bounds might be interesting as well. Although, there have been proved some of them in [12], they are still far away of existing lower bounds even over $\mathbb{R}$.

This also holds for the second decision problem that arises in dynamic geometry, this is, the so-called *Tracing Problem*. In addition, lower complexity bounds of this problem have only been established for relatively easy sets of allowed operations (see [48]).

Geometry and Numerical Integration

Applying geometric ideas to numerical integration of ordinary differential equations enabled us to increase the order of convergence of an underlying method due to a superconvergence effect, which occurred in the presence of certain complex time grids. In addition, this approach provided a geometric access to an entire class of composition methods.

Our studies in this field were originally triggered by detouring singularities in this context, while the gain of convergence appeared as a side effect. However, staying away from singularities also reduces the resolution of an adaptive step size controlled method and thus the computational effort as well. For this reason, the detection of singularities could be used to adapt the remaining path of integration in a beneficial way. On the one hand, singularities lying on the path can be detoured, while on the other hand, the path can be chosen so that the distance of it to nearby lying singularities is optimized in a certain sense. Several approaches in order to detect nearby lying singularities have already been discussed in [8, 9].

Besides the computational effort that one can gain by using complex detours, there is at least one interesting implementation-related issue as well, i.e. the problem of tracing. During a computation, one has to ensure that the used method reflects the analytic character of the function it approximates. As a result, the correct branch must be chosen in each computation step in each iteration. So it would be nice to have a software package that provides this feature automatically, in other words, a package containing primitive operations that have got a history to choose self-controlled the correct branch with respect to the previous computations. A corresponding java library has already been developed by Michael Schmid and the author.

5

Bibliography

[1] Ahlfors, L. V. *Complex Analysis*, 3 ed. International Student Edition. McGraw-Hill, Singapore, 1979.

[2] Arenstorf, R. Periodic solutions of the restricted three body problem representing analytic continuations of Keplerian elliptic motions. *American Journal of Mathematics 85* (1963), 27–35.

[3] Berlekamp, E., McEliece, R., and van Tilborg, H. On the inherent intractability of certain coding problems. *IEEE Transactions on Information Theory 24*, 3 (1978), 384–386.

[4] Blum, L., Cucker, F., Shub, M., and Smale, S. *Complexity and real computation.* Springer-Verlag, New York, Berlin, Heidelberg, 1997.

[5] Bornemann, F., Laurie, D., Wagon, S., and Waldvogel, J. *The SIAM 100-Digit Challenge: A Study in High-Accuracy Numerical Computing.* Society of Industrial Applied Mathematics (SIAM), Philadelphia, 2004.

[6] Bürgisser, P., Clausen, M., and Shokrollahi, M. *Algebraic complexity theory.* Springer-Verlag, New York, Berlin, Heidelberg, 1997.

[7] Calogero, F., Gmoez-Ullate, D., Santini, P., and Sommacal, M. The transition from regular to irregular motion explained as travel on Riemann surfaces. *Journal of Physics A: Mathematical and General 38*, 41 (1980), 8873–8896.

[8] Chang, Y., and Corliss, G. Ratio-like and recurrence relation tests for convergence of series. *IMA Journal of Applied Mathematics 25*, 4 (1980), 349–359.

[9] Corliss, G. Integrating ode's in the complex plane - pole vaulting. *Mathemtics of Computation 35*, 152 (1980), 1181–1189.

[10] Cox, D., Little, J., and O'Shea, D. *Using Algebraic Geometry.* No. 185 in Graduate Texts in Mathematics. Springer-Verlag, New York, Berlin, Heidelberg, 1998.

[11] Denner-Broser, B. On the decidability of tracing problems in dynamic geometry. In *Automated Deduction in Geometry* (2006), H. Hoon and D. Wang, Eds., no. 3763 in Springer Lecture Notes in Artificial Intelligence, Springer-Verlag, Heidelberg, pp. 111–129.

[12] Denner-Broser, B. *Tracing-Problems in Dynamic Geometry.* PhD thesis, Freie Universität Berlin, 2008.

