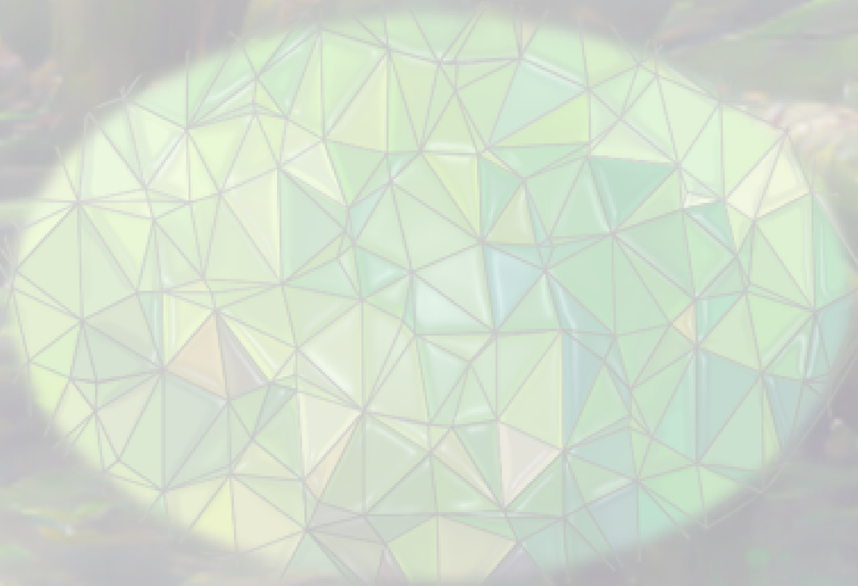
Tame Geometry and the String Landscape

Mick J. van Vliet



Utrecht 2022

Supervisors:
prof. dr. Thomas W. Grimm
prof. dr. Gil Cavalcanti

Utrecht University

Tame Geometry and the String Landscape

Mick J. van Vliet

Master Thesis

Theoretical Physics and Mathematical Sciences

Abstract

The idea of tame geometry, made precise by the notion of an o-minimal structure, has recently provided many new insights into mathematics and physics. On the mathematical side it has led to promising new results in Hodge theory, and on the physical side it has inspired a deeper understanding of the landscape of string theory vacua. In this thesis we review these exciting developments, focusing on the tameness of the period map in Hodge theory and on the proposal of a new conjectured aspect of low-energy effective theories arising from quantum gravity, called the Tameness Conjecture. We provide further evidence for the Tameness Conjecture in the setting of higher supergravity theories. In particular, we show that in such theories the geometry of the scalar field space is tame, and we argue for the tameness of gauge couplings functions. We point out several subtleties arising from the interaction between tameness and arithmetic aspects.

Acknowledgements

There are numerous people who contributed to the completion of this project to whom I wish to express my appreciation. I would like to start by sincerely thanking my supervisors Thomas Grimm and Gil Cavalcanti. I am grateful to Thomas for providing me with the opportunity to work on an incredibly exciting contemporary research topic. The novelty of this topic made this a challenging journey and it would not have been possible without his guidance and support. I am grateful to Gil for being willing to explore a topic outside of his area of research, and I greatly enjoyed our discussions about mathematical aspects of the topic and writing for physicists and mathematicians.

Next, I would like to extend my gratitude to my daily supervisor Damian van de Heisteeg for insightful discussions, helping me with various questions, and providing feedback. It was a great pleasure to collaborate on some aspects of this thesis.

Throughout the year I have benefited from many enjoyable and engaging discussions with fellow master students doing research in string theory. In particular, I would like to thank Maarten Rottier, Christoph Sieling, Thomas van Vuren and Sam Woldringh. Our almost weekly meetings were always a pleasure, and this year would not have been the same without them.

I would like to further thank the members of Thomas' research group, Brice Bastian, Arno Hoefnagels, Stefano Lanza, Chongchuo Li, Jeroen Monnee, and Lorenz Schlechter, for many valuable conversations and for providing an inspiring environment to work in.

I am grateful for the friends I made throughout my studies. In particular, I thank Lumen Eek, Frank Imbens, Lucas Smits, and Sam Woldringh, whose friendship brought great joy during the last few years. Finally, I thank my friends and family for their unconditional support.

Contents

Introduction

One of the main unifying themes of physics and mathematics is *geometry*. At the heart of geometry lies the notion of space, and there are many flavors of geometry depending on the class of spaces that one wishes to study. Given the success of geometry at describing phenomena in nature, a question that emerges is:

what is the most natural geometric framework for physics?

This is an ambitious and nebulous question, and the answer depends on which area of physics we are interested in. To take a first step in answering this question, let us go on a short tour through various types of geometry.

If we are interested in the coarsest notion of space, we enter the realm of *topology*. In this framework one studies topological spaces, and the focus lies on properties of spaces that are preserved under continuous deformations. In physics, topology is of interest whenever we are interested in the most essential properties of a space, but also when we study physical objects that are invariant under continuous deformations in a similar way as topological spaces. An example of the latter is the advent of topological insulators in condensed matter physics.

The generality of topology allows for many 'wild' phenomena to occur. As a simple but striking example, let us consider the *topologist's sine curve*. This space may be obtained as the graph of the function $x \mapsto \sin(1/x)$, defined for $x > 0$, as shown below.

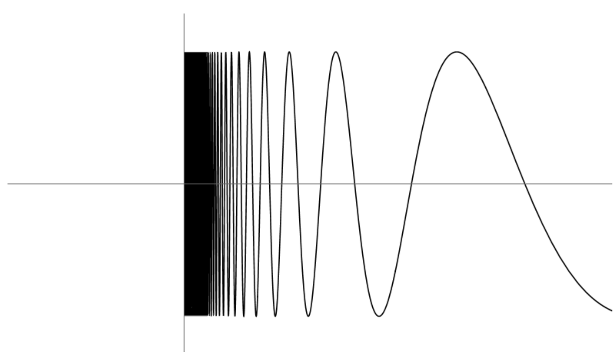


Figure 1: The topologist's sine curve.

Viewed purely as a topological space, this space is perfectly acceptable. However, from a geometric point of view, it clearly has several strange properties. For instance, near $x = 0$, the graph of this function is so wild that the notion of dimension becomes obscure. We do not often encounter such spaces in physics.

A much more well-behaved class of spaces is encountered in the field of differential geometry, where the spaces of interest are smooth. This type of geometry is prominent in physics, for instance as the framework of Einstein's theory of general relativity. If the area of physics that we wish to focus on is classical gravity, then the answer to the question posed in the first paragraph is certainly differential geometry. As a slight upgrade to differential geometry, we can consider spaces which are analytic. Upon combining this analyticity with the complex numbers, we enter complex geometry. Due to the ubiquity of complex numbers in physics, complex geometry has many physical applications.

As a final example of an important branch of geometry, we consider algebraic geometry. In its simplest form, algebraic geometry studies spaces which are defined through the vanishing of a set of polynomials. The fact that polynomials have finitely many terms causes algebraic geometry to have a certain 'finite complexity'. As a result, the structure of spaces in algebraic geometry is tame, in stark contrast with the example of the topologist's sine curve. The apparent absence of such wild spaces in physics, leads to another natural question:

could 'tameness' be a general principle in certain areas of physics?

That this could indeed be the case was recently proposed in [1], and this proposal will be one of the central topics of this thesis. For the moment we remain vague about what exactly 'tameness' means, and give a precise characterization later. The main message that we extract from the enumeration of geometries given above, is that the more we constrain the spaces under consideration, the tamer the geometry becomes. A simple way to organize these geometries is to look at the class of functions that are allowed to live on spaces. For example, in topology these are the continuous functions, and in algebraic geometry these are the polynomials. This idea is schematically depicted as follows:

	Type of Geometry	**Type of Functions**
	Topology	Continuous
tameness	Differential Geometry	Smooth
	Analytic/Complex Geometry	Analytic/Holomorphic
	Algebraic Geometry	Polynomial

Algebraic geometry lies at the tamest end of the spectrum, but as a geometric framework for physics it is too rigid. For instance, in physics we frequently encounter functions which are not polynomials, such as the real exponential. On the other hand, when passing to the more general analytic geometry, much of this tameness is lost, since analytic spaces need not have a finite geometric complexity, as present on algebraic spaces. We are then led to ask:

is there a flexible generalization of algebraic geometry which preserves its tameness?

The answer to this question is provided by a geometric framework called *tame geometry*. Originally, this type of geometry was motivated by the idea that wildly behaving objects, such as the topologist's sine curve, should somehow be banned from geometry. In its modern conception, tame geometry is formulated within the branch of mathematics called logic.

The idea of tame geometry is to constrain the allowed spaces and functions to be defined within a so-called *o-minimal structure*. This is a type of logical structure which is precisely suitable to encode the idea of tameness, since it includes the finite complexity of geometric objects as an axiom. In the context of the schematic illustration on the previous page, the type of functions allowed on these spaces are called *definable*. Remarkably, the notion of an o-minimal structure is capable of providing a framework similar to algebraic geometry, but with greatly increased flexibility. For instance, it is general enough to allow the real exponential function and compactly supported analytic functions to be included in the class of definable functions. This makes it viable for physics, as opposed to algebraic geometry. Despite its modest origins in logic, tame geometry has recently led to many exciting far-reaching developments in mathematics and physics [1–4]. Exploring these developments is the main topic of this thesis.

The area of physics in which we will explore the notion of tame geometry is centered around the theme of *quantum gravity*. The aim of such a theory is to unify all forces of nature into a single framework, for which the greatest challenge is the unification of quantum field theory with general relativity. At present, one of the strongest candidates for such a theory is *string theory*. The foundational principle of string theory is that point particles are replaced by strings. These are one-dimensional extended objects, and the vibrations of these objects are postulated to represent the fundamental degrees of freedom of nature. This basic idea leads to a beautiful theory which includes gravity and gauge theories in a natural way.

One of the most striking consequences of having strings as the fundamental objects of nature is that the dimension of spacetime must be exactly ten in order for the theory to be consistent. On the one hand, it is elegant that the dimension of spacetime is predicted, as quantum field theory and general relativity themselves can in principle be consistently formulated in any number of dimensions. On the other hand however, it is uncomfortable that this number does not agree with the number of macroscopically observed dimensions. The solution to this apparent problem is the idea of *compactification*. It proposes that the six extra dimensions are so small that they escape detection at the currently experimentally accessible energy scales. A great challenge posed by string theory is that there is an enormous range of options for the structure of the extra dimensions. Each consistent background for string theory defines a *vacuum*, and the wide range of possibilities for these backgrounds has led to the picture that there is a vast *landscape* of string vacua. Understanding the structure and size of the landscape is a central problem in string theory.

This leads to our first encounter with tame geometry in string theory. A significant part of the landscape of string theory vacua consists of a class known as *flux vacua*. These vacua are characterized by the presence of a background flux of certain fields in the theory, supported on the extra dimensions. For a fixed topology of the extra dimensions, these flux vacua can be thought of as choice of size of the extra dimensions and a choice of flux, which together minimize a certain energy potential. The set of flux vacua is assembled into a large space, and the geometry of this space has remained mysterious for many years. The most central question surrounding the space of flux vacua is whether this space is finite in any way; this was conjectured to be the case in [5, 6]. Recently, it has been shown that the geometry of the set of flux vacua is tame, in the sense that it is definable in an o-minimal structure [4]. This confirms the long-standing conjecture that there are finitely many flux vacua. The search for our universe among the huge number of flux vacua may still be a search for a needle in a haystack, but this new result shows that the haystack is not infinitely big.

Upon considering low energy scales, each vacuum of string theory gives rise to an *effective field theory* describing the remaining degrees of freedom. Closely related to the landscape of string vacua, there is thus a landscape of effective field theories that can arise from compactifications of string theory. More generally, in this landscape one includes all the consistent effective field theories that can arise as a low energy description of a theory of quantum gravity. This landscape is huge, reflecting the huge number of string vacua, but it does not contain every possible effective field theory. In fact, the effective field theories that do not arise as consistent low-energy descriptions of quantum gravity constitute a much larger set known as the *swampland* [7–9]. To uncover general low-energy imprints of quantum gravity, it is therefore of great importance to understand what distinguishes a theory in the landscape from one in the swampland. This is the task of the *swampland program.* The idea of this program is to formulate criteria, known as swampland conjectures, which are conjectured to hold for any effective field theory that can be consistently completed to quantum gravity.

In the context of the swampland program, we find our second encounter with tame geometry in physics. For all effective field theories arising from flux vacua, the geometry of the theory appears to be tame, in the sense that all spaces and functions which appear in the formulation of the effective field theory are definable in an o-minimal structure. This has motivated the proposal of a new swampland conjecture, called the *Tameness Conjecture* [1]. The precise statement of the conjecture is that all field spaces, coupling functions, and parameter spaces appearing in an effective field theory consistent with quantum gravity should be definable in an o-minimal structure. To partially answer the question stated in the opening paragraph: in the context of effective field theories arising from quantum gravity, the natural geometric framework appears to be tame geometry, as pictorially illustrated in the figure below. Explaining the details of this answer is the main aim of this thesis.

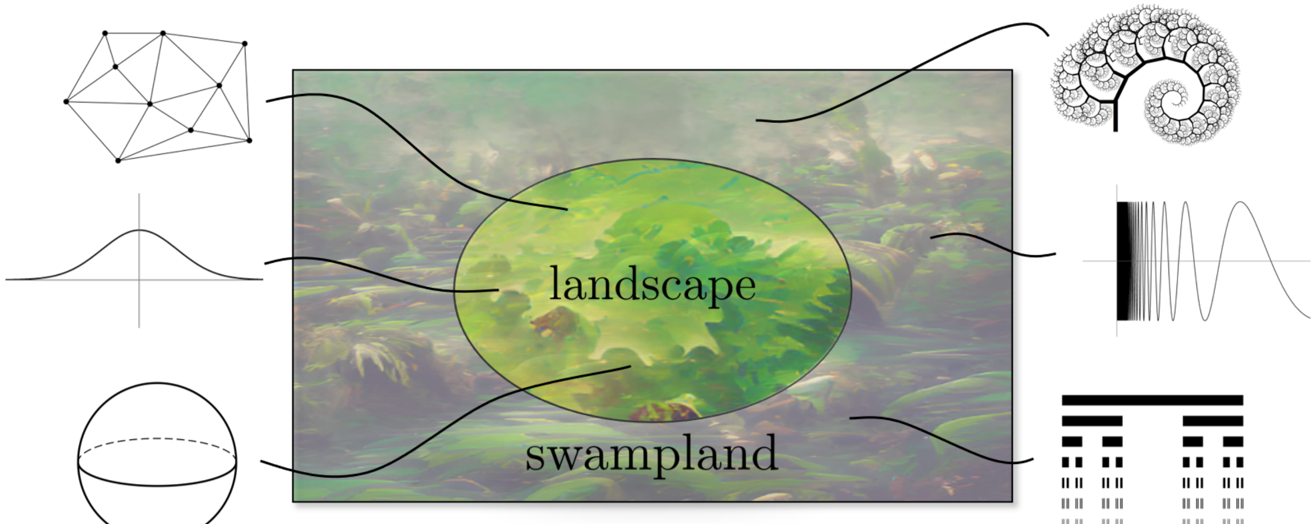


Figure 2: Pictorial interpretation of the landscape and the swampland of effective field theories. According to the Tameness Conjecture, one only finds tame spaces and functions in the landscape, whereas wildly behaving objects like the topologist's sine curve are banished to the swampland.

Most of the recent developments in applying tame geometry in physics originate from a branch of mathematics called *Hodge theory*. In a certain sense, the objective of Hodge theory is to study algebraic spaces by means of analytic methods. Regarding tame geometry as a type of geometry lying in between algebraic and analytic, Hodge theory therefore seems like a natural setting for tame geometry to appear. That this is indeed the case has recently been proven in [2], where it was shown that one of the most important objects in Hodge theory, the *period map*, is definable in an o-minimal structure. Understanding the definability of the period map is therefore an important aspect of this thesis.

This takes us to the outline of this thesis. The topic of this thesis rests on three pillars, namely Hodge theory, string theory, and tame geometry. Each of these pillars deserves its own comprehensive chapter, and these three chapters constitute the first part of the thesis. In Chapter 1 we therefore start with Hodge theory, and the main goal of the chapter is to get acquainted with the notion of a period map. Along the way we will learn about *variations of Hodge structures*, which provides a useful framework for studying compactifications of string theory. We then move on to Chapter 2 on string theory. Though we start from a general perspective, the emphasis of the chapter lies on the landscape of string vacua that arises from flux compactifications of string theory, as well as the swampland of effective field theories that are inconsistent with quantum gravity. With the language of Hodge theory at our disposal, we can formulate certain aspects of string compactifications in a mathematical way, which will be useful for later chapters. In Chapter 3, we then finally introduce tame geometry in the form of o-minimal structures. Here the main purpose is to become familiar with the basics of o-minimal structures, and to learn how the resulting framework of geometry is tame. We also discuss connections with tame geometry and other types of geometry, such as algebraic and differential geometry. This third chapter is independent of the first two, and the reader who wishes to learn about tame geometry right away may start there.

The second part of the thesis consists of exploring connections between the three pillars of the first part. In Chapter 4 we discuss the recent applications of tame geometry to Hodge theory, of which the most important application is the definability of the period map. This requires us to also understand the tame geometry of the target space of the period map, which is a type of space known as an *arithmetic quotient*. The chapter concludes with a brief discussion of how tame geometry has provided new insights into the Hodge conjecture. Building further on the connections between tame geometry and Hodge theory, Chapter 5 is centered around the tameness of the landscape of flux vacua and the proposal of the Tameness Conjecture. Finally, in Chapter 6 we provide new evidence for this conjecture in the setting of *higher supergravity* theories. In such theories, the arithmetic quotients examined in Chapter 4 are encountered as target spaces of the scalar fields, confirming the first part of the Tameness Conjecture. In this discussion, we find potential connections to other swampland conjectures. Subsequently, we discuss the definability of the coupling functions for the scalar and gauge field sectors, which validates the second part of the conjecture in this setting. The chapter ends with a discussion of the parameter space of higher supergravity theories.

We conclude with a summary and provide an outlook on directions for future research. The surge of recent developments inspired by tame geometry has been nothing short of exciting, and in this thesis we hope to convey some of that excitement.

Chapter 1

Hodge Theory

One of the beautiful aspects of complex algebraic geometry is that powerful analytic methods are available to study algebraic objects. Hodge theory, a prominent example of this idea, is based on the fact that the cohomology of complex algebraic manifolds has an additional layer of structure which is not visible to the underlying topology alone. Studying this structure has become profoundly important in geometry and has led to many insights and ideas, culminating in the famous *Hodge Conjecture*.

The aim of this chapter is to give an overview of Hodge theory, and the story is conceptually divided into the following three aspects:

(i) the cohomology groups of compact Kähler manifolds carry a *Hodge structure*;

(ii) in families of compact Kähler manifolds, this structure varies and gives rise to a *variation of Hodge structure* on the parameter space of the family;

(iii) near regions in the parameter space where the compact Kähler manifold degenerates, the variation of Hodge structures shows interesting phenomena captured by *asymptotic Hodge theory*.

The first aspect is the topic of Sections 1.1 and 1.2. We begin by recalling a number of foundational results of Hodge theory, and then proceed by studying the notion of a Hodge structure. The material discussed is mostly based on [10–12]. In Section 1.3 and 1.4 we then discuss point (ii), and the most important result will be that a variation of Hodge structure is encoded in a *period map* taking values in the classifying space of Hodge structures [13, 14]. The third point is the topic of 1.5, and the central topic will be the asymptotic behavior of the period map, captured by the orbit theorems of Schmid [15]. Finally, we comment on the Hodge conjecture in Section 1.6.

Throughout the chapter, we will frequently visit the torus as an example. We assume familiarity with some foundational concepts of complex geometry and algebraic geometry, for which excellent introductory texts are [11, 16, 17].

1.1 The Hodge decomposition

1.1.1 Complex Manifolds and Kähler Manifolds

The complex structure that is present on a complex manifold X has many intriguing linear-algebraic consequences. The most basic one is the fact that the complex tangent bundle $TX \otimes \mathbb{C}$ decomposes into a holomorphic and anti-holomorphic part as

$$TX \otimes \mathbb{C} = T^{1,0}X \oplus T^{0,1}X, \quad \text{with} \quad \overline{T^{1,0}X} = T^{0,1}X.$$

Here the notion of (anti-)holomorphic can be characterized by the $\pm i$-eigenspaces of J, where J is the underlying almost complex structure of X. This basic decomposition induces a similar decomposition on bundles that are constructed from TX in a linear-algebraic way. In particular, for each k the bundle of complex differential forms $\Omega^k(X)$ splits as

$$\Omega^k(X) = \bigoplus_{p+q=k} \Omega^{p,q}(X).$$

Here the $\Omega^{p,q}(X)$ are formed by taking anti-symmetric products

$$\Omega^{p,q}(X) = \bigwedge^p \Omega^{1,0}(X) \otimes \bigwedge^q \Omega^{0,1}(X)$$

of the components $\Omega^{1,0}(X)$ and $\Omega^{0,1}(X)$ of the cotangent bundle. Sections of $\Omega^{p,q}(X)$ are called (p,q)-forms, and locally they take the form

$$f_{i_1,\dots,i_p,\bar{\jmath}_1,\dots,\bar{\jmath}_q} dz^{i_1} \wedge \cdots \wedge dz^{i_p} \wedge d\bar{z}^{\bar{\jmath}_1} \wedge \cdots \wedge d\bar{z}^{\bar{\jmath}_q}.$$

The Hodge decomposition theorem now tells us that for some spaces, this decomposition persists on the level of cohomology.

Theorem 1.1.1. *Let X be a compact Kähler manifold. Then for each k, the k-th complex cohomology group decomposes as*

$$H^k(X;\mathbb{C}) = \bigoplus_{p+q=k} H^{p,q}(X), \tag{1.1}$$

with $H^{p,q}(X)$ consisting of cohomology classes represented by forms of degree (p,q). Moreover, $\overline{H^{p,q}}(X) = H^{q,p}(X)$.

It is not within the scope of this thesis to prove this theorem. However, we will provide a rough outline of the proof following [10], as it illustrates some of the main ideas of classical Hodge theory and at the same time allows us to review some geometric concepts that will be important later.

1.1.2 Harmonic Forms and Laplacians

In the following we assume that X is a compact Kähler manifold. We denote the space of sections of $\Omega^k(X)$, i.e. the differential k-forms on X, by $A^k(X)$. The metric on X allows us to put an L^2-inner product on each of the $A^k(X)$, defined as

$$(u,v)_{L^2} = \int_X \langle u, v\rangle_x \,\text{vol}. \tag{1.2}$$

Here $\langle u, v\rangle_x$ is constructed using the Kähler metric and the evaluation of differential forms at a point. With respect to this inner product, we can consider the formal adjoint d^* of the de Rham differential $d : A^k(X) \to A^{k+1}(X)$, which is characterized by

$$(du, v)_{L^2} = (u, d^* v)_{L^2} \quad \text{for all } u \in A^k(X),\, v \in A^{k+1}(X).$$

Using techniques from the functional analysis of differential operators on manifolds[1], it is possible decompose the space of k-forms as

$$A^k(X) = \operatorname{im} d \oplus (\operatorname{im} d)^{\perp} = \operatorname{im} d \oplus \ker d^*,$$
$$A^k(X) = \ker d \oplus (\ker d)^{\perp} = \ker d \oplus \operatorname{im} d^*.$$

Combining these two decompositions and using the fact that $\operatorname{im} d \subseteq \ker d$, we arrive at

$$A^k(X) = \operatorname{im} d \oplus \operatorname{im} d^* \oplus \ker d \cap \ker d^*.$$

In this splitting, we identify the special subspace $\mathcal{H}^k(X) := \ker d \cap \ker d^*$. It is a complementary subspace of $\operatorname{im} d$ inside $\ker d$, which quite remarkably means that it is indentified with the cohomology group $H^k(X; \mathbb{C}) = \ker d / \operatorname{im} d$ under the projection $\ker d \to \ker d/\operatorname{im} d$. Forms in $\mathcal{H}^k$ are called *harmonic forms*, and we have now established that cohomology classes on X may be identified with harmonic forms. More precisely, we have

Theorem 1.1.2. *Let X be a compact Kähler manifold. Then the map*

$$\begin{aligned} \mathcal{H}^k(X) &\to H^k(X; \mathbb{C}) \\ u &\mapsto [u] \end{aligned} \tag{1.3}$$

is an isomorphism. Moreover, the harmonic representative of $[u]$ is the unique form u that minimizes the norm $\|u\|^2_{L^2} = (u, u)_{L^2}$.

In fact, it is worth noting that this result holds for compact oriented Riemannian manifolds, and in this more general setting it is known as the Hodge theorem [10].

Proof. The first part of the theorem follows from the discussion above. For the second part, let $[u] \in H^k(X; \mathbb{C})$. Let u be the harmonic representative of $[u]$, i.e. $du = 0$ and $d^* u = 0$. Any other representative of $[u]$ can be written as $u + dv$, and we have

$$\begin{aligned} \|u + dv\|^2_{L^2} &= \|u\|^2_{L^2} + \|dv\|^2_{L^2} + (u, dv)_{L^2} + (dv, u)_{L^2} \\ &= \|u\|^2_{L^2} + \|dv\|^2_{L^2} + (d^* u, dv)_{L^2} + (v, d^* u)_{L^2} \\ &\geq \|u\|^2_{L^2}, \end{aligned}$$

which completes the proof. □

An important alternative characterization of harmonic forms can be given in terms of the *Laplacian*, which is a differential operator acting on forms, defined by

$$\Delta = dd^* + d^* d. \tag{1.4}$$

[1] The techniques required concern elliptic differential operators and their properties.

From this we see that a form u is harmonic if and only if $\Delta u = 0$. Later we will see harmonic forms make an appearance when we identify the massless modes in the string spectrum.

On a complex manifold X, the de Rham differential d decomposes as $d = \partial + \bar{\partial}$, where the operators ∂ and $\bar{\partial}$ characterized by sending a form of bidegree (p, q) to a form of bidegree $(p+1, q)$ and $(p, q+1)$, respectively. One can now again use the L^2-inner product defined above to construct formal adjoint operators ∂^* and $\bar{\partial}^*$. From these we define new Laplacian operators

$$\Delta_{\partial} = \partial\partial^* + \partial^*\partial, \quad \Delta_{\bar{\partial}} = \bar{\partial}\bar{\partial}^* + \bar{\partial}^*\bar{\partial}.$$

An essential result in Kähler geometry relates these Laplacians to the original Laplacian.

Proposition 1.1.3. *Let X be a compact Kähler manifold. Then*

$$\Delta = 2\Delta_{\partial} = 2\Delta_{\bar{\partial}}. \tag{1.5}$$

Assuming this result and the Hodge theorem, it is possible to prove Theorem 1.1.1. The key observation is that Δ_{∂} and $\Delta_{\bar{\partial}}$ preserve the type (p, q) of pure (p, q)-forms, and hence by Proposition 1.1.3 so does the Laplacian Δ. This implies that the (p, q)-components of a harmonic form are harmonic as well, and hence the space of harmonic forms decomposes as

$$\mathcal{H}^k(X) = \bigoplus_{p+q=k} \mathcal{H}^{p,q}(X),$$

where $\mathcal{H}^{p,q}(X)$ are the harmonic forms of type (p, q). By making the identification $H^{p,q}(X) \cong \mathcal{H}^{p,q}(X)$ using Theorem 1.1.2, the Hodge decomposition is obtained.

Before we conclude this section about classical Hodge theory, we review two more concepts that appear later, namely *primitive cohomology* and *polarizations*. Let X be a Kähler manifold with Kähler form ω. Taking the wedge product with the Kähler class $[\omega]$ induces an operator L on cohomology called the Lefschetz operator. Explicitly, it is defined as

$$\begin{aligned} L : H^k(X; \mathbb{C}) &\to H^{k+2}(X; \mathbb{C}), \\ [u] &\mapsto [\omega \wedge u]. \end{aligned} \tag{1.6}$$

If n is the complex dimension of X, then the kth primitive cohomology group is defined by

$$H^k_{\mathrm{p}}(X; \mathbb{C}) := \ker(L^{n-k+1} : H^k(X; \mathbb{C}) \to H^{2n-k+2}(X; \mathbb{C})). \tag{1.7}$$

Here we defined primitive cohomology with complex coefficients, but it can be defined over different coefficient rings depending on the Kähler class $[\omega]$. For instance, if the Kähler class is integral, i.e. $[\omega] \in H^k(X; \mathbb{Z})$, the Lefschetz operator L can be defined on integral cohomology and we set

$$H^k_{\mathrm{p}}(X; \mathbb{Z}) := \ker(L^{n-k+1} : H^k(X; \mathbb{Z}) \to H^{2n-k+2}(X; \mathbb{Z})). \tag{1.8}$$

The complex primitive cohomology groups, being a subspaces of the complex cohomology groups, have a natural Hodge decomposition

$$H^k_{\mathrm{p}}(X; \mathbb{C}) = \bigoplus_{p+q=k} H^{p,q}_{\mathrm{p}}(X), \tag{1.9}$$

where $H^{p,q}_{\mathrm{p}}(X) = H^k_{\mathrm{p}}(X;\mathbb{C}) \cap H^{p,q}(X)$. This decomposition has a special property that the original Hodge decomposition does not have. For a given $0 \leq k \leq 2n$, consider the bilinear form defined on $H^k(X;\mathbb{C})$ by

$$q_k([u],[v]) = (-1)^{k(k-1)/2} \int_X \omega^{n-k} \wedge u \wedge v. \tag{1.10}$$

Because of how the wedge product interacts with degrees of forms, q_k is symmetric if k is even and anti-symmetric if k is odd, and we say that q_k is $(-1)^k$-*symmetric*. Using the same techniques that are required in proving the Hodge decomposition theorem, it is possible to prove the following result for the primitive cohomology.

Theorem 1.1.4 (Hodge-Riemann Bilinear Relations)**.** *The bilinear form q_k satisfies the following properties.*

(i) The Hodge decomposition $H^k_{\mathrm{p}}(X;\mathbb{C}) = \bigoplus_{p+q=k} H^{p,q}_{\mathrm{p}}(X)$ is orthogonal with respect to the form $q_k(-,\overline{-})$, i.e.

$$q_k([u],\overline{[v]}) = 0 \quad \textit{for } [u] \in H^{p,q}(X)_{\mathrm{p}},\ [v] \in H^{p',q'}_{\mathrm{p}} \textit{ unless } p = p',\ q = q';$$

(ii) The bilinear form $i^{p-q} q_k(-,\overline{-})$ is positive definite on the subspace $H^{p,q}_{\mathrm{p}}(X) \subseteq H^k_{\mathrm{p}}(X;\mathbb{C})$.

For our purposes, the second bilinear relation is what makes primitive cohomology relevant. With this property, $H^k_{\mathrm{p}}(X;\mathbb{C})$ is said to be *polarized*[2] by q_k.

The 2-torus $\mathbb{T}$ is one of the simplest examples of a compact Kähler manifolds, and it is able to illustrate many of the definitions and results that we have encountered.

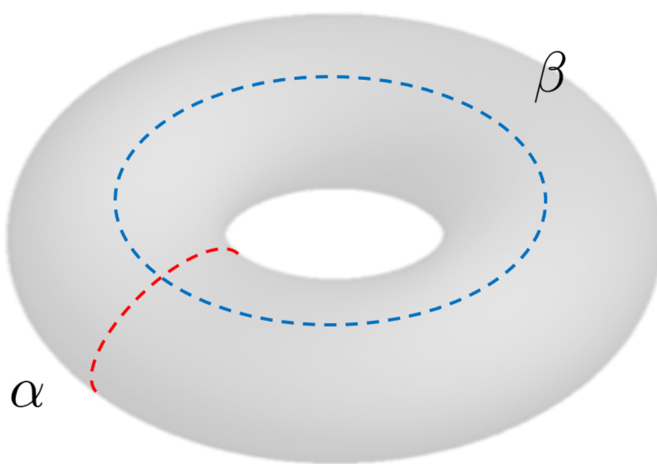


Figure 1.1: The torus $\mathbb{T}$, with two homology cycles α and β forming a basis of $H_1(\mathbb{T};\mathbb{Z})$. The basis is canonical in the sense that the intersection pairing of α and β is equal to 1.

[2]A precise definition of polarization is given in the next section.

Example 1.1.5. The degree 1 cohomology of the torus is given by $H^1(\mathbb{T};\mathbb{C}) \cong \mathbb{C}^2$. As a basis, we choose two cohomology classes α^*, β^* which are Poincaré dual to a canonical homology basis $\alpha, \beta \in H_1(\mathbb{T};\mathbb{Z})$, as illustrated in Figure 1.1. According to the Hodge decomposition theorem, the cohomology splits as

$$H^1(\mathbb{T};\mathbb{C}) = H^{1,0} \oplus H^{0,1}, \quad \text{with} \quad \overline{H^{1,0}} = H^{0,1}.$$

The space $H^{1,0}$ is spanned by the cohomology class of the (up to scaling) unique holomorphic form $\Omega \in \Gamma(\Omega^{1,0}(X))$, and in terms of the basis introduced above we can express

$$[\Omega] = A\alpha^* + B\beta^*,$$

where the complex coefficients A and B are given by the *periods* of Ω,

$$A = \int_\alpha \Omega, \quad B = \int_\beta \Omega.$$

The Lefschetz operator vanishes for degree reasons, so we have $H^1_{\mathrm{p}}(X;\mathbb{C}) = H^1(X;\mathbb{C})$, and the polarization form is simply given by integration. The Hodge-Riemann bilinear relations are then captured by the identities

$$\int_\mathbb{T} \Omega \wedge \Omega = 0, \quad \int_\mathbb{T} \overline{\Omega} \wedge \overline{\Omega} = 0, \quad i\int_\mathbb{T} \Omega \wedge \overline{\Omega} = 2\,\mathrm{Im}(B\overline{A}) > 0.$$

1.2 Hodge Structures

We now turn to a more conceptual point of view, and instead of working directly with spaces whose cohomology admits a Hodge decomposition, we study an abstracted version of this concept. This leads us to the definition of a Hodge structure.

Definition 1.2.1. An (*integral*) *weight k Hodge structure* consists of a free abelian group of finite rank $H_\mathbb{Z}$ whose complexification $H_\mathbb{C} = H_\mathbb{Z} \otimes \mathbb{C}$ carries a Hodge decomposition

$$H_\mathbb{C} = \bigoplus_{p+q=k} H^{p,q} \tag{1.11}$$

with $\overline{H^{p,q}} = H^{q,p}$. The dimensions $h^{p,q} = \dim H^{p,q}$ of a Hodge structure are called the *Hodge numbers*.

The key example of a Hodge structure to keep in mind is of course the one that is present on the cohomology of compact Kähler manifolds as discussed in the previous section. The data of a Hodge structure can be equivalently characterized by a *Hodge filtration*. This is a finite decreasing filtration $F^\bullet$ given by

$$F^p = \bigoplus_{r \geq p} H^{r,k-r}. \tag{1.12}$$

The components of the Hodge decomposition are recovered by $H^{p,q} = F^p \cap \overline{F^q} = F^p \cap \overline{F^{k-p}}$, as illustrated in the diagram

$$\underbrace{H^{k,0} \oplus H^{k-1,1} \oplus \cdots \oplus H^{p+1,q-1} \oplus \overbrace{H^{p,q}}^{}}_{F^p}\!\!\overbrace{\oplus H^{p-1,q+1} \oplus \cdots \oplus H^{1,k-1} \oplus H^{0,k}}^{\overline{F^q}}.$$

From this diagram we also infer that the Hodge symmetry condition $\overline{H^{p,q}} = H^{q,p}$ takes the form

$$H_{\mathbb{C}} = F^p \oplus \overline{F^{k-p+1}}$$

in terms of the filtration $F^\bullet$. The point of view of Hodge filtrations will be important when we consider the classifying space of Hodge structures. If we want to emphasize the underlying Hodge structure, we also write $F^\bullet H_{\mathbb{C}}$.

Definition 1.2.2. A weight k Hodge structure on $H_{\mathbb{C}}$ is *polarized* if $H_{\mathbb{Z}}$ is equipped with a $(-1)^k$-symmetric quadratic form $q_{\mathbb{Z}}$, such that the Hermitian form h defined by

$$h(u,v) = q_{\mathbb{C}}(u, C\overline{v}), \quad \text{with} \quad C = \bigoplus_{p+q=k} i^{p-q}\mathrm{id}_{H^{p,q}} \tag{1.13}$$

makes the Hodge decomposition orthogonal with respect to h, and h is positive definite.

The bilinear form h appearing in this definition is called the *Hodge form*, and the two conditions that we impose on h are precisely the bilinear relations that we encountered in Theorem 1.1.4. The operator C is known as the Weil operator, and it will play a crucial role later. Examples of polarized integral Hodge structures are given by the primitive cohomology groups $H^k_{\mathrm{p}}(X;\mathbb{C})$ of compact Kähler manifolds for which the Kähler class $[\omega]$ is integral. This integrality is needed to have integral primitive cohomology $H^k_{\mathrm{p}}(X;\mathbb{Z})$ on which the polarization form q_k is defined, as mentioned in the discussion in the previous section. We mention the following result because it gives an abundance of polarized Hodge structures [12].

Theorem 1.2.3. *Let X be a smooth complex projective variety. Then X is a compact Kähler manifold with integral Kähler class*[4]*.*

Proof. The assumptions imply that X is a complex submanifold of $\mathbb{CP}^m$ for some $m > 0$. The complex projective space $\mathbb{CP}^m$ is a Kähler manifold with integral Kähler class given by the Fubini-Study form ω^{FS}, and the restriction of ω^{FS} to X defines an integral Kähler class on X. □

[3] It is interesting to note that the converse is also true. If X is a compact Kähler manifold whose Kähler class is integral, then X is a smooth complex projective variety. This is the Kodaira Embedding Theorem.