[13] Deuflhard, P., and Bornemann, F. *Numerische Mathematik II. Gewöhnliche Differentialgleichungen*, 2 ed. de Gruyter, New York, Berlin, 2002.

[14] Filatov, D. M. On complex-stepped Runge-Kutta methods for exact time integration of linear PDEs. *HAIT Journal of Science and Engineering C 4*, 1–2 (2006), 128–135.

[15] Fischer, G. *Lehrbuch der Algebra*, 1 ed. Vieweg, Wiesbaden, 2008.

[16] Fischer, W., and Lieb, I. *Ausgewählte Kaptiel aus der Funktionentheorie.* No. 48 in Vieweg-Studium Aufbaukurs Mathematik. Vieweg, Braunschweig, Wiesbaden, 1988.

[17] Forster, O. *Riemannsche Flächen.* Springer-Verlag, Berlin, Heidelberg, 1977.

[18] Fung, T. Third order complex-time-step methods for transient analysis. *Computer methods in applied mechanics and engineering 190*, 22–23 (2001), 2789–2802.

[19] Fung, T. Solving non-linear problems by complex time step methods. *Communications in Numerical Methods in Engineering 18* (2002), 287–303.

[20] Garey, M., and Johnson, D. *Computers and Intractability: A Guide to the Theory of NP-Completeness.* Mathematical Sciences. Freeman, 1979.

[21] Gödel, K. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme. *Monatshefte für Mathematik und Physik 38* (1931), 173–198.

[22] Gray, J. *Worlds Out of Nothing – A Course in the History of the 19th Century.* Springer Undergraduate Mathematics Series. Springer-Verlag, London, 2007.

[23] Hairer, E., Lubich, C., and Wanner, G. *Geometric Numerical Integration. Structure-Prevserving Algorithms for Ordinary Differential Equations*, first ed. No. 31 in Springer Series in Computational Mathematics. Springer-Verlag, Berlin, 2002.

[24] Hansen, E., and Ostermann, A. Higher order splitting methods for analytic semigroups exists. *Submitted* (2008).

[25] Hille, E. *Ordinary Differential Equations in the Complex Domain.* Dover Pubn Inc, 1997.

[26] Jänich, K. *Topologie*, 6 ed. Springer-Verlag, New York, Berlin, Heidelberg, 1999.

[27] Klein, F. *Vorlesungen über nicht-euklidische Geometrie.* Springer-Verlag, Berlin, 1928.

[28] Klein, F. *Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert.* Springer-Verlag, Berlin, Heidelberg, New York, 1979.

[29] Kleiner, I. Principle of continuity. A brief histroy. *The Mathematical Intelligencer 28*, 4 (2006), 49–57.

[30] Kortenkamp, U. *Foundations of Dynamic Geometry.* PhD thesis, ETH Zürich, 1999.

[31] Kortenkamp, U., and Richter-Gebert, J. Decision complexity in dynamic geometry. In *Automated Deduction in Geometry* (2000), J. Richter-Gebert and D. Wang, Eds., no. 2061 in Springer Lecture Notes in Artificial Intelligence, Springer-Verlag, Heidelberg, pp. 216–220.

[32] Kortenkamp, U., and Richter-Gebert, J. Grundlagen dynamischer Geometrie. *Zeichnung - Figur - Zugfigur Mathematische und didaktische Aspekte Dynamischer Geometrie-Software* (2001), 123–145.

[33] Krick, T. Straight-line programs in polynomial equation solving. In *Foundations of Computational Mathematics* (2002), F. Cucker, R. DeVore, P. Olver, and S. A., Eds., no. 312 in London Mathematical Society Lecture Notes Series, Cambridge University Press, pp. 96–136.

[34] LAMOTKE, K. *Riemannsche Flächen.* Springer-Verlag, Berlin, Heidelberg, 2005.

[35] LINDEMANN, F. Über die Zahl π. *Mathematische Annalen 20* (1882), 213–225.

[36] MARTIN, G. *Geometric Constructions.* Undergraduated Texts in Mathematics. Springer-Verlag, New York, Berlin, Heidelberg, 1997.