1.3 The Classifying Space of Hodge Structures

Now that we have set up the definition of a Hodge structure, it is natural to ask whether there is a nice way to classify all the different Hodge structures that we can put on a fixed vector space. The answer is that this can indeed be done for polarized Hodge structures, and in fact, that all the polarized Hodge structures with fixed Hodge numbers that a vector space can have can be assembled into a classifying space by a construction of Griffiths [14]. Traditionally, such a classifying space is called a *period domain* and denoted by $\mathcal{D}$, and this section will be devoted to constructing these spaces and understanding their geometry. This understanding is of significant importance for the theory of *variations of Hodge structure*, and consequently period domains are among the main objects in this thesis. In the construction below we follow Voisin and Schmid [10, 15].

Before moving on to the general construction of period domains, we look at a simple example.

Example 1.3.1. For the torus, the Hodge structure on $H^1(\mathbb{T};\mathbb{C})$ is fully determined by the one-dimensional subspace $H^{1,0}$, since $H^{0,1} = \overline{H^{1,0}}$. This subspace is spanned by the cohomology class of the holomorphic form Ω. For a fixed homology basis $\alpha, \beta \in H_1(\mathbb{T};\mathbb{Z})$, we have seen that the class $[\Omega]$ is determined by its periods $A = \int_\alpha \Omega$ and $B = \int_\beta \Omega$, which are both non-zero as a result of the bilinear relations. Since the span $H^{1,0}$ is independent of rescalings of Ω, it only depends on the period ratio $\tau = B/A$. The condition

$$\operatorname{Im}(B\overline{A}) > 0$$

that we found earlier now implies that τ has positive imaginary part. Therefore, a Hodge structure on the middle cohomology of the torus is completely determined by a point τ in the upper half-plane $\mathbb{H}$. In other words, $\mathcal{D} = \mathbb{H}$ is the period domain for weight 1 Hodge structures on a two-dimensional vector space.

The precise data that we need to construct a general period domain is:

(i) a vector space $H_\mathbb{C}$ and a finitely generated abelian group $H_\mathbb{Z}$ with $H_\mathbb{C} = H_\mathbb{Z} \otimes \mathbb{C}$;

(ii) a non-degenerate $(-1)^k$-symmetric form $q_\mathbb{Z}$ on $H_\mathbb{Z}$;

(iii) a weight $k \in \mathbb{Z}$ and a set of prescribed Hodge numbers $\{h^{p,q}\}_{p+q=k}$ with $h^{p,q} = h^{q,p}$.

The starting point for the construction of the period domain $\mathcal{D}$ is the observation that a weight k polarized Hodge structure is fully determined by a filtration $F^k \subseteq F^{k-1} \subseteq \cdots \subseteq F^0$ satisfying the Hodge symmetry condition

$$H_\mathbb{C} = F^p \oplus \overline{F^{k-p+1}}, \tag{1.14}$$

with $H^{p,q} = F^p \cap \overline{F^q}$ for all $p+q = k$, on which the form $h(-,-) := q_\mathbb{C}(-, C\overline{-})$ satisfies the two bilinear relations. A filtration that has certain fixed Hodge numbers is said to be of *fixed type*. Equivalently, a filtration is of fixed type if for each p,

$$\dim F^p = f^p := \sum_{r \geq p} h^{r,k-r}. \tag{1.15}$$

The set of filtrations $F^\bullet$ of fixed type is a flag variety and can naturally be viewed as a closed subset of a product of Grassmannians[4]

$$\check{\mathcal{F}} \subseteq \mathrm{Gr}_{\mathbb{C}}(f_0, H_{\mathbb{C}}) \times \cdots \times \mathrm{Gr}_{\mathbb{C}}(f_k, H_{\mathbb{C}})$$

In fact, it can straightforwardly be shown that the equations defining $\check{\mathcal{F}}$ are algebraic so that $\check{\mathcal{F}}$ is a complex projective variety. By a symmetry argument[5] it follows that $\check{\mathcal{F}}$ is smooth. The filtrations in $\check{\mathcal{F}}$ are not yet valid Hodge decompositions, because we did not yet enforce the condition of equation (1.14). The filtrations satisfying this condition define an open subset $\mathcal{F} \subseteq \check{\mathcal{F}}$, and this subset parametrizes the (unpolarized) Hodge structures on $H_{\mathbb{C}}$ whose Hodge numbers are given by $\{h^{p,q}\}$.

In order to construct the period domain, we now need to impose the bilinear relations. Recall that the first bilinear relation is the orthogonality condition

$$h(H^{p,q}, H^{p',q'}) = 0, \quad \text{unless} \quad p = p',\, q = q'.$$

In terms of the Hodge filtration, this is equivalent to $h(F^p, \overline{F^{p-k+1}}) = 0$ for all p. This is an algebraic condition on points in $\check{\mathcal{F}}$, and hence defines a closed complex subvariety $\check{\mathcal{D}} \subseteq \check{\mathcal{F}}$. By the same argument as above, $\check{\mathcal{D}}$ is in fact smooth. The second bilinear relation is the positive definiteness requirement

$$h(v, v) > 0 \quad \text{for all} \quad v \in H_{\mathbb{C}},\, v \neq 0. \tag{1.16}$$

The set of filtrations for which this holds, in addition to the other constraints discussed above, finally defines the period domain $\mathcal{D} \subseteq \check{\mathcal{D}}$. It is an open subset of $\check{\mathcal{D}}$, which essentially follows from the fact that it is defined through the inequality (1.16). Since $\check{\mathcal{D}}$ is smooth, we conclude that the period domain is an open complex submanifold of $\check{\mathcal{D}}$.

We have now constructed an abstract complex manifold $\mathcal{D}$ that classifies the polarized weight k Hodge structures of fixed type. Owing to the symmetry of the construction, we can actually construct a more geometric model for the period domain. Consider the linear algebraic group

$$G_{\mathbb{C}} := \mathrm{Aut}(H_{\mathbb{C}}, q_{\mathbb{C}}) = \{g \in \mathrm{GL}(H_{\mathbb{C}}) \,|\, q_{\mathbb{C}}(gu, gv) = q_{\mathbb{C}}(u, v) \ \text{for all} \ u, v \in H_{\mathbb{C}}\}. \tag{1.17}$$

This group acts on the Grassmann manifold $\mathrm{Gr}_{\mathbb{C}}(f_0, H_{\mathbb{C}}) \times \cdots \times \mathrm{Gr}_{\mathbb{C}}(f_k, H_{\mathbb{C}})$ by rotating subspaces of $H_{\mathbb{C}}$ according to the action of an element $g \in \mathrm{GL}(H_{\mathbb{C}})$. More explicitly, g acts on an element $(F^0, \ldots, F^k) \in \mathrm{Gr}_{\mathbb{C}}(f_0, H_{\mathbb{C}}) \times \cdots \times \mathrm{Gr}_{\mathbb{C}}(f_k, H_{\mathbb{C}})$ as

$$g \cdot (F^0, \ldots, F^k) = (gF^0, \ldots, gF^k) =: gF^\bullet. \tag{1.18}$$

Since $G_{\mathbb{C}}$ preserves the form $q_{\mathbb{C}}$, it also preserves the first bilinear relation, and hence the group action restricts in a well-defined way to $\check{\mathcal{D}}$. The following elementary result was first shown in [14].

---

[4]Recall that a Grassmannian is a manifold parametrizing subspaces of fixed dimension in a given vector space. The simplest example is projective space, for which we have $\mathrm{Gr}_{\mathbb{C}}(1, \mathbb{C}^{m+1}) = \mathbb{CP}^m$. For a more detailed description of Grassmannians we refer to Section 10.1 of [10].

[5]The precise statement is that the group $\mathrm{GL}(H_{\mathbb{C}})$ acts transitively on $\check{\mathcal{F}}$, so that it is homogeneous and therefore nonsingular.

Proposition 1.3.2. *The group $G_{\mathbb{C}}$ acts transitively on $\check{\mathcal{D}}$, i.e. for any two filtrations $F^\bullet, \widetilde{F}^\bullet \in \check{\mathcal{D}}$ there exists a $g \in G_{\mathbb{C}}$ with $gF^\bullet = \widetilde{F}^\bullet$.*

The proof is a matter of straightforward but slightly tedious linear algebra, and therefore we omit it and refer to [13] for details. The main point is to construct a type of basis for $H_{\mathbb{C}}$ that is adapted to the polarization form $q_{\mathbb{C}}$ and a given filtration $F^\bullet \in \check{\mathcal{D}}$, which allows for the construction of an element g as above. Along the same lines it can be shown that a similar result holds for the smaller group

$$G_{\mathbb{R}} := \mathrm{Aut}(H_{\mathbb{R}}, q_{\mathbb{R}}) = \{g \in \mathrm{GL}(H_{\mathbb{R}}) \,|\, q_{\mathbb{R}}(gu, gv) = q_{\mathbb{R}}(u, v) \text{ for all } u, v \in H_{\mathbb{R}}\}. \tag{1.19}$$

Here $H_{\mathbb{R}} := H_{\mathbb{Z}} \otimes \mathbb{R}$ and $q_{\mathbb{R}}$ is the resulting bilinear form on $H_{\mathbb{R}}$. We usually simply write G instead of $G_{\mathbb{R}}$. Note that G is a Lie group, realized as a closed subgroup of $\mathrm{GL}(H_{\mathbb{R}})$. Since this group consists of real transformations, it preserves the second bilinear relation (positivity) as well, and hence it acts naturally on $\mathcal{D}$.

Proposition 1.3.3. *The group G acts transitively on the period domain $\mathcal{D}$.*

These two propositions imply that $\check{\mathcal{D}}$ and $\mathcal{D}$ are homogeneous spaces which can be realized as quotients of $G_{\mathbb{C}}$ and G. Let us fix a basepoint filtration $F_0^\bullet \in \mathcal{D}$, and define the stabilizer subgroups

$$B = (G_{\mathbb{C}})_{F_0^\bullet} = \{g \in G_{\mathbb{C}} \,|\, gF_0^\bullet = F_0^\bullet\}, \quad H = (G)_{F_0^\bullet} = B \cap G_{\mathbb{R}}. \tag{1.20}$$

From the transitivity we then indeed conclude that the spaces $\check{\mathcal{D}}$ and $\mathcal{D}$ are realized as quotients $\check{\mathcal{D}} \cong G_{\mathbb{C}}/B$ and $\mathcal{D} \cong G/H$. For later purposes, we note the following.

Proposition 1.3.4. *The stabilizer subgroup $H \subseteq G$ is compact.*

Proof. Since H is a group of real transformations and $H_0^{p,q} = F_0^p \cap \overline{F}_0^q$, H fixes the Hodge subspaces $H_0^{p,q}$. It therefore also fixes corresponding the Weil operator $C_0 = \bigoplus_{p+q=k} i^{p-q}\mathrm{id}_{H_0^{p,q}}$. It follows that H is a subgroup of the orthogonal group of the positive definite form h_0 defined by $h_0(u, v) = q_{\mathbb{C}}(u, C_0\overline{v})$. This orthogonal group is compact because h_0 is positive definite, and hence H is compact. □

We now revisit the torus example to see how this construction works explicitly.

Example 1.3.5. The polarization form on $H^1(\mathbb{T}; \mathbb{C})$ is non-degenerate and anti-symmetric, and therefore symplectic. We may thus identify $H^1(\mathbb{T}; \mathbb{C}) = \mathbb{C}^2$ equipped with the standard basis $\{\alpha, \beta\}$ and assume that $q_{\mathbb{C}}$ is given by the standard symplectic form

$$q_{\mathbb{C}}(\alpha, \alpha) = 0, \quad q_{\mathbb{C}}(\alpha, \beta) = 1, \quad q_{\mathbb{C}}(\beta, \beta) = 0.$$

We then have $G_{\mathbb{C}} = \mathrm{Sp}(2, \mathbb{C}) = \mathrm{SL}(2, \mathbb{C})$ and $G = \mathrm{Sp}(2, \mathbb{R}) = \mathrm{SL}(2, \mathbb{R})$. Hodge structures on $H^1(\mathbb{T}; \mathbb{C})$ are now fully determined by a choice of one-dimensional subspace $H^{1,0} = F^1 \subseteq \mathbb{C}^2$, and hence parametrized by the complex projective space $\mathbb{CP}^1$. The first bilinear relation $q_{\mathbb{C}}(F^1, F^1) = 0$ is automatically satisfied since F^1 is one-dimensional and $q_{\mathbb{C}}$ is anti-symmetric, so we find $\check{\mathcal{D}} = \mathbb{CP}^1$. Imposing the second bilinear relation again leads to $\mathcal{D} = \mathbb{H}$, and as a homogeneous space we have $\mathcal{D} = \mathrm{SL}(2, \mathbb{R})/\mathrm{U}(1)$.

Using linear algebraic techniques, we can in fact classify period domains [13].

Proposition 1.3.6. *Let $\mathcal{D} = G/H$ be the period domain for weight k Hodge structures with Hodge numbers $\{h^{p,q}\}$ on $H_{\mathbb{C}}$ with polarization form $q_{\mathbb{C}}$ and* $\dim H_{\mathbb{C}} = 2n$*.*

(i) If the weight $k = 2m+1$ is odd, then $q_{\mathbb{C}}$ is anti-symmetric and

$$G \cong \mathrm{Sp}(2n, \mathbb{R}), \quad H = \prod_{p \le m} \mathrm{U}(h^{p,q}), \tag{1.21}$$

and the quotient G/H is connected and non-compact.

(ii) If the weight $k = 2m$ is even, then $q_{\mathbb{C}}$ is symmetric and

$$G \cong \mathrm{SO}(s, t), \quad H = \prod_{p < m} \mathrm{U}(h^{p,q}) \times \mathrm{SO}(h^{m,m}), \tag{1.22}$$

and the quotient G/H has two connected components and is compact. Here $s = \sum_{p \text{ even}} h^{p,q}$ and $t = \sum_{p \text{ odd}} h^{p,q}$.

1.4 Variations of Hodge Structure and Period Maps

In geometry we often encounter the situation that a certain space is not alone, but comes in a family of spaces depending smoothly on variables in some parameter space. This happens for instance if we are interested in deformations of geometric structures on a space, such as the complex structure or the Kähler structure. We can think of such a family of spaces as a bundle, whose base space is the parameter space and whose fibers are the deformed copies of the space that we started with. If these spaces are compact Kähler manifolds, each fiber will have a Hodge structure, and we can now ask how this Hodge structure changes as we move in the base space. This is the main idea behind a *variation of Hodge structure*. Let us now make this more precise.

1.4.1 Geometric Setting

Definition 1.4.1. A *family of complex manifolds* is a surjective proper holomorphic submersion $\pi : \mathcal{X} \to \mathcal{M}$ of connected complex manifolds.

Here we think of $\mathcal{M}$ as the parameter space of the family. Demanding that π is surjective and holomorphic implies that the fibers $X_z = \pi^{-1}(z)$ are complex submanifolds. Furthermore, the condition that π is proper ensures that all the X_z are compact. By invoking a famous result due to Ehresmann, we are able to conclude that from a differential-geometric viewpoint, all the fibers X_z are the same. The precise statement is as follows [12].

Theorem 1.4.2 (Ehresmann)**.** *Let $\pi : \mathcal{X} \to \mathcal{M}$ be a family of connected complex manifolds. Then π is a smooth fiber bundle, and in particular all the fibers X_z for $z \in \mathcal{M}$ are diffeomorphic.*

Since we are interested in spaces whose cohomology carries a Hodge structure, we specialize to the case that the fibers X_z are n-dimensional Kähler manifolds with integral Kähler forms ω_z that vary in a locally constant way[6]. This setting occurs for instance if $\mathcal{X}$ and $\mathcal{M}$ are smooth quasi-projective varieties and π is algebraic with projective fibers. Since the fibers are then compact Kähler, the cohomology of every fiber X_z has a Hodge structure

$$H^k(X_z;\mathbb{C}) = \bigoplus_{p+q=k} H^{p,q}(X_z),$$

and the Hodge structure on the primitive cohomology $H^k_{\mathrm{p}}(X_z;\mathbb{C})$ is polarized by the blinear form

$$q_k(u,v) = (-1)^{k(k-1)/2}\int_{X_z} \omega_z^{n-k}\wedge u\wedge v.$$

The spaces $\{H^k_{\mathrm{p}}(X_z;\mathbb{C})\}_{z\in\mathcal{M}}$ together form a complex vector bundle $\mathcal{H}^k_{\mathbb{C}}\to\mathcal{M}$ with distinguished subbundles $\mathcal{H}^k_{\mathbb{R}}$ and $\mathcal{H}^k_{\mathbb{Z}}$ with fibers $H^k_{\mathrm{p}}(X_z;\mathbb{R})$ and $H^k_{\mathrm{p}}(X_z;\mathbb{Z})$ respectively[7]. It is possible to show that the Hodge numbers $h^{p,q}$ of the Hodge structure on the fibers remain constant when varying $z\in\mathcal{M}$ (see e.g. Chapter 9 of [10]), and therefore $\mathcal{H}^k_{\mathbb{C}}$ comes with smooth subbundles $\mathcal{H}^{p,q}$ whose fibers are the (p,q)-parts of the Hodge structure on the fibers of $\mathcal{H}^k_{\mathbb{C}}$. The key observation is now the following.

Observation 1.4.3. The geometry of the bundles $\mathcal{H}^{p,q}$ encodes how the Hodge structure on $H^k_p(X_z;\mathbb{C})$ varies with $z\in\mathcal{M}$.

The lattice bundle $\mathcal{H}^k_{\mathbb{Z}}$ can be identified with a local system, and hence the complex vector bundle $\mathcal{H}^k_{\mathbb{C}} = \mathcal{H}^k_{\mathbb{Z}}\otimes\mathcal{O}_{\mathcal{M}}$ comes equipped with a flat connection ∇ called the Gauss-Manin connection (cf. Appendix **??**). This connection has a non-trivial behavior with respect to the bundles $\mathcal{H}^{p,q}$. To see this, we must again rephrase the situation in terms of the Hodge filtrations $(F^\bullet)_z = F^\bullet H^k_{\mathrm{p}}(X_z;\mathbb{C})$, which now give rise to a filtration of vector bundles

$$\mathcal{H}^k_{\mathbb{C}} = \mathcal{F}^0 \supseteq \mathcal{F}^1 \supseteq \cdots \supseteq \mathcal{F}^k.$$

One of the foundational results in variations of Hodge structure is the following Theorem of Griffiths [14].

Theorem 1.4.4. *In the present setting, the bundles $\mathcal{F}^p\to\mathcal{M}$, $0\le p\le k$ are holomorphic vector bundles, and*

$$\nabla\underline{\mathcal{F}}^p \subseteq \underline{\mathcal{F}}^{p-1}\otimes\Omega^1_{\mathcal{M}} \tag{1.23}$$

where ∇ is the Gauss-Manin connection.

Here $\underline{\mathcal{F}}^p$ denotes the sheaf of sections of $\mathcal{F}^p$. The first part of the theorem is already non-trivial, since in general the vector bundles $\mathcal{H}^{p,q}$ are smooth but not holomorphic. For us this will be the main advantage of working with the Hodge filtration instead of the equivalent decomposition in terms of (p,q)-spaces. The second part of the theorem is a property called Griffiths transversality, and it captures the fact that the Hodge filtration does not vary arbitrarily over the base $\mathcal{M}$. More explicitly,

[6] The precise statement is that we demand that the ω_z form a section of the direct image sheaf $R^2\pi_*(\mathbb{Z})$ [15]. Although it takes some work to define such a sheaf, the intuitive picture is that it is associated to the presheaf $U\mapsto H^2(\pi^{-1}(U);\mathbb{Z})$ by sheafification.

[7] Note that $\mathcal{H}^k_{\mathbb{R}}$ is a vector bundle whereas $\mathcal{H}^k_{\mathbb{Z}}$ is not. The latter is referred to as a lattice bundle.

Griffiths transversality says that if u is a section of $\mathcal{F}^p$ and ξ is a holomorphic vector field, then $\nabla_\xi u$ does not take arbitrary values in $\mathcal{H}^k_{\mathbb{C}}$, but takes values in the smaller bundle $\mathcal{F}^{p-1}$.

The objects arising in the setting discussed above are now collected into the following definition.

Definition 1.4.5. A *variation of Hodge structure* of weight k on a connected complex manifold $\mathcal{M}$ consists of the following data:

(i) a local system $\mathcal{H}_{\mathbb{Z}}$;

(ii) a $(-1)^k$-symmetric form q on $\mathcal{H}_{\mathbb{Z}}$;

(iii) a holomorphic decreasing filtration $\mathcal{F}^\bullet$ of the flat complex vector bundle $\mathcal{H}_{\mathbb{C}} = \mathcal{H}_{\mathbb{Z}} \otimes \mathcal{O}_{\mathcal{M}}$;

such that Griffiths transversality holds and for every $z \in \mathcal{M}$, the data $(\mathcal{H}_{\mathbb{Z}}, q, \mathcal{F}^\bullet)$ evaluated at z determines a weight k Hodge structure.

In later chapters we will encounter various variations of Hodge structure, and for now we only provide one example [13].

Example 1.4.6. (Legendre Family) Consider the algebraic equation

$$y^2 = x(x-1)(x-z), \tag{1.24}$$

where x and y are affine coordinates on $\mathbb{CP}^2$ and z is interpreted as a parameter in $\mathbb{CP}^1$. For $z \neq 0, 1, \infty$, this equation defines a curve in $\mathbb{CP}^2$ which is diffeomorphic to a torus, and hence we obtain a family of tori $\mathcal{E} \to \mathcal{M}$, where

$$\mathcal{E} = \{((x:y:1), z) \in \mathbb{CP}^2 \times \mathbb{CP}^1 \,|\, y^2 = x(x-1)(x-z)\} \quad \text{and} \quad \mathcal{M} = \mathbb{CP}^1 \setminus \{0, 1, \infty\}. \tag{1.25}$$

The setup is illustrated in Figure 1.2. The Hodge structure on $H^1(\mathcal{E}_z; \mathbb{C})$ is polarized by the form

$$q_z(u, v) = \int_{\mathcal{E}_z} u \wedge v. \tag{1.26}$$

It can be shown that the holomorphic form whose cohomology class spans the subspace $H^{1,0}_z \subseteq H^1(\mathcal{E}_z; \mathbb{C})$ is given by

$$\omega_z = \frac{dx}{y} = \frac{dx}{\sqrt{x(x-1)(x-z)}}, \tag{1.27}$$

and the variation of ω_z with z defines the subbundle $\mathcal{H}^{1,0} = \mathcal{F}^1 \subseteq \mathcal{H}^2_{\mathbb{C}}$ which fully encodes the variation of Hodge structure of this family of tori. The holomorphicity of this bundle follows from the holomorphicity of ω_z in the parameter z. For $z \in \{0, 1, \infty\}$, the equation $y^2 = x(x-1)(x-z)$ describes a singular variety, interpreted as a degeneration of the torus.

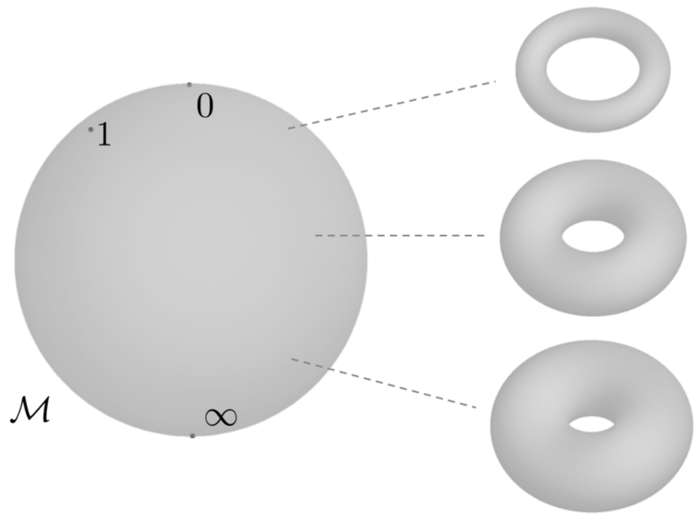


Figure 1.2: Illustration of the Legendre family of tori $\mathcal{E} \to \mathcal{M}$. The base space is $\mathcal{M} = \mathbb{CP}^1 \setminus \{0, 1, \infty\}$, and the fibers $\mathcal{E}_z$ give rise to tori of different shapes. The shape is encoded by the period ratio $\tau(z) = B(z)/A(z)$.

1.4.2 The Period Map and Monodromy

The connection ∇ that is present in a variation of Hodge structure enables us to parallel transport data along paths in the base space. Fixing a basepoint $z_0 \in \mathcal{M}$ and denoting $H_0 = H^k(X_{z_0}; \mathbb{C})$, every choice of $z \in \mathcal{M}$ and path γ from z to z_0 determines a polarized Hodge structure $\{H_\gamma^{p,q}\}$ on the fixed vector space H_0 by parallel transport of $H_z^{p,q}$ along γ. Because the Gauss-Manin connection is flat, the spaces $H_\gamma^{p,q}$ depend only on the homotopy class of γ. Parallel transport along loops based at z_0 may also lead to distinct Hodge structures, and this is described by a representation

$$\rho : \pi_1(\mathcal{M}, z_0) \to \mathrm{GL}(H_0) \tag{1.28}$$

called the monodromy representation. This leads to the following observation.

> **Observation 1.4.7.** Every point $z \in \mathcal{M}$ determines a weight k polarized Hodge structure on H_0 with fixed Hodge numbers $\{h^{p,q}\}$, unique up to monodromy.

To see the concept of monodromy in action, we return to the torus.

Example 1.4.8. (Monodromy in the Legendre Family) To understand how monodromy works in the context of the Legendre family, we have to understand more precisely how the equation

$$y^2 = x(x-1)(x-z), \quad z \in \mathcal{M} \tag{1.29}$$

gives rise to a torus. For x away from any of the points $0, 1, \infty, z \in \mathbb{CP}^1$, this equation has two solutions for y. We can therefore construct the solution set by assembling two copies of $\mathbb{CP}^1$ along two branch cuts, which we choose to go from 0 to z and from 1 to ∞. The two copies of $\mathbb{CP}^1$ with the branch cuts removed are called *sheets*. In this way, we obtain a torus, as shown in Figure 1.3.

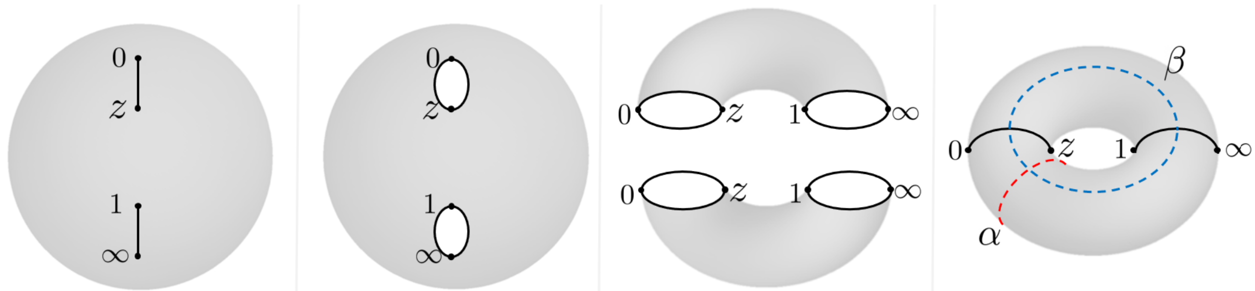


Figure 1.3: Construction of a torus as the solution set of $y^2 = x(x-1)(x-z)$. The leftmost figure shows $\mathbb{CP}^1$ with two branch cuts running from 0 to z and from 1 to ∞. Proceeding to the right, the branch cuts are opened and attached to another copy of $\mathbb{CP}^1$ with the same branch cuts. The rightmost figure shows the resulting torus, together with the cycles $\alpha, \beta \in H_1(\mathbb{T};\mathbb{Z})$ of the canonical homology basis.

The base space of the Legendre family $\mathcal{M} = \mathbb{CP}^1 \setminus \{0, 1, \infty\}$ is not simply connected, and has a fundamental group generated by two loops winding around the punctures. Let us consider a loop γ in $\mathcal{M}$ based at z, winding around the puncture at 0. The action of the resulting monodromy operator $\rho(\gamma)$ on $H^1(\mathcal{E}_z;\mathbb{C})$ can be understood by looking at the homology basis $\alpha, \beta \in H_1(\mathcal{E}_z;\mathbb{Z})$. If we take z along the path γ, the cycles α and β are continuously deformed, and after z has encircled 0 the cycles have been transformed nontrivially. This is illustrated in Figure 1.4.

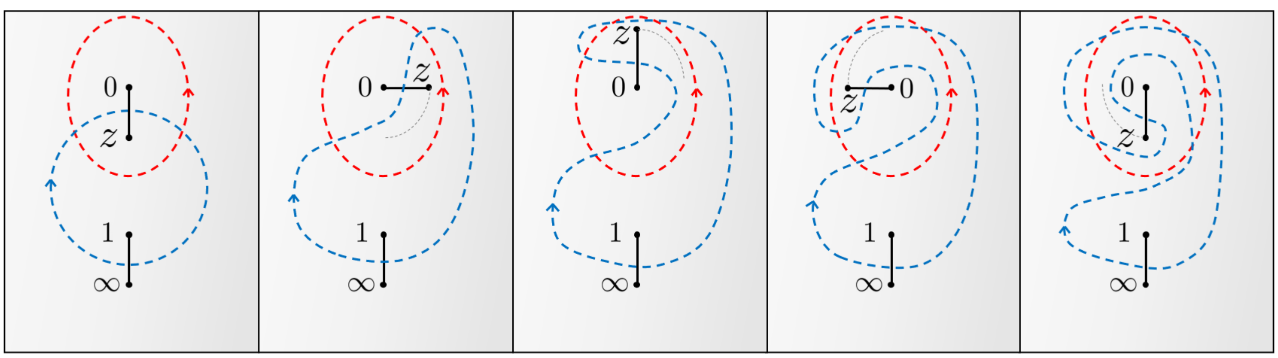


Figure 1.4: An illustration of monodromy in the Legendre family, inspired by [13]. The leftmost panel shows the canonical homology basis $\alpha, \beta \in H_1(\mathcal{E}_z;\mathbb{Z})$ of figure 1.1, indicated by the red and blue cycle, respectively. A cycle crossing a branch cut is understood to traverse from one sheet into the other. In the leftmost panel the torus is shown for a certain $z \in \mathcal{M}$, and from left to right the parameter z encircles the puncture 0. This rotates the branch cut, and the cycles are deformed continuously. In the rightmost panel, z reaches its original value after fully encircling the singularity. The red α-cycle is unchanged, but the blue β-cycle has transformed nontrivially: it has undergone a monodromy transformation.

Upon carefully examining Figure 1.4, for instance by counting intersection numbers, we find that monodromy acts on the cycles α and β as $\alpha \mapsto \alpha$, $\beta \mapsto \alpha + 2\beta$. The periods $A = \int_\alpha \Omega$ and $B = \int_\beta \Omega$ then transform accordingly, and from this we find how the monodromy operator $\rho(\gamma)$ acts on the Hodge structure on $H^1(\mathcal{E}_z;\mathbb{C})$.

In general, let us denote the image of the monodromy representation by $\Gamma \subseteq \mathrm{GL}(H_0)$. Elements in the monodromy group Γ preserve the form $q_0 = q_{z_0}$, as it is constructed from integration and the wedge product, and therefore Γ is a subgroup of $G_{\mathbb{C}} = \mathrm{Aut}(H_0, q_0)$. In fact, the connection ∇ comes from a local system and hence parallel transport preserves integrality, so that Γ is a subgroup $\Gamma \subseteq G_{\mathbb{Z}} = \mathrm{Aut}((H_0)_{\mathbb{Z}}, q_0)$ with $(H_0)_{\mathbb{Z}} = H^k_{\mathrm{p}}(X_{z_0}; \mathbb{Z})$. Observation 1.4.7 now implies that every z determines a point in the period domain $\mathcal{D}$ of weight k polarized Hodge structures on H^0 with Hodge numbers $\{h^{p,q}\}$, up to monodromy. We thus obtain a map

$$\Phi : \mathcal{M} \to \Gamma \backslash \mathcal{D} \tag{1.30}$$

called the *period map*[8]. The period map contains all the information of the variation of Hodge structure [15], and the rest of this chapter is devoted to understanding the properties of this map. Writing the period domain as $\mathcal{D} = G/H$, the period map assigns to a point z a class $[g] \in \Gamma \backslash G/H$, where g is such that

$$g \cdot F_0^\bullet = F_z^\bullet. \tag{1.31}$$

Here $F_0^\bullet$ is the reference Hodge filtration on H_0. Instead of taking Γ to be the monodromy group, it is customary to take $\Gamma = G_{\mathbb{Z}}$, which contains the monodromy group. The space $\Gamma \backslash G/H$ then parametrizes polarized Hodge structures up to integral isomorphism. More generally, Γ is taken to be any finite index subgroup of $G_{\mathbb{Z}}$ containing the monodromy group. Such a subgroup is called an arithmetic subgroup, and in this setting the double quotient space $\Gamma \backslash G/H$, which forms the target space of the period map, is called an *arithmetic quotient*. These spaces take a central role in this thesis and will be studied in more detail in Chapter 4. From now on, we assume that Γ is an arithmetic subgroup of G.

Recall that the quotient $\mathcal{D} = G/H$ has a complex structure, determined by its embedding into the space of filtrations on H_0. The action of Γ on $\mathcal{D}$ is in general not free, but its stabilizers are finite as they lie in the intersection of $\Gamma \subseteq G_{\mathbb{Z}}$ and the compact subgroup H [18]. The arithmetic quotient $\Gamma \backslash G/H$ is therefore not in general smooth, but it is an *analytic space* whose singularities are relatively mild[9]. On an analytic space, the notion of being holomorphic is well-defined, and as a consequence of the holomorphicity for the bundles we have the following important result of Griffiths for the period map [14].

Theorem 1.4.9. *The period map $\Phi : \mathcal{M} \to \Gamma \backslash \mathcal{D}$ associated to a variation of Hodge structure on $\mathcal{M}$ is holomorphic.*

A further important property of the period map is that its differential takes values in a special 'horizontal' subbundle of the tangent bundle on $\mathcal{D}$. It is the analogue of Griffiths transversality of the bundles $\mathcal{F}^\bullet$. This property can be used to *define* a variation of Hodge structure in terms of a period map. Though it is an essential property, we will not make use of it in the following and therefore not discuss it further. We refer to [15] for details.

[8] We write the quotient by the monodromy group by $\Gamma \backslash \mathcal{D}$ instead of $\mathcal{D}/\Gamma$ for later convenience.

[9] An analytic space is defined analogously to an algebraic variety, with the difference being that it is locally defined by the vanishing of analytic functions rather than polynomials. The precise statement formalizing the mildness of the singularities of $\Gamma \backslash G/H$ is that it is a *normal* analytic space [18].

1.5 Asymptotic Hodge Theory

It is often the case that the family of smooth projective manifolds $\mathcal{X} \to \mathcal{M}$ under consideration is contained a larger family $\overline{\mathcal{X}} \to \overline{\mathcal{M}}$, where the fibers of the points $z \in \mathcal{M}_{\text{sing}} := \overline{\mathcal{M}} \setminus \mathcal{M}$ are singular.

> **Example 1.5.1.** The base space of the Legendre family is $\mathcal{M} = \mathbb{CP}^1 \setminus \{0, 1, \infty\}$, and the family can be extended to a family $\overline{\mathcal{X}} \to \overline{\mathcal{M}}$, where $\overline{\mathcal{M}} = \mathbb{CP}^1$. The fibers of the points $\{0, 1, \infty\} = \mathcal{M}_{\text{sing}}$ are singular elliptic curves.

On such singular fibers the Hodge decomposition theorem no longer holds, and the Hodge structure on the cohomology breaks down. It is then natural to ask how the variation of Hodge structure behaves near the singular locus $\mathcal{M}_{\text{sing}}$, and this is the subject of asymptotic Hodge theory.

As in [15], we focus on the case that $\mathcal{M}_{\text{sing}}$ is a codimension one subvariety and assume that $\overline{\mathcal{M}}$ is smooth, containing $\mathcal{M}_{\text{sing}}$ as a divisor whose singularities are at worst normal crossings. For general $\mathcal{M}$ and $\overline{\mathcal{M}}$, it is argued in [15] always reduce to this case by resolving singularities [19]. Hence, the inclusion $\mathcal{M} \subseteq \overline{\mathcal{M}}$ locally takes the form

$$(\Delta^*)^l \times \Delta^{m-l} \subseteq \Delta^k, \tag{1.32}$$

where $\Delta = \{z \in \mathbb{C} \,|\, |z| < 1\}$ is the unit disk and $\Delta^* = \Delta \setminus \{0\}$ is the punctured disk. To study the asymptotic behavior of the variation of Hodge structure, we proceed locally and consider period maps

$$\Phi : (\Delta^*)^l \times \Delta^{m-l} \to \Gamma \backslash \mathcal{D}, \tag{1.33}$$

and we are in particular interested in the limiting behavior of Φ near the singular loci defined by $z_i = 0$ with $(z_1, \dots, z_m) \in \Delta^m$. The fact that there is no Hodge structure for the fibers on these singular loci manifests itself in the singularity of Φ. However, as we will see below, the singularities of the period map are moderate, and the period map is in fact 'tame' in a precise sense. This will be the main content of Chapter 4.