[37] MATIYASEVICH, Y. *Hilbert's 10th Problem.* Foundations of Computing. The MIT Press, 1993.

[38] ORENDT, T. A complex viewpoint on the newtonian two-body problem. Diploma thesis, Technische Universität München, 2007.

[39] ORENDT, T., RICHTER-GEBERT, J., AND SCHMID, M. Geometry of numerical complex time integration, 2009. E-print arXiv:0903.1585v1.

[40] PAPADIMITRIOU, C. *Computational complexity.* Addison-Wesley Reading, 1994.

[41] PLAISTED, D. Sparse complex polynomials and polynomial reducibility. *Journal of Computer and System Science 14*, 2 (1977), 210–221.

[42] PLAISTED, D. Some polynomial and integer divisibility problems are NP-hard. *SIAM Journal on Computing 7*, 4 (1978), 458–464.

[43] PLAISTED, D. New NP-hard and NP-complete polynomial and integer divisibility problems. *Theoretical Computer Science 31* (1984), 125–138.

[44] PONCELET, J. *Traité des propriétés projectives des figures.* Gauthier-Villars, 1822.

[45] POONEN, B. Undecidability in number theory. *Notices of the American Mathematical Society 55*, 3 (2008).

[46] RICHTER-GEBERT, J. *Perspectives on Projective Geometry.* in preparation.

[47] RICHTER-GEBERT, J., AND KORTENKAMP, U. *Cinderella – the interactive geometry software.* Springer-Verlag, Berlin, Heidelberg, New York, 2000.

[48] RICHTER-GEBERT, J., AND KORTENKAMP, U. Complexity issues in dynamic geometry. In *Foundations of computational mathematics* (2000), M. Rojas and F. Cucker, Eds., World Scientific, pp. 355–404.

[49] RICHTER-GEBERT, J., AND KORTENKAMP, U. Dynamic aspects in computational geometry. In *Proceedings of the EACA 2000* (2000), A. Montes, Ed., pp. 51–61.

[50] RICHTER-GEBERT, J., AND KORTENKAMP, U. A dynamic setup for elementary geometry. In *Multimedia Tools in Comunicating Mathematics* (2000), J. Borwein, M. Morales, and K. Polthier, Eds., Springer-Verlag, Heidelberg, pp. 203–221.

[51] RICHTER-GEBERT, J., AND ORENDT, T. *Geometriekalküle.* Springer-Verlag, Berlin, Heidelberg, 2009.

[52] ROSENFELD, B. The analytic principle of continuity. *The American Mathematical Monthly 112*, 8 (2005), 743–748.

[53] SCHAEFER, T. The complexity of satisfiability problems. In *Proceedings of the tenth ACM symposium on Theory of computing* (1978), ACM, pp. 216–226.

[54] Schäfke, F., and Schmidt, D. *Gewöhnliche Differentialgleichungen - Die Grundlagen der Theorie im Reellen und Komplexen.* Springer-Verlag, Berlin, Heidelberg, New York, 1973.

[55] Schmid, M. Inside $\dot{x} = x$. Diploma thesis, Technische Universität München, 2008.

[56] Smale, S. Mathematical problems for the next century. *The Mathematical Intelligencer 20*, 2 (1998), 7–15.

[57] Vardy, A. Algorithmic complexity in coding theory and the minimum distance problem. In *Proceedings of the twenty-ninth annual ACM Symposium on Theory of computing* (1997), ACM, pp. 92–109.

[58] Vilmart, G. *Étude d'intégrateurs géométriques pour des équations différentielles.* PhD thesis, Université de Genève, 2008.

[59] von Staudt, K. *Geometrie der Lage.* Bauer & Raspe, Nürnberg, 1847.

[60] von Staudt, K. *Beiträge zur Geometrie der Lage.* Bauer & Raspe, Nürnberg, 1856.

[61] Yoshida, H. Construction of higher order symplectic integrators. *Physics Letter A 150*, 8 (1990), 262–268.