The universal cover of the punctured disk Δ^* is the upper half-plane $\mathbb{H}$ with the covering map

$$\begin{aligned} p : \mathbb{H} &\to \Delta^*, \\ t &\mapsto e^{2\pi i t}. \end{aligned} \tag{1.34}$$

On the universal cover there is no monodromy, and hence the period map lifts to a map

$$\widetilde{\Phi} : \mathbb{H}^l \times \Delta^{m-l} \to \mathcal{D} \tag{1.35}$$

such that the diagram

$$\begin{array}{ccc} \mathbb{H}^l \times \Delta^{m-l} & \xrightarrow{\widetilde{\Phi}} & \mathcal{D} \\ \downarrow & & \downarrow \\ (\Delta^*)^l \times \Delta^{m-l} & \xrightarrow{\Phi} & \Gamma \backslash \mathcal{D} \end{array}$$

commutes. As a remnant of the monodromy experienced by Φ, this map has the property

$$\widetilde{\Phi}(t_1, \dots, t_i + 1, \dots, t_l, z_{l+1}, \dots, z_m) = T_i \cdot \widetilde{\Phi}(t_1, \dots, t_l, z_{l+1}, \dots, z_m),$$

where the elements $T_i \in \Gamma$ together generate the monodromy group. Note that the fundamental group of a product of puncured disks is abelian, so the elements $\{T_i\}_{i=1,\dots,l}$ are mutually commuting. It can be shown that the monodromy operators T_i are quasi-unipotent [15], i.e. there are integers m_i and n_i such that

$$(T_i^{m_i} - 1)^{n_i} = 0.$$

Under a finite covering transformation $\Delta^* \to \Delta^*$, $z_i \mapsto z_i^{m_i}$, the monodromy operator T_i is mapped to $T_i^{m_i}$, and therefore by applying such a transformation to each z_i we may assume that the T_i are unipotent. Let $\mathfrak{g}$ be the Lie algebra of G. We then define the log-monodromy operators $N_i \in \mathfrak{g}$ by

$$N_i := \log T_i = \sum_{j=1}^{\infty} (-1)^{j-1} \frac{1}{j} (T_i - 1)^j, \tag{1.36}$$

where the latter is a finite sum because T_i is unipotent. This in turn implies that the N_i are nilpotent. The log-monodromy operators can be used to construct a map $\mathbb{H}^l \times \Delta^{k-l} \to \check{\mathcal{D}}$ as

$$\widetilde{\Psi}(t_1, \dots, t_l, z_{l+1}, \dots, z_m) := g_N(t_1, \dots, t_l) \cdot \widetilde{\Phi}(t_1, \dots, t_l, z_{l+1}, \dots, z_m), \tag{1.37}$$

where $g_N : \mathbb{H}^l \to G_{\mathbb{C}}$ is a group-valued map defined by

$$g_N(t_1, \dots, t_l) := \exp\left(- \sum_{j=1}^{l} t_j N_j \right). \tag{1.38}$$

Let us explain the meaning of $\widetilde{\Psi}$. The map g_N exponentiates a linear combination of the log-monodromy operators in $\mathfrak{g}$, and because the coefficients of the linear combination are the complex coordinates of $\mathbb{H}^l$, the resulting group element lands in the complex group $G_{\mathbb{C}}$. With this group element we act on $\widetilde{\Phi}$, and since the period domain $\mathcal{D}$ is only preserved by the real group G, the result lies in the larger space $\check{\mathcal{D}}$. By construction, the map $\widetilde{\Phi}$ is invariant under translations $t_i \mapsto t_i + 1$, and thus the action of the element $g_N(t_1, \dots, t_l)$ can be thought of as 'unwinding the monodromy'. The invariance under $t_i \mapsto t_i + 1$ implies that $\widetilde{\Phi}$ descends to a map $\Psi : (\Delta^*)^l \times \Delta^{m-l} \to \check{\mathcal{D}}$. In terms of the map Ψ, we can state the nilpotent orbit theorem of Schmid, a central result in the study of asymptotic Hodge theory [15]. Below we abbreviate $(t_1, \dots, t_l, z_{l+1}, \dots, z_m)$ by (t, z) and set $t_i = x_i + iy_i$.

> **Theorem 1.5.2.** *(Nilpotent Orbit Theorem) The map Ψ extends to a holomorphic map $\Delta^m \to \check{\mathcal{D}}$. Let $\Psi_\infty(z) = \Psi(0, z)$ denote the value of Ψ on the singular locus. Then the lifted period map $\widetilde{\Phi}$ is asymptotic to a nilpotent orbit, i.e.*
>
> $$\widetilde{\Phi}(t, z) \sim \Psi_{\text{nil}}(t, z) := \big(g_N(t)\big)^{-1} \cdot \Psi_\infty(z). \tag{1.39}$$
>
> *More precisely, for any $0 < \eta < 1$ there exist $\alpha, \beta \geq 0$ such that*
>
> $$d(\widetilde{\Phi}(t, z), \Psi_{\text{nil}}(t, z)) \leq \left(\prod_{j=1}^{l} y_j \right)^{\beta} \sum_{j=1}^{l} e^{-2\pi y_j} \quad \textit{and} \quad \Psi_{\text{nil}}(t, z) \in \mathcal{D} \tag{1.40}$$
>
> *for $y_i \geq \alpha$, $|z_i| \leq \eta$. Here d is any G-invariant Riemannian distance function on $\mathcal{D}$.*

The first part of the theorem says that the map Ψ that we constructed is non-singular, and that it extends across the singular locus $\mathcal{M}_{\text{sing}}$. By acting with the inverse monodromy factor g_N^{-1}, we obtain the nilpotent orbit Ψ_{nil}. Viewed as a multi-valued function on $(\Delta^*)^l \times \Delta^{m-l}$ via the relation

$$t_i = \frac{1}{2\pi i} \log(z_i) \quad \text{for} \quad 1 \leq i \leq l,$$

we see that the nilpotent orbit is singular at $z_i \to 0$. However, since the log-monodromy operators N_i are nilpotent, the function g_N is in fact a polynomial in the variables $\log(z_i)$, and as a result the singularities of the nilpotent orbit are mild. Near the singular locus $z_i = 0$ $(1 \leq i \leq l)$, the theorem states that the period map is well-approximated by the nilpotent orbit, and this therefore tells us about the nature of the singularities of the period map.

There are several refinements of the nilpotent orbit theorem that describe the singularities of the period map in more detail. The most important one is the *SL(2)-orbit theorem*. It owes its name to a certain embedding $\mathrm{SL}(2,\mathbb{R}) \to G$ which plays a major role in the statement of the theorem. The first version of this theorem is formulated for a single variable $(l = 1)$ and due to Schmid [15], and the multi-variable generalization was obtained by Cattani, Kaplan and Schmid [20]. The multi-variable version is significantly more complicated, which is already suggested by the fact that it was proved more than a decade later than the one-variable version.

We will not state the SL(2)-orbit theorem because of its technical nature, but we will discuss several of its consequences. The first of these consequences gives an interpretation to the limit $\Psi_\infty(z) \in \check{\mathcal{D}}$. Recalling the construction of $\check{\mathcal{D}}$, we see that $\Psi_\infty(z)$ defines a filtration $F^\bullet_\infty(z)$ on a reference vector space H_0, but since it generally does not lie in the period domain $\mathcal{D} \subseteq \check{\mathcal{D}}$ it does not define a polarized Hodge structure on H_0; this captures the idea that the Hodge structure 'breaks down' on a singular fiber. However, a result of the SL(2)-orbit theorem is that $\Psi_\infty(z)$ gives rise to a weaker type of Hodge structure [21].

Definition 1.5.3. A *mixed Hodge structure* consists of

(i) a free abelian group of finite rank $H_\mathbb{Z}$;

(ii) an increasing filtration $W_\bullet$ on $H_\mathbb{Z}$;

(iii) a decreasing filtration $F^\bullet$ on $H_\mathbb{C} = H_\mathbb{Z} \otimes \mathbb{C}$;

such that the filtration induced by $F^\bullet$ on each graded quotient $\mathrm{Gr}_p^W H_\mathbb{C} := (W_p)_\mathbb{C}/(W_{p-1})_\mathbb{C}$ defines a Hodge structure of weight p on $\mathrm{Gr}_p^W H_\mathbb{Z}$.

The significance of mixed Hodge structures is that they make sense on singular spaces.

Example 1.5.4. Consider the degenerate surface X obtained from shrinking one cycle of a genus 2 surface, as shown in Figure 1.5. The cohomology group $H^1(X;\mathbb{Z})$ has rank 3, and is generated by the Poincaré duals $\alpha^*, \beta^*, \gamma^*$ of the cycles $\alpha, \beta, \gamma \in H_1(X;\mathbb{Z})$ shown in Figure 1.5. Purely for dimensional reasons, we thus see that $H^1(X;\mathbb{C})$ cannot have a weight 1 Hodge structure, as it would in the non-singular case.

There is one holomorphic 1-form on X, whose class Ω lies in the subspace $\mathbb{C}\alpha^* \oplus \mathbb{C}\beta^*$. Consider the spaces

$$W_0 = \mathbb{Z}\gamma^*, \quad W_1 = H^1(X;\mathbb{Z}),$$
$$F^1 = \mathbb{C}\Omega, \quad F^0 = H^1(X;\mathbb{C}).$$

The filtrations $W_\bullet$ and $F^\bullet$ together form a mixed Hodge structure on $H^1(X;\mathbb{Z})$. The content of this statement is the fact that the graded quotient

$$\mathrm{Gr}_1^W = W_1/W_0 = \mathbb{Z}\alpha^* \oplus \mathbb{Z}\beta^*$$

carries a weight 1 Hodge structure determined by $H^{1,0} = F^1$ and $H^{0,1} = \overline{F^1}$.

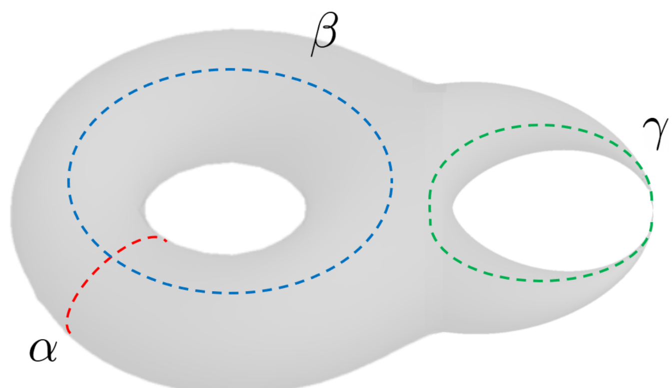


Figure 1.5: A degenerate complex curve of genus 2. The first homology group $H_1(X;\mathbb{Z})$ is generated by the cycles α, β and γ.

In general, we can construct a certain filtration $W_\bullet$ on $(H_0)_\mathbb{Z}$ associated to a singular limit $z_i \to 0$ as follows [22]. The log-monodromy N_i operator determines a filtration $W_\bullet$ by the formula

$$W_p = \bigoplus_{q \geq 1, p-k+1} \ker N_i^q \cap \operatorname{im} N_i^{q-p+k-1},$$

which is a finite filtration because N_i is nilpotent. The key result is now the following.

Theorem 1.5.5. *The triple $\big((H_0)_\mathbb{Z}, W_\bullet, F^\bullet_\infty(z)\big)$ defines a mixed Hodge structure.*

Here we stated the result for a single variable $z_i \to 0$, but it holds in greater generality for any number of variables on the punctured polydisk going to zero. As with the SL(2)-orbit theorem, this single variable result is due to Schmid [15], and the highly non-trivial generalization to multiple variables is due to Cattani, Deligne, and Kaplan [22].

1.6 The Hodge Conjecture

To conclude the chapter on Hodge theory, we briefly explore the *Hodge Conjecture*, one of the major unsolved problems in geometry[10]. In the story of Chapter 4 we will refer back to the material discussed here. For more details we refer to Chapter 11 of [10].

The Hodge Conjecture is a problem concerning the subvarieties of a smooth complex projective variety X, and in particular the interaction of such a subvariety with the cohomology of X. Assume that X is n-dimensional, and consider a smooth subvariety $Z \subseteq X$ of dimension k. If u is $2k$-form on X, we can pull it back along the inclusion $\iota : Z \hookrightarrow X$ and consider the value of the integral

$$\int_Z \iota^* u. \tag{1.41}$$

This integral vanishes unless u is of a special type. To see this, let us first assume that u is a pure form of degree (p, q). Locally, we can choose adapted holomorphic coordinates for Z, i.e. coordinates $z_1, \ldots, z_n$ on X such that Z is defined by the vanishing of $z_{k+1}, \ldots, z_n$. If $(p, q) \neq (k, k)$, then u must contain either a dz_j or a $d\bar{z}_{\bar{j}}$ with $j > k$, which vanishes when restricted to the subvariety Z. This shows that the integral can only be non-zero when u is of type (k, k).

With some work, this story can be translated to cohomology, and this leads to the conclusion that the Poincaré dual $Z^* \in H^{2k}(X; \mathbb{Z})$ has type (k, k) in the Hodge decomposition. Cohomology classes in $H^{2k}(X; \mathbb{Q})$ of type (k, k) are called *Hodge classes*, and we thus find that classes associated to subvarieties of X are Hodge classes. The Hodge Conjecture asks whether the converses is true.

Conjecture 1.6.1. *(Hodge Conjecture) Let X be a smooth complex projective variety. Then every Hodge class on X is a rational linear combination of the cohomology classes of subvarieties of X.*

Though a proof of the Hodge Conjecture appears to be out of reach at this moment, there is evidence in favor of the conjecture. The strongest evidence comes from variations of Hodge structure. For a smooth algebraic family $\mathcal{X} \to \mathcal{M}$ with associated variation of Hodge structure $(\mathcal{H}_{\mathbb{Z}}, \mathcal{F}^\bullet)$, we can define the so-called Hodge locus

$$\mathrm{HL}(\mathcal{M}) := \{z \in \mathcal{M} \mid (\mathcal{H}_{\mathbb{C}})_z \text{ contains a non-zero Hodge class}\}. \tag{1.42}$$

Due to the complicated nature of variations of Hodge structure, the Hodge locus is generally a complicated subset of $\mathcal{M}$. The holomorphicity of the period map implies that $\mathrm{HL}(\mathcal{M})$ is a countable union of analytic subvarieties of $\mathcal{M}$. The Hodge Conjecture predicts a much stronger result, namely

$$\text{the Hodge locus } \mathrm{HL}(\mathcal{M}) \text{ is a countable union of algebraic subvarieties of } \mathcal{M}, \tag{1.43}$$

a statement which is often simply phrased as the 'algebraicity of Hodge loci'. This means that, locally, the regions in the parameter space $\mathcal{M}$ where Hodge classes occur are determined by algebraic equations rather than analytic ones.

[10] In fact, it is one of the seven Millenium Prize Problems selected by the Clay Mathematics Institute.

The remarkable evidence for the conjecture comes from the fact that Cattani, Deligne, and Kaplan prove the algebraicity of Hodge loci unconditionally[11] [22]. The proof is often regarded as highly delicate and technical, and it requires the full analytic machinery of asymptotic Hodge theory, most notably the multi-variable SL(2)-orbit theorem [15, 21]. In Chapter 4, we will see how tame geometry provides a new perspective on this evidence for the conjecture. However, before we explore this further, we will first introduce the area of physics in which tame geometry makes a natural appearance, and this brings us to the next chapter on string theory.

[11] The word 'unconditionally' has two meanings here: i) irrespective of whether the Hodge conjecture holds or not; ii) for any variation of Hodge structure, not just those arising from an algebraic family $\mathcal{X} \to \mathcal{M}$.

Chapter 2

String Theory - Landscape and Swampland

Over the years, string theory has become a huge field with many intriguing aspects. As alluded to in the introduction, it is arguably the strongest present candidate for a theory that describes everything in nature, and for that reason it drives a significant amount of the research in high energy physics of today. In this chapter we provide an overview of some aspects of string theory relevant for this thesis, with an emphasis on the notions of *landscape* and *swampland*.

We start this chapter by briefly touching upon the basic starting point of string theory, including its supersymmetric formulation in terms of superstrings. This quickly leads us to five prominent variants of superstring theory which are formulated in ten dimensions. The passage from the ten dimensions required in string theory to the four dimensions that are observed is done by the idea of *compactification*, which is a major theme of this chapter. We are then led to consider the geometry of the extra dimensions, and find that deformations of this geometry appear as degrees of freedom in the four-dimensional theory. In addition, we will see that there is a vast amount of choices for the topology and geometry of the extra dimensions, leading to the idea that there is a landscape of four-dimensional theories arising from string theory. The story then proceeds with a discussion of a modern view on this landscape, in the form of the swampland program. The swampland program will form the setting in which tameness is introduced to physics.

The selection of material presented in this chapter is mostly based on appearance in later chapters. It therefore discusses mostly aspects of string theory which are relevant for the research surrounding this thesis, and gives an incomplete view of what the field is about. Nonetheless, we hope that the story of this chapter is clear and that it gives an impression of what some of the main questions and objects in string theory are. For a detailed introduction to string theory, we refer to [23].

2.1 Strings and Superstrings

2.1.1 Bosonic Strings

Fundamentally, string theory is a theory that explores the consequences of replacing point particles by extended objects. Among the extended objects the most basic ones are the one-dimensional objects: the strings. Whereas point particles can be thought of as tracing out a line in spacetime, strings will trace out a surface. In this way, the evolution of a string is described by a map

$$\Sigma \to M_d$$

from some surface Σ, called the worldsheet, into a d-dimensional spacetime M_d. There is a natural action that can be associated to such a map, namely the area of Σ as measured using the pullback of the spacetime metric. From this point of view we obtain a field theory on the worldsheet Σ whose degrees of freedom are the spacetime coordinates corresponding to the map $\Sigma \to M_d$. Since the spacetime coordinates are bosonic in nature, the theory formulated in this way is known as *bosonic string theory*. A truly remarkable feature of bosonic string theory is that, as a quantum theory, it is only consistent for a single value of the spacetime dimension, namely $d = 26$.

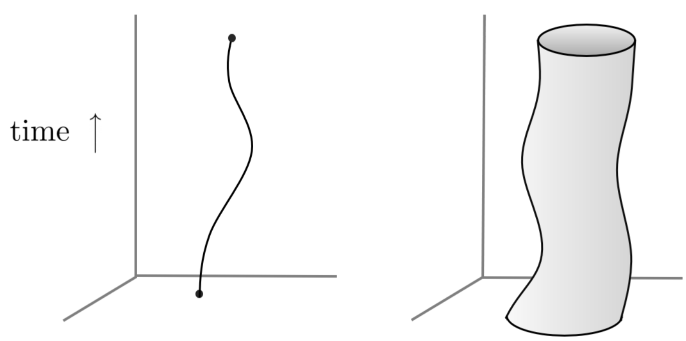


Figure 2.1: A particle and a string, tracing out a worldline and a worldsheet in spacetime, respectively.

The natural way to proceed is to study the equation of motion for the string. The solutions that one finds constitute a spectrum of vibrational modes, which quantum-mechanically are interpreted as states of the string. The various states of the strings may effectively be described as fields on the spacetime M_d. These fields have a variety of masses and spins, and among them there is the *graviton*, which is a massless spin-2 field that describes the propagation of gravity. This is a miraculous result, since it implies that a quantum theory of one-dimensional extended objects automatically incorporates gravity. This is one of the features that makes string theory so promising as a candidate theory of quantum gravity.

Unfortunately, there are a number of severe problems with the bosonic string that make it unsuitable as a theory of nature. Firstly, its spectrum contains a tachyon, i.e. a state with negative mass squared, which signals an instability in the theory. Secondly, the theory has no fermions, and a theory must have fermionic degrees of freedom for realistic matter to exist. It turns out that these problems can be simultaneously resolved by including fermions in the worldsheet field theory in a *supersymmetric* way. Since supersymmetry plays a major role in string theory, we briefly digress into a discussion of supersymmetry.

2.1.2 Supersymmetry

Supersymmetry is an important concept that pervades high energy physics. There is much to say about this topic, but will only review some essential aspects. We refer to [24] for a comprehensive overview of the topic and to [25] for a reference that emphasizes the geometric aspects underlying supersymmetry.

The starting point for supersymmetry is the general idea of symmetries in a physical theory. Two essential classes of symmetries are the symmetries of spacetime, encoded in the Poincaré group, and gauge symmetries, which play a crucial role in the transmission of forces. Since these classes are so ubiquitous in physics, it is natural to ask whether these are the only type of symmetries that a theory can have. It turns out that there is a way to incorporate additional symmetries if we modify how symmetries are generated. Symmetries in physics are generated by *charges*, which are usually bosonic objects. The modification required to generalize the possible symmetries is to allow the charges to be fermionic, and such charges are called *supercharges*. Symmetries generated by supercharges are incredibly constrained, and the resulting type of symmetry is supersymmetry. A necessary consequence of the fermionic nature of the supercharges is that they interchange bosonic and fermionic degrees of freedom:

$$\text{bosons} \xleftrightarrow{\text{supersymmetry}} \text{fermions.}$$

Since the generators of supersymmetry are fermions, they are spinors, and we generally denote them by Q^i_α. The type of spinor (e.g. Dirac, Weyl, Majorana) strongly depends on the spacetime dimension of the theory, since different spacetimes admit different spin structures. The index α labels the components of the spinor, and each component represents an individual supercharge. The upper index i counts the number of spinors; their total number is denoted by $\mathcal{N}$. This leads to two common ways of counting the amount of supersymmetry: the number $\mathcal{N}$ or the amount of supercharges. The number of components of a spinor depends on the dimension of spacetime that is considered, so the relation between these two counts also depends on the dimension. This idea will play a role later in Chapter 6.

The amount of supersymmetry cannot grow indefinitely. Acting on a field with successive supersymmetry transformations increases the spin of the field, there are no known consistent formulations of local field theories whose elementary fields have spin greater than 2. As a result one finds that the maximum allowed number of supercharges is 32. In four dimensions, this corresponds to $\mathcal{N} = 8$, and in ten dimensions this corresponds to $\mathcal{N} = 2$.

For phenomenological purposes, $\mathcal{N} = 1$ is the most relevant in four dimensions, since it is the minimal amount of supersymmetry and therefore the smallest deviation from the Standard Model. It provides a possible explanation or solution to several phenomenological problems, including the hierarchy problem, grand unification and dark matter [26].

As done in gauge theories, it is possible to promote supersymmetry from a global symmetry to a local symmetry. A remarkable consequence of this procedure is that in such a theory, gravity is inevitable. This can be seen from a part of the supersymmetry algebra, namely the anti-commutator of a supercharge with its conjugate, which schematically reads

$$\{Q, Q^\dagger\} \sim \gamma^\mu P_\mu. \tag{2.1}$$

Here γ^μ is a γ-matrix representing the Clifford algebra, and P_μ is the momentum operator. Momentum is the charge that generates translations, and making supersymmetry a local symmetry implies through equation (2.1) that the theory must include *local* translations as a symmetry. In other words, invariance under diffeomorphisms becomes a symmetry, which means that the theories include general relativity. The resulting theories are called *supergravity theories*, and we will have more to say about them later in this thesis.

Our discussion of supersymmetry is not complete without mentioning the fact that supersymmetry has not yet been observed in nature. It therefore remains a hypothetical aspect of theories describing our universe. The way to justify supersymmetry as a realistic symmetry is to assume that it is *broken* at the energy scales that we can access. The phenomenological incentive for supersymmetry, together with the fact supersymmetric theories are very well-understood due to their mathematical structure, provides enough motivation to continue the study of supersymmetric theories.

2.1.3 Superstring Theories

We now resume our discussion of string theory. In the beginning of this section we mentioned that we can cure the problems with the bosonic string by including fermions and making the worldsheet theory supersymmetric. The resulting theory describes *superstrings*. There are several consistent ways to implement the fermions in a supersymmetric way, depending on a handful of discrete choices one can make. However, in any case, the resulting superstring theory is only consistent in exactly *ten* dimensions.

Fact 2.1.1. There are five superstring theories in ten dimensions. These are known as

$$\text{Type I}; \quad \text{Type IIA}; \quad \text{Type IIB}; \quad \text{Heterotic } \mathrm{E}_8 \times \mathrm{E}_8; \quad \text{Heterotic } \mathrm{SO}(32).$$

To be precise, these five theories arise from demanding supersymmetry in spacetime, in addition to supersymmetry on the worldsheet. For our purposes it is not important how exactly the various superstring theories are constructed, and we refer to [23] for a detailed explanation. In this work, we focus on arguably the most prominent superstring theory, namely the Type IIB superstring[1].

As in the case of the string, the superstring comes with a spectrum, and the states in the spectrum may be interpreted as spacetime fields. Crucially, the spectrum no longer contains a tachyon, which was one of the objectives that the inclusion of supersymmetry was supposed to achieve. For later reference we write down the massless bosonic fields that appear from the spectrum of the IIB string:

(i) a scalar field ϕ called the *dilaton*;

(ii) a 2-form gauge field B_2;

(iii) a metric tensor $g_{\mu\nu}$ called the *graviton*;

(iv) for $p = 0, 2, 4$, a p-form gauge field C_p. The field C_0 is called the *axion*.

[1]It is worth noting that these five theories do not stand completely on their own, but are related by a network of *dualities* [23]. Though this is an important fact, it will not play a big role in this thesis.

These massless fields are the effective low-energy bosonic degrees of freedom of the Type IIB string, and their dynamics is governed[2] by the action

$$S_{\mathrm{IIB}} \sim \int_{M_{10}} \mathcal{R} * 1 - \frac{1}{2}\frac{d\tau \wedge *d\bar{\tau}}{(\operatorname{Im}\tau)^2} + \frac{G_3 \wedge \bar{G}_3}{\operatorname{Im}\tau} + \frac{1}{2}\tilde{F}_5 \wedge *\tilde{F}_5 + C_4 \wedge H_3 \wedge F_3. \tag{2.2}$$

Here $\mathcal{R}$ is the Ricci scalar of the spacetime metric $g_{\mu\nu}$, and τ is a complex scalar field taking values in the upper-half plane $\mathbb{H}$ defined by $\tau = C_0 + i\,e^{-\phi}$. It combines the axion and the dilaton, and is therefore known as the axio-dilaton. Furthermore, the action S_{IIB} is formulated in terms of the field strengths

$$G_3 = F_3 - \tau H_3, \quad \tilde{F}_5 = F_5 - \frac{1}{2}C_2 \wedge H_3 + \frac{1}{2}B_2 \wedge F_3, \tag{2.3}$$

where $H_3 = dB_2$ and $F_{p+1} = dC_p$. The reason to write the action in terms of these combinations of fields is that it makes a certain invariance under an action[3] of $\mathrm{SL}(2,\mathbb{Z})$ manifest. This symmetry can be exploited in an elegant way, as discussed below.

2.1.4 F-theory

There exists another important formulation of superstring theory, called *F-theory*, which may be interpreted as a geometrization of the Type IIB string [27]. The original motivation for such a geometrization stems from the $\mathrm{SL}(2,\mathbb{Z})$-invariance of the action S_{IIB}. What makes this symmetry special is that it is believed to be a non-perturbative symmetry of the full Type IIB string theory. Due to this symmetry, the axio-dilaton τ may be interpreted[4] the period ratio $\tau = B/A \in \mathbb{H}$ of an auxiliary torus (as in Example 1.3.1). Combining this auxiliary torus with the ten-dimensional spacetime, one is led to a twelve-dimensional theory. Since two of the dimensions are interpreted as a torus, the twelve-dimensional space must be *elliptically fibered*, meaning that it is the total space of a bundle whose fibers are (possibly singular) elliptic curves.

Through its relations with algebraic geometry and inclusion of non-perturbative effects, this twelve-dimensional F-theory has become a powerful and elegant tool to analyze questions in Type IIB string theory [28, 29]. Properly setting up F-theory requires a formal detour through *M-theory* and various dualities, and for the purposes of this thesis, an understanding of such details is not required. It will be sufficient to view F-theory as a twelve-dimensional description of Type IIB, and we refer to [28] for a detailed discussion.

Regardless of whether we now choose to work with any of the five ten-dimensional superstring theories or with the twelve-dimensional F-theory, making contact with our four-dimensional universe requires us to make sense of the extra dimensions in string theory. This is the task of the next section.

[2]To correctly extract the dynamics from this action, the self-duality condition $*\tilde{F}_5 = \tilde{F}_5$ must be imposed manually in addition to the equations of motion specified by S_{IIB}.

[3]More precisely, an element of $\mathrm{SL}(2,\mathbb{Z})$ acts on τ by modular transformations, and on the pair (F_3, H_3) in the fundamental representation.

[4]Recall that the period ratio $\tau \in \mathbb{H}$ also carries an $\mathrm{SL}(2,\mathbb{Z})$-action due to monodromy.

2.2 Compactifications of String Theory

As alluded to in the introduction, the apparent conflict in the number of dimensions observed in nature and predicted by superstring theory can be resolved by the idea of *compactification*. In this proposed solution, the spacetime M_d is assumed to factor as

$$M_d = M_4 \times X_D, \tag{2.4}$$

where M_4 is a four-dimensional Lorentzian manifold representing the dimensions we observe, and X_D is a compact manifold called the *internal space*, whose size is sufficiently small to escape detection at the length scales we are currently able to probe. Though X_D is supposedly undetectable, the geometry of X_D has enormous implications for the theory that is perceived on M_4. The ten-dimensional theory is said to be *compactified on* X_D. For phenomenological purposes, we are mostly interested in compactifications where M_4 is a maximally symmetric space, i.e. Minkowski space, de Sitter space or anti-de Sitter space.

The idea of compactification appears to be tailored towards the problem of hiding extra dimensions, but in fact it has already been proposed long before by Kaluza and Klein as a way to unify gravity and electrodynamics. They studied a four-dimensional theory obtained by compactifying a five-dimensional theory on a circle. Since circle compactification already illustrates many of the essential features of compactifications, it is worthwhile to discuss it in some detail.

2.2.1 Circle Compactification

Instead of looking at a compactification of string theory, we will focus on a field theory in this subsection. We will comment on this distinction later in Section 2.6. Consider a free massless scalar field ϕ defined on a five-dimensional manifold of the form $M_5 = M_4 \times S^1$. We denote the coordinates on M_5 by $x^{\hat{\mu}} = (x^\mu, \theta)$, with $\theta \sim \theta + 2\pi$. The metric on M_5 is assumed to factor as

$$g_{\hat{\mu}\hat{\nu}}\, dx^{\hat{\mu}}\, dx^{\hat{\nu}} = \eta_{\mu\nu}\, dx^\mu\, dx^\nu + R^2 d\theta^2, \tag{2.5}$$

where R is the radius of the circle. The action of such a theory is

$$S[\phi] = -\frac{1}{2}\int_{M_5} d^5x\, \sqrt{-g}\, g^{\hat{\mu}\hat{\nu}}\, \partial_{\hat{\mu}}\phi\, \partial_{\hat{\nu}}\phi. \tag{2.6}$$

The equation of motion specified by this action is

$$\partial_\mu \partial^\mu \phi + \frac{1}{R^2}\partial_\theta^2 \phi = 0. \tag{2.7}$$

Let us now see how this theory leads to an emergent theory in four dimensions. The compactness of the circle factor of M_5 implies that any field $\phi(x, \theta)$ admits a discrete mode expansion

$$\phi(x, \theta) = \sum_{n\in\mathbb{Z}} \phi_n(x)\, e^{in\theta}, \tag{2.8}$$

and upon inserting this expansion into the equation of motion (2.7) we find that each ϕ_n satisfies

$$\left(\partial_\mu \partial^\mu - \frac{n^2}{R^2}\right)\phi_n(x) = 0. \tag{2.9}$$

In this context, the expansion in equation (2.8) is called the Kaluza-Klein (KK) expansion and the fields ϕ_n are called the KK modes. Equation 2.9 shows that each KK mode ϕ_n obeys precisely the equation of motion of a massive scalar on M_4 with mass n/R. This shows that the physics in four dimensions is determined by the geometry of the compactification, in this case the radius R.

Now suppose that we couple the theory to gravity; in other words, we allow the metric $g_{\hat{\mu}\hat{\nu}}$ to become a dynamical field of the theory. The dynamics is generated by the Einstein-Hilbert action

$$S_{\text{EH}}[g] \sim \int_{M_5} d^5x \sqrt{-g}\,\mathcal{R}_{(5)}, \tag{2.10}$$

where $\mathcal{R}_{(5)}$ is the five-dimensional Ricci scalar. Since the radius R appears as a component of the metric (equation (2.5)), it becomes dynamical as well. This implies in particular that R is now a field whose value may vary on the four-dimensional spacetime M_4, as illustrated below in Figure 2.2.

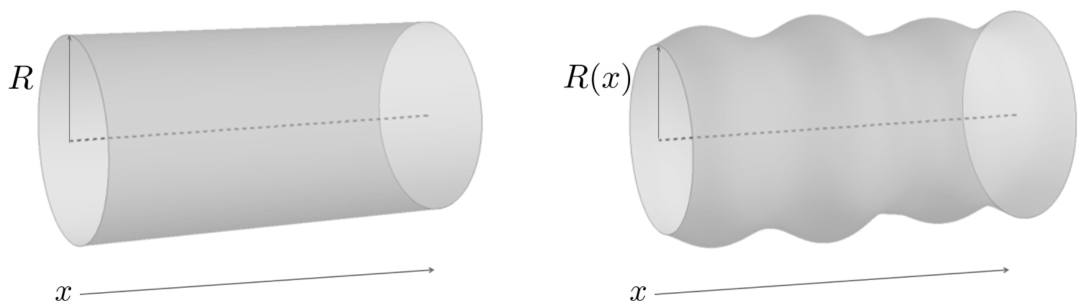


Figure 2.2: Compactification on a circle S^1. The space $M_5 = M_4 \times S^1$ is depicted as a cylinder, in which the non-compact spacetime M_4 is represented by a line with coordinate x. On the left, the radius R is fixed. On the right, gravity is turned on, allowing the radius R to fluctuate as a function of the spacetime coordinate x.

The equation of motion for the field $R(x)$ is hidden in the action in (2.10). Expressing the five-dimensional Ricci scalar $\mathcal{R}^{(5)}$ in terms of the four-dimensional Ricci scalar $\mathcal{R}^{(4)}$ and the radius R, assuming the metric ansatz of equation (2.5), one finds that R is a *massless* free scalar field taking values in the space $\mathbb{R}_{>0}$. It is our first example of a *modulus*: a parameter encoding the internal geometry of the compactification which appears as a free massless scalar field in the lower-dimensional theory. Since moduli have no mass, they should be visible at any energy scale, in particular at the energy scale of physics around us.

However, no such massless scalar fields are observed in experiment, seemingly leading to a direct contradiction with the idea of compactification. To make compactifications realistic, we therefore have to remove these moduli, for instance by incorporating a mechanism that makes them acquire a mass. This issue is called *moduli stabilization*, and it is a central theme of research on string compactifications. We will revisit this idea later in this section.

To conclude this discussion of circle compactifications, we record several general features that we learn from this example.

Observation 2.2.1. Compactifications are accompanied by the following phenomena:

(i) the geometry of the internal space controls the physics of the compactified theory;

(ii) compactifications give rise to moduli parametrizing the geometry of the internal space;

(iii) in order for the compactification to be realistic, all moduli must be stabilized.

2.2.2 Calabi-Yau Compactification

In the case of a one-dimensional compactification, the geometry is nearly uniquely determined to be a circle. For the case at hand, where six dimensions have to be compactified, there is an enormous range of possibilities for the geometry of the internal space X. The desired presence of supersymmetry improves the situation, since supersymmetry strongly restricts the permissible internal geometries. Demanding $\mathcal{N} = 1$ supersymmetry in four dimensions, it can be argued that the internal space X must admit the existence of a certain spinor, which constrains the holonomy group of X. The spaces that satisfy this constraint are the *Calabi-Yau manifolds* (see e.g. [23, p. 441]).

Definition 2.2.2. A *Calabi-Yau manifold* is a compact Kähler manifold X of dimension n whose canonical bundle $\Omega^n(X)$ is trivial.

Calabi-Yau manifolds have several interesting properties, most of which we state here without proof (we again refer to [23] for a discussion).

Proposition 2.2.3. *Let X be a Calabi Yau n-fold. Then*

(i) X admits a global non-vanishing holomorphic $(n, 0)$-form Ω;

(ii) the first Chern class $c_1(TX)$ vanishes;

(iii) the holonomy group of X is $\mathrm{SU}(n)$;

(iv) X admits a Ricci-flat Kähler metric, and this metric is uniquely determined by the Kähler class of X.

The first of these properties is an immediate consequence of the triviality of $\Omega^n(X)$. Property (iii) plays an important role in arguing why one requires a Calabi-Yau for $\mathcal{N} = 1$ supersymmetry in four dimensions, and can actually be used as a defining property for Calabi-Yaus. The Ricci-flat metric that follows from (iv) is crucial for studying the deformation theory of a Calabi-Yau, as we will see below in Section 2.3. It is worth noting that Ricci-flatness of the metric,

$$\mathcal{R}_{\mu\nu} = 0, \tag{2.11}$$

is precisely the Einstein equation in vacuum[5].

[5] This equation also emerges as a consistency condition from the cancellation of the conformal anomaly, at least to first order in perturbation theory.

We can now state how Calabi-Yau manifolds can be used to compactify the ten-dimensional string theories from Fact 2.1.1 to an $\mathcal{N} = 1$ supersymmetric theories in four dimensions [29].

Fact 2.2.4. The following compactifications give $\mathcal{N} = 1$ supersymmetry in four dimensions.

- Type I or heterotic string theory on a Calabi-Yau threefold;
- Type IIA or Type IIB string theory on a Calabi-Yau orientifold;
- F-theory on an elliptically fibered Calabi-Yau fourfold.

Let us briefly comment on the notion of *orientifold*. This is a type of quotient which operates partly on the level of geometry of X and partly on the level of the worldsheet of the string. More precisely, it typically involves a quotient by an involution of X paired with an operator that flips the orientation of the string. The setting that results from this procedure is a string theory on a Calabi-Yau orientifold, and a consequence of taking the quotient is that half of the spectrum is removed, reducing the $\mathcal{N} = 2$ supersymmetry of Type II to $\mathcal{N} = 1$.

An enormous number of Calabi-Yau manifolds is known, and at present string theory provides no mechanism that selects one choice of Calabi-Yau over another. We therefore proceed with a generic Calabi-Yau. As illustrated with the circle, performing a compactification causes moduli to emerge. This leads us to consider the moduli of Calabi-Yau manifolds.

2.3 The Calabi-Yau Moduli Space

To see how moduli arise in Calabi-Yau compactifications, we require an understanding of the Calabi-Yau moduli space. We begin by studying the geometry of this moduli space locally by looking at deformations of the Kähler metric, following [30]. We then proceed to discuss global aspects of the moduli space geometry and connect this to the setting of Chapter 1. In both subsections we focus on the essential aspects needed for later chapters in this work, and some details are omitted.

2.3.1 Local Moduli: Deformations

As already seen in the circle compactification example, moduli emerge as parameters encoding the geometry[6] of the compactification manifold X. More precisely, they appear as deformation parameters of the metric. Let us therefore start by considering infinitesimal deformations δg of the Calabi-Yau metric g. Two conditions have to be imposed on δg. Firstly, the deformed metric has to remain Ricci-flat so that $(X, g + \delta g)$ is still Calabi-Yau, leading to the condition

$$\mathcal{R}_{\mu\nu}(g + \delta g) = 0. \tag{2.12}$$

Secondly, we require that the transformation $g \mapsto g + \delta g$ is not generated by a diffeomorphism, since then (X, g) and $(X, g + \delta g)$ should be considered isomorphic. It turns out that this can be done by imposing the coordinate condition $\nabla^\mu \delta g_{\mu\nu} = 0$.

[6]There are also other types of moduli, arising for instance from the presence of D-branes. We will not consider these moduli in this thesis.

Expanding equation (2.12) to first order in δg and assuming this coordinate condition, one finds the so-called Lichnerowicz equation, given by

$$\nabla^\rho \nabla_\rho \delta g_{\mu\nu} + 2\mathcal{R}_\mu{}^\rho{}_\nu{}^\sigma \delta g_{\rho\sigma} = 0. \tag{2.13}$$

This equation simplifies by virtue of the fact that X is Kähler. The most important simplification is that it decouples into two seperate equations for 'pure' and 'mixed' deformations. By this we mean that, when splitting an index μ into holomorphic and anti-holomorphic indices as $\mu = (i, \bar{\imath})$, the metric deformation $\delta g_{\mu\nu}$ decomposes into δg_{ij}, $\delta g_{i\bar{\jmath}}$, and complex conjugate components. The Lichnerowicz equations for δg_{ij} and $\delta g_{i\bar{\jmath}}$ are then independent as a result of the vanishing of certain components of the Riemann tensor on a Kähler manifold. Let us now discuss these deformations separately.

Consider a mixed deformation of the form $\delta g_{i\bar{\jmath}}$. From the components $\delta g_{i\bar{\jmath}}$ we can construct a real $(1,1)$-form

$$\delta\omega = i\, \delta g_{i\bar{\jmath}}\, d\zeta^i \wedge d\bar{\zeta}^{\bar{\jmath}}, \tag{2.14}$$

where the ζ^i denote complex coordinates on X. Since the Kähler form associated to the original metric $g_{\mu\nu}$ is given by

$$\omega = i\, g_{i\bar{\jmath}}\, d\zeta^i \wedge d\bar{\zeta}^{\bar{\jmath}}, \tag{2.15}$$

we find that the deformation of type $\delta g_{i\bar{\jmath}}$ correspond to deformations of the Kähler structure of X. With some work, it can then be shown that the Lichnerowicz equation is equivalent to

$$\Delta\, \delta\omega = 0, \tag{2.16}$$

whose solutions are the harmonic forms encountered in Chapter 1. In view of the Hodge theorem (1.1.2) relating harmonic forms to cohomology classes, we thus conclude the following.

> **Observation 2.3.1.** Infinitesimal deformations of the Kähler structure of X are in one-to-one correspondence with elements in $H^2(X;\mathbb{R}) \cap H^{1,1}(X)$.

Any real variables parametrizing the $\delta g_{i\bar{\jmath}}$ are referred to as *Kähler moduli.*

Let us now proceed with deformations of the form δg_{ij} (and its conjugate $\delta g_{\bar{\imath}\bar{\jmath}}$). What is the interpretation of these deformations? Due to their pure index structure, the metric $g + \delta g$ appears a priori not to be Hermitian. However, this notion of Hermitian depends on the distinction between holomorphic and anti-holomorphic indices, which is determined by the complex structure of X. The metric $g + \delta g$ can thus be made Hermitian by a change of complex structure, and we are led to interpret δg_{ij} and $\delta g_{\bar{\imath}\bar{\jmath}}$ as deformations of the complex structure.

These complex structure deformations can be compared to certain (p,q)-forms in a similar way as above, although the comparison is slightly less direct. Let us illustrate this explicitly for the case of Calabi-Yau threefolds. Let

$$\Omega = \Omega_{ijk}\, d\zeta^i \wedge d\zeta^j \wedge d\zeta^k \tag{2.17}$$

be the (up to rescaling) unique $(3,0)$-form. From a pure metric deformation we construct in a one-to-one manner a $(2,1)$-form

$$\delta\eta = \Omega_{ij}{}^{\bar{l}}\, \delta g_{\bar{l}\bar{k}}\, d\zeta^i \wedge d\zeta^j \wedge d\bar{\zeta}^{\bar{k}}. \tag{2.18}$$

With some work it can be shown that the Lichnerowicz equation for pure deformations is equivalent to

$$\Delta\, \delta\eta = 0. \tag{2.19}$$

Again, we see harmonic forms appear as solutions. We are led to conclude the following.

> **Observation 2.3.2.** Infinitesimal deformations of the complex structure of a Calabi-Yau threefold X are in one-to-one correspondence with elements in $H^{2,1}(X)$.

A similar analysis can be performed for Calabi-Yau fourfolds, and the complex structure deformations are then identified with $H^{3,1}(X)$. Any complex variables parametrizing the $\delta g_{i\bar{\jmath}}$ are referred to as *complex structure moduli*. These moduli will be the most important in this thesis.

In this subsection we have mostly followed the physics literature on infinitesimal deformations of Calabi-Yaus, and we note that a more precise formalism for such deformations is *Kodaira-Spencer theory*; this approach is taken in e.g. [10]. The next step is to consider the global description of moduli.

2.3.2 Global Moduli: Weil-Petersson Geometry

We have now seen that the local deformation theory of a Calabi-Yau metric has a nice structure, but a priori it is not at all clear whether the infinitesimal deformations 'integrate' to global deformations. However, fortunately it turns out that in the Calabi-Yau setting this is indeed the case, and the global deformations can be assembled into a *moduli space* [31, 32]. This moduli space has several important properties, which we now discuss.

The local considerations of the previous subsection tell us that, at least locally, the moduli space factors into a Kähler moduli space and a complex structure moduli space. For the threefold case, these are of real dimension $h^{1,1}$ and $2h^{2,1}$, respectively. When studying compactifications of string theory, it is customary to study these separately. In this thesis, we will exclusively focus on the complex structure moduli space. The main motivation for this is that there the tools of Chapter 1 are available, allowing for a detailed mathematical understanding of complex structure moduli[7].

Before abandoning Kähler moduli, we note that they are not entirely unrelated to complex structure moduli. The phenomenon of *mirror symmetry*, which is conjectured to relate any Calabi-Yau X to a certain 'mirror' Calabi-Yau $\hat{X}$, links the Kähler moduli of X to the complex structure moduli of $\hat{X}$ and vice versa [33]. One sign of this is the fact that, in the threefold case,

$$h^{1,1}(X) = h^{2,1}(\hat{X}) \quad \text{and} \quad h^{2,1}(X) = h^{1,1}(\hat{X}). \tag{2.20}$$

This connection further justifies the emphasis on complex structure moduli.

We now turn to a discussion of the complex structure moduli space $\mathcal{M}$. This space can be viewed as a geometric object classifying[8] all the distinct complex structures on a fixed Calabi-Yau manifold X. The first significant result on the geometry of $\mathcal{M}$ is the following theorem [35].

[7] Another important motivation is that, in some sense, the complex structure moduli is 'non-perturbative', in the sense that its geometry does not receive corrections from string perturbation theory.

[8] Formally, $\mathcal{M}$ is a 'coarse' moduli space, and it can be constucted using geometric invariant theory as a quotient of a Hilbert scheme [34]. For the purposes of this thesis we will not concern ourselves with such details, and it suffices to have an intuitive notion of moduli space.

Theorem 2.3.3. *Let X be a polarized Calabi-Yau manifold. Then the complex structure moduli space $\mathcal{M}$ of X is a quasi-projective variety.*

The word *polarized* refers to the integrality of the Kähler form which gives rise to a polarization of the Hodge structure; it is a technical condition which is always satisfied in our case. As seen in Chapter 1, it is customary to resolve possible singularities of $\mathcal{M}$ using Hironaka's Theorem [19] and work with a smooth moduli space. From this point onward we will therefore assume that $\mathcal{M}$ is smooth. This brings us to the setting of Section 1.4.

Observation 2.3.4. The complex structure deformations of a Calabi-Yau manifold X give rise to a family of Calabi-Yau manifolds $\mathcal{X} \to \mathcal{M}$, in the sense of Definition 1.4.1.

In fact, more is true, and in [35] it is shown that the moduli space $\mathcal{M}$ is a subvariety of a projective variety $\overline{\mathcal{M}}$, in such a way that the family $\mathcal{X} \to \mathcal{M}$ extends to a family $\overline{\mathcal{X}} \to \overline{\mathcal{M}}$. The fibers of the locus $\mathcal{M}_{\text{sing}} := \overline{\mathcal{M}} \setminus \mathcal{M}$ correspond to singular Calabi-Yaus. This is precisely the setting we worked with in Chapter 1.5, and we thus find the following.

Observation 2.3.5. The complex structure moduli space $\mathcal{M}$ of a Calabi-Yau manifold X carries a variation of Hodge structure.

This observation enables the use of the Hodge-theoretic techniques discussed in the previous chapter and will play a central role in drawing physical conclusions from powerful mathematical results.

It is an essential fact that the complex moduli space $\mathcal{M}$ carries a natural Kähler metric $\mathcal{K}_{i\bar{\jmath}}$. In fact, this metric is *physical* in the sense that it appears in the effective action of the four-dimensional theory as the coupling function of the complex structure moduli [30]. More precisely, the four-dimensional supergravity action arising from a Calabi-Yau compactification contains a term

$$\int_{M_4} d^4x\, \sqrt{-g}\, \mathcal{K}_{i\bar{\jmath}}\, \partial_\mu z^i\, \partial^\mu \bar{z}^{\bar{\jmath}}, \tag{2.21}$$

where the z^i are coordinates on $\mathcal{M}$ interpreted as fields on spacetime M_4. The metric $\mathcal{K}_{i\bar{\jmath}}$ is called the Weil-Petersson metric [34], and its Kähler potential is given by

$$\mathcal{K}(z,\bar{z}) = -\log i^n \int_X \Omega(z) \wedge \overline{\Omega(z)}. \tag{2.22}$$

Here $\Omega(z)$ is a smoothly varying choice of holomorphic $(n,0)$ form on the fibers X_z. In the language of variations of Hodge structure, the cohomology classes $[\Omega(z)]$ define a section of the holomorphic line bundle $\mathcal{F}^n = \mathcal{H}^{n,0}$.

The formula for the Kähler potential does not specify which section of $\mathcal{F}^n$ we should take, and we are free to rescale $\Omega(z)$ by any holomorphic function on $\mathcal{M}$. This leads to a *Kähler transformation* of the Kähler potential,

$$\Omega(z) \mapsto f(z)\,\Omega(z); \quad \mathcal{K}(z,\bar{z}) \mapsto \mathcal{K}(z,\bar{z}) - \log f(z) - \log \overline{f(z)}, \tag{2.23}$$

and the dependence on $f(z)$ disappears when computing the Kähler metric

$$\mathcal{K}_{i\bar{\jmath}} = \partial_i \overline{\partial}_{\bar{\jmath}} \mathcal{K}. \tag{2.24}$$

We will see this metric appear later in Chapter 5.

2.4 String Vacua

As foreshadowed in the introduction, a challenge of string theory compactification is that it produces an enormous amount of *string vacua*. There is no fixed definition of string vacuum, but roughly speaking it refers to a background through which the string can consistently propagate. Such a background amounts to a consistent choice of internal geometry and background values for the physical objects in the theory, including for instance *branes* and *fluxes*. The abundance of string vacua leads us to the first appearance of a *landscape* of string vacua, as featured in the title of this work. In this section we introduce a prominent class of string vacua named *flux vacua*. Two comprehensive reviews on this topic are [29] and [36].

2.4.1 Moduli Stabilization with Fluxes

In order to produce realistic string vacua, we should first address the problem of moduli stabilization introduced in Section 2.2. This subsection is devoted to an important mechanism that can be used to stabilize complex structure moduli, assuming that we have made a choice of Calabi-Yau manifold X to compactify on. Recall from Section 2.1 that from the excitations of the superstring, several fields emerge in the low-energy description. So far we have mostly ignored these, but they actually have important phenomenological implications. Giving these fields a background value is known as *turning on fluxes*, in analogy with the fluxes of field strengths in electromagnetism. The idea is that these fluxes generate an energy potential for the moduli, which forces them towards a minimum of the potential. Let us now make this more precise.

We focus on the setting of F-theory to streamline the discussion. The field of interest is the 4-form field strength G_4, which is closely related to the field strength G_3 which appeared earlier in Section 2.1. Setting G_4 to a non-zero value, a scalar potential for the complex structure moduli is induced, and it can be shown to take the form[9]

$$V(z, G_4) = \int_X G_4 \wedge *G_4 - \int_X G_4 \wedge G_4. \tag{2.25}$$

The dependence on the complex structure moduli, generically denoted by z, is hidden in the Hodge star, as it is defined in terms of the metric on X. Before we are able to continue with this scalar potential, it is necessary to discuss a number of crucial physical constraints on G_4 (these are summarized in e.g. [38]). We will not attempt to derive any of them, but we do briefly indicate their physical origin.

Although we spoke rather loosely about giving G_4 a background value, we cannot do so arbitrarily. Firstly, since G_4 is a field strength, it has its own dynamics. Imposing the equation of motion for G_4, one finds that G_4 must be a harmonic form. In view of the Hodge theorem discussed in Chapter 1, we may therefore identify G_4 with a cohomology class in $H^4(X;\mathbb{R})$. The second condition comes from considering the theory at the quantum level. The *Dirac quantization* condition implies that the fluxes must be discrete, and viewing the flux as a cohomology class we may conclude[11] that $G_4 \in H^4(X;\mathbb{Z})$ [39].

[9]The more common way of displaying the scalar potential uses the $\mathcal{N}=1$ supergravity formulation, where the potential is written in terms of the Kähler potential $\mathcal{K}$ and the so-called Gukov-Vafa-Witten superpotential [37][10]. Since this form of the action does not appear later in the thesis, we do not present it.

[11]There is a subtlety here, and to be completely precise, G_4 has to be shifted by a certain characteristic class before it takes values in $H^4(X;\mathbb{Z})$. This subtlety has no impact for the arguments in this work.

The third physical requirement on G_4 comes from consistency of having fluxes on a compact space. The intuitive physical picture to have in mind is as follows. A charged object emits flux lines, and inside a compact space these cannot go to infinity and must therefore end on another (oppositely) charged object. This forces the total charge on a compact space to be neutral. The precise statement for the G_4-flux is the so-called *tadpole condition*, and it reads [40]

$$\frac{1}{2}\int G_4 \wedge G_4 = -N_{\mathrm{D3}} + \frac{\chi(X)}{24}. \tag{2.26}$$

Here $\chi(X)$ is the Euler characteristic of X, and N_{D3} is the number of (spacetime-filling) D3-branes in the theory.

Next, it can be shown that, assuming certain conditions on the Kähler moduli, G_4 has to be primitive, i.e.

$$\omega \wedge G_4 = 0, \tag{2.27}$$

where ω is the Kähler form of X [41]. The final constraint that we mention comes from imposing supersymmetry. In order to obtain a supersymmetric string vacuum, it turns out that the (p,q)-type of G_4 is restricted to be $(2,2)$.

We summarize the physical conditions on the G_4-flux in the following observation.

Observation 2.4.1. A physically permissible G_4-flux in F-theory on a Calabi-Yau fourfold X is an integral cohomology class $G_4 \in H^4_{\mathrm{p}}(X;\mathbb{Z})$ satisfying

$$\int_X G_4 \wedge G_4 = t_0 \tag{2.28}$$

for a rational number t_0 determined by the compactification. If supersymmetry is imposed, then the flux takes values in $H^4_{\mathrm{p}}(X;\mathbb{Z}) \cap H^{2,2}(X)$.

Curiously, note that in the supersymmetric case, G_4 is a Hodge class, as discussed in Section 1.6. We will return to this remarkable fact later.

We have now seen that it is possible allow for non-zero values of G_4, thereby generating a scalar potential for the complex structure moduli which in principle has the ability to stabilize moduli. In order to obtain string vacua, the next step is to analyze the minima of the potential. This is the task of the next section.

Before we proceed, it should be remarked that it is not universally agreed that flux compactifications are a good mechanism for moduli stabilization, and that this mechanism is currently under heavy scrutiny. Most notably, a *Tadpole Conjecture* was recently proposed which challenges the idea that fluxes satisfying the tadpole condition of equation (2.26) are sufficient to stabilize *all* complex structure moduli [40]. The evidence in favor of this conjecture is growing, and includes arguments from Hodge theory [42]. Nonetheless, string vacua obtained from flux compactifications remain an important class of vacua, and we proceed to study them in the next subsection.

2.4.2 Flux Vacua

The observation that the complex structure moduli space comes with a variation of Hodge structure allows us to rephrase the scalar potential generated by a G_4-flux. We viewed G_4 as an element of $H^4_{\mathrm{p}}(X;\mathbb{Z})$, but adopting the point of view that our Calabi-Yau comes in a deformation family, it should be viewed as an element of $H^4_{\mathrm{p}}(X_z;\mathbb{Z})$ where $z \in \mathcal{M}$ is a given point in the complex structure moduli space. Recall that the spaces $H^4_{\mathrm{p}}(X_z;\mathbb{Z})$ are assembled into the local system $\mathcal{H}_{\mathbb{Z}}$ on $\mathcal{M}$, and that this local system comes equipped with a pairing $q : \mathcal{H}_{\mathbb{Z}} \otimes \mathcal{H}_{\mathbb{Z}} \to \mathbb{Z}$ representing the polarization form. Associated to this local system, we defined the Hodge bundle $\mathcal{H}_{\mathbb{C}} \to \mathcal{M}$ which has fibers $H^4_{\mathrm{p}}(X_z;\mathbb{C})$, and the local system may be viewed as a lattice subbundle of $\mathcal{H}_{\mathbb{C}}$. On a fiber $(\mathcal{H}_{\mathbb{C}})_z$, the pairing q_z is simply given by integration:

$$q_z(u,v) = \int_{X_z} u \wedge v. \tag{2.29}$$

Recalling that the Hodge star is given by the Weil operator C when interpreted as an operator on cohomology, we conclude the following.

Observation 2.4.2. The flux-induced scalar potential may be interpreted as the map

$$\begin{aligned} V &: \mathcal{H}_{\mathbb{C}} \to \mathbb{C}, \\ (z, G_4) &\mapsto q_z(G_4, C_z G_4) - q_z(G_4, G_4). \end{aligned} \tag{2.30}$$

Here (z, G_4) is now interpreted as an element in the vector bundle $\mathcal{H}_{\mathbb{C}}$. To obtain the physical scalar potential, one restricts G_4 according to the constraints of Observation 2.4.1.

In order to analyze the minima of V, let us first further investigate its structure. For a given fiber X in the family $\mathcal{X}$, the primitive cohomology splits as

$$H^4_{\mathrm{p}}(X;\mathbb{C}) = H^{4,0} \oplus H^{3,1} \oplus H^{2,2} \oplus H^{1,3} \oplus H^{0,4}. \tag{2.31}$$

The Weil operator C is defined by $C|_{H^{p,q}} = i^{p-q}\mathrm{id}_{H^{p,q}}$, and this simply becomes

$$Cv = \begin{cases} +v & \text{for } v \in H^{4,0} \oplus H^{2,2} \oplus H^{0,4}, \\ -v & \text{for } v \in H^{3,1} \oplus H^{1,3}. \end{cases} \tag{2.32}$$

An element v is called *self-dual* if $Cv = v$ and *anti-self-dual* if $Cv = -v$. Any $v \in H^4_{\mathrm{p}}(X;\mathbb{C})$ splits into a self-dual and anti-self-dual part as

$$v = v^+ + v^-, \quad Cv^\pm = \pm v^\pm, \tag{2.33}$$

where $v^+ = (v + Cv)/2$ and $v^- = (v - Cv)/2$. In this notation, we have the following.

Lemma 2.4.3. *Restricting G_4 to be real, the flux-induced scalar potential is positive-definite.*

Proof. First we rewrite V via a simple computation:

$$\begin{aligned} V(z, G_4) &= q_z(G_4, C_z G_4) - q_z(G_4, G_4) = q_z(G_4, C_z G_4 - G_4) \\ &= q_z(G_4, C_z(G_4 - C_z G_4)) = 2q_z(G_4, C_z G_4^-) \\ &= 2q_z(G_4^-, C_z G_4^-). \end{aligned}$$

The last line uses the bilinear relations from Theorem 1.1.4 to conclude that self-dual and anti-self-dual classes are orthogonal to each other. Assuming that G_4 is real, we see that this is the Hodge norm squared of the anti-self-dual part G_4^-, which is positive definite. □

This proof immediately reveals what the minima of the scalar potential are: they are precisely the self-dual classes. If one now imposes the physical conditions on the G_4-flux, then minima of the potential in principle correspond to valid string vacua. With this in mind we give the following definition.

Definition 2.4.4. Let X be a Calabi-Yau fourfold with complex structure moduli space $\mathcal{M}$, let $\mathcal{H}_{\mathbb{C}} \to \mathcal{M}$ be the Hodge bundle, and let t_0 be a rational number. The *locus of flux vacua* is the subset of $\mathcal{H}_{\mathbb{C}}$ defined by

$$\mathscr{F}_X(t_0) := \big\{(z, G_4) \in \mathcal{H}_{\mathbb{C}} \,\big|\, G_4 \text{ is integral, self-dual, and } q_z(G_4, G_4) = t_0\big\}. \tag{2.34}$$

Note that in physical settings, the 'tadpole constant' t_0 is not arbitrary but given by

$$t_0 = -2N_{\mathrm{D3}} + \frac{\chi(X)}{12}. \tag{2.35}$$

Leaving the number of D3-branes arbitrary, we could therefore also leave t_0 out of the definition and consider the locus

$$\mathscr{F}_X := \big\{(z, G_4) \in \mathcal{H}_{\mathbb{C}} \,\big|\, G_4 \text{ is integral, self-dual, and } q_z(G_4, G_4) \leq \chi(X)/12\big\}. \tag{2.36}$$

So far we have worked with a fixed Calabi-Yau fourfold X, but there appears to be no natural choice of Calabi-Yau fourfold to compactify on. Indeed, there is as of yet no known mechanism that selects one Calabi-Yau over the other. It is known that there exists a huge number of topologically distinct Calabi-Yaus, and it remains an open problem of whether there are finitely or infinitely many[12].

Formally, we could now attempt to define the *landscape of F-theory flux vacua* as the set

$$\mathscr{F} = \coprod_{\text{CYs } X} \mathscr{F}_X. \tag{2.37}$$

To understand how our universe may arise from string theory, it is essential to gain an understanding of $\mathscr{F}$. However, without any clear handle on a classification of Calabi-Yau fourfolds, it is hardly sensible to study this object. We therefore settle for a more modest goal, and study the individual $\mathscr{F}_X(t_0)$ or $\mathscr{F}_X$ for a fixed X. This already turns out to be a highly non-trivial task. One of the most important questions surrounding the locus of flux vacua is whether the number of flux vacua is finite, and a definite answer to this question has been given only very recently [1, 4].

[12] It can however be argued that there are only finitely many topological types that can be consistent with our universe [5]. This argument is based on Cheeger's theorem [43] and assumes upper bounds on the vacuum energy and the volume of the internal space, and a lower bound on the mass scale of the lightest KK tower.

The answer found in [4] is that there is indeed a certain sense in which $\mathscr{F}_X(t_0)$ is finite. The more precise statement is that it has a 'tame geometry'; this idea will be discussed at length in the upcoming chapters. The proof will be explored in Chapter 5, but let us already comment on what makes finiteness non-trivial. Suppose for the moment that we fix $z \in \mathcal{M}$ and $t_0 \in \mathbb{Q}$. We now ask: *are there finitely many integral self-dual* $G_4 \in (\mathcal{H}_\mathbb{C})_z$ *with* $q_z(G_4, G_4) = t_0$? If G_4 is self-dual, then the tadpole condition becomes $q_z(G_4, C_z G_4) = t_0$, and this is nothing but the statement that the Hodge norm squared of G_4 is equal to t_0. In other words, the question is whether there are finitely many points on a lattice that have a given norm, and this is of course the case.

The difficulty now lies in concluding that such a finiteness statement holds uniformly in $z \in \mathcal{M}$. Indeed, when approaching the singular locus $\mathcal{M}_{\text{sing}} = \overline{\mathcal{M}} \setminus \mathcal{M}$, the Hodge norm may decay asymptotically, meaning that the number of G_4 on the lattice $(\mathcal{H}_\mathbb{Z})_z$ could grow without a bound, as illustrated in Figure 2.3.

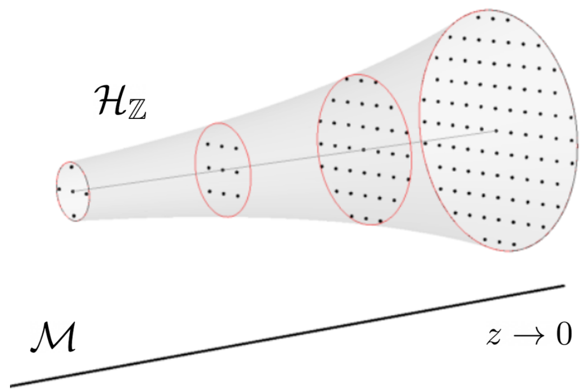


Figure 2.3: Schematic illustration of a growing number of flux vacua. The moduli space $\mathcal{M}$ is represented by a line, and a limit $z \to 0$ is taken. The vertical slices represent fibers of the Hodge bundle $\mathcal{H}_\mathbb{C}$, and in these fibers a circle of radius $\sqrt{t_0}$ is shown. In the scenario shown here, the Hodge norm decays as $z \to 0$, leading to a growth of the number of integral self-dual points inside the circle.

The scenario shown in Figure 2.3 suggests that it could be possible that the number of flux vacua is infinite, due to an unbounded growth associated to a singular limit $z \to 0$. Fortunately, one can show that situations of this type never occur. This will be discussed further in Chapter 5.

Now that we have a basic acquaintance with the idea that string theory provides a landscape of vacua, we proceed to further discuss interpretations of this landscape. This will first require us to introduce *effective field theories*.

2.5 Effective Field Theories and Supergravity

Up to this point we have worked loosely with the idea that the low-energy excitations in the spectrum of the string are described by a field theory on spacetime. In preparation for the next section, we aim to make this slightly more precise and we briefly discuss the general concept of effective field theory. For a detailed discussion, we refer to [44] and Chapter 16 of [23].

The essential idea behind effective field theory is that, in order to describe physics at a certain length scale, one does not require a complete description of the physics taking place at much smaller length scales. For example, it is not very sensible to compute string scattering amplitudes to describe the flow of water in a river. In this spirit, an *effective* theory can be obtained by *integrating out* degrees of freedom corresponding to length scales smaller than those of interest. Equivalently, this amounts to integrating out modes above[13] a certain energy scale. In a field-theoretic context, this notion of 'integrating out' can be made precise by performing a path integral over modes with an energy greater than a certain energy scale Λ_{EFT}. It is not too far-fetched to say that effective field theories govern most of the world around us. For instance, note that the Standard Model of particle physics is an effective field theory, with the scale Λ_{EFT} set by the energy scale that we can experimentally access.

In string theory, a similar procedure can be performed, and at very low energy scales, the intricate nature of the string reduces to an effective field theory [23]. The Type IIB string, for instance, reduces to the effective action of equation (2.2) which is obtained by integrating out all the massive modes[14]. This description is valid for Λ_{EFT} below the mass scale of the string, which is assumed to be much larger than energy scales that are currently accessible in particle accelerators. It should be noted, however, that part of the reason that we study the massless modes is that the higher energy modes are often very difficult to describe. Nonetheless, studying the low-energy limits of string theory has fortunately proven to be a fruitful endeavour, and much of the landscape of string theory vacua can already be understood from this effective perspective.

The type of effective field theory that arises from string theory is special. Since the theory is supersymmetric and the massless modes of the string include the graviton, the resulting low-energy effective theories are ten-dimensional supergravity theories. For example, the action in equation (2.2) describes the bosonic part of the so-called Type IIB supergravity[15]. This theory has $\mathcal{N} = 2$ supersymmetry, corresponding to 32 supercharges. The next step is to perform the compactification procedure, leading to a four-dimensional effective field theory. This is a non-trivial task, as it depends on detailed information of the internal geometry. For Calabi-Yau compactifications, results were found in e.g. [45, 46]. Whereas there are only a handful of ten-dimensional effective field theories arising from string theory, in four dimensions the situation is totally different: every string vacuum in principle gives rise to a different effective field theory.

[13] Recall that energy scales inversely with length.

[14] This first requires one to pass from strings to fields. One way to do this is to compute string scattering amplitudes for the massless states of the string, and to construct a field theory action which reproduces these amplitudes. This procedure is described in Section 16.3 of [23].

[15] The name is of course inspired by the Type IIB string. Similarly, there is a Type IIA and Type I supergravity, and there are two heterotic supergravities.

2.6 Landscape and Swampland

From the previous sections we have learned that string vacua give rise to low-energy effective field theories. Each of the resulting effective theories can be thought of as having a quantum-gravitational origin, namely string theory. We can reverse the story, starting with an effective field theory, and asking whether it is the low-energy description of a consistent quantum gravity theory. This is certainly not always the case; there is a vast amount of effective theories which can consistently incorporate gravity at energy scales below their cut-off scale, but break down when attempting find a higher-energy description. The terminology is that such an effective field theory admits no *UV-completion*[16] to a theory of quantum gravity. This idea partitions the space of consistent[17] effective field theories into two regions [7–9].

> **Definition 2.6.1.** The *landscape* is the set of effective field theories which admit a UV-completion to quantum gravity. The *swampland* is the set of consistent effective field theories which do not lie in the landscape.

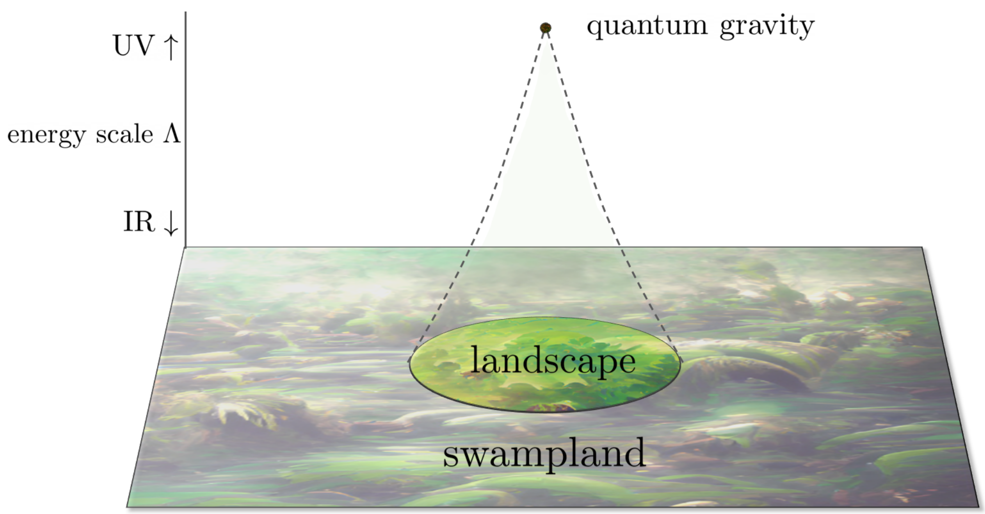


Figure 2.4: Schematic illustration of the landscape and the swampland of effective field theories. In the UV there is a single theory of quantum gravity, and as the energy scale is lowered towards the infrared (IR), effective field theories emerge from quantum gravity, illustrated by the cone. The horizontal slice represents a space of effective field theories at a given energy scale Λ, and it intersects the cone in the landscape, whereas the remaining theories belong to the swampland. As suggested by the figure, the landscape is tiny in comparison to the swampland.

This definition is illustrated in Figure 2.4. We now have two notions of landscape: a landscape of vacua in string theory and a landscape of UV-completable effective theories. There is a map from the former into the latter, but it is a priori not clear if these landscapes are the same, since effective field theories may have a UV-description in terms of a quantum gravity theory other than string theory. However, it is conjectured that the two landscapes are essentially equal; this is called *String Universality* or the *String Lamppost Principle* [9].

[16] In theoretical physics, 'UV' (ultraviolet) is typically synonymous with high energy.

[17] By 'consistent' we mean that the theory is anomaly-free, since we want to rule out theories that are bad to begin with.

> **Conjecture 2.6.2** (String Universality). *Every effective field theory in the landscape arises from string theory.*

A central question that emerges from the definition of the landscape is whether we can formulate criteria that distinguish the landscape from the swampland. Such criteria should be formulable for a generic effective field theory, and if a given effective theory does not obey it one should be able to conclude that it lies in the swampland. What makes this idea very powerful is that these hypothetical criteria may then be interpreted as general *predictions* of quantum gravity for effective theories that can describe our universe. For instance, we may not be able to directly observe strings, but through the swampland program we may find imprints of string theory on physics at our energy scales.

The program that aims to identify these criteria is called the *swampland program*, initiated by Vafa in [7]. The criteria are often conjectural in nature and known as *swampland conjectures*. The swampland program has led to a surge of activity in the high energy physics community and the list of swampland conjectures is steadily growing. It should be noted that these conjectures are not completely of the same nature as mathematical conjectures, since the framework of quantum gravity is not complete and fully understood. Nonetheless, in many cases the conjectures can be formulated mathematically in string theory settings, which in principle allows for a mathematical proof or disproof in these settings. Below we discuss a few important swampland conjectures, some of which will reappear later in this thesis. The discussion will not be completely self-contained and we will be somewhat vague about the statements of the conjectures. Two excellent reviews on the swampland program are [8] and [9].

The first swampland conjecture that we briefly discuss concerns global symmetries in a theory of quantum gravity. Roughly speaking, a global symmetry is a transformation of the theory that leaves the physics of the theory invariant while acting non-trivially on the states of the theory. This conjecture is called the No Global Symmetries (NGS) Conjecture is the oldest and most well-established of the conjectures.

> **Conjecture 2.6.3** (No Global Symmetries Conjecture). *There are no global symmetries in a theory of quantum gravity; any symmetry must be either broken or gauged.*

Contrary to what we stated above, this conjecture does not say something about effective theories but rather about quantum gravity itself. Though it is the conjecture with the most support, it is therefore also the one with the least amount of phenomenological consequences for effective field theories. However, there are several ways in which the NGS Conjecture manifests itself at the level of effective field theories, for instance the so-called *Cobordism Conjecture* [47,48].

There is much evidence for the NGS Conjecture. Arguably the simplest evidence is an argument based on the semi-classical properties of black holes [8, p. 32], but there are also many arguments based on string theory. Without going into any detail of these arguments, note that they are of a different type: the black hole argument is a ‘bottom-up’ argument in the sense that it only assumes some general phenomenology, namely the existence and properties of black holes. On the other hand, evidence from string theory is ‘top-down’ since it is based on a full theory of quantum gravity. This is a general theme in the swampland program, and the most favored conjectures come with a mixture of bottom-up and top-down evidence.

A closely related conjecture is the *Completeness Hypothesis*. It concerns gauge symmetries instead of global symmetries, and it essentially says that the spectrum of a theory must be complete.

Conjecture 2.6.4 (Completeness Hypothesis)**.** *An effective field theory which is coupled to gravity and contains gauge symmetries must contain physical states with all possible gauge charges consistent with Dirac quantization.*

More in-depth explanations of this conjecture can be found in [8] and [9]; for our purposes we do not require the details. Later in Chapter 6 we will mention it in an argument, and therefore we state the conjecture for the sake of completeness.

A conjecture which is more geometric in nature is the *Distance Conjecture*, proposed in [49]. It is a statement about the moduli in an effective field theory, and in particular about asymptotic regions in the moduli space.

Conjecture 2.6.5 (Distance Conjecture)**.** *Consider an effective field theory coupled to gravity with a scalar field space $\mathcal{M}$ as the target space of a number of moduli fields ϕ^i. Then for any point $P \in \mathcal{M}$, there exists a point Q_∞ such that the geodesic distance $d(P, Q_\infty)$ is infinite. Moreover, for any Q approaching Q_∞ there exists an infinite tower of states with characteristic mass scale M such that*

$$M(Q) \sim M(P)\, e^{-\alpha\, d(P,Q)},$$

for a certain positive constant $\alpha > 0$.

Formally, the point Q is not required to be in $\mathcal{M}$, and may be interpreted as a point at infinity. For instance, for the complex structure moduli space $\mathcal{M}$, Q may be a point in the larger space $\overline{\mathcal{M}}$ which contains $\mathcal{M}$. Purely in terms of the geometry of $\mathcal{M}$, the statement is that there is always a point Q at arbitrarily large distance from P.

The physical significance of the conjecture is that in such an infinite distance limit, the presence of an infinite tower of light states signals a breakdown of the effective field theory description. We can immediately see an example of this conjecture by recalling the circle compactification from earlier in this chapter.

Example 2.6.6. The compactification on a circle gave rise to a single modulus, namely the circle radius R taking values in the moduli space $\mathcal{M} = \mathbb{R}_{>0}$. From the KK mode expansion on the circle, an infinite tower of fields ϕ_n emerged with a characteristic mass scale $M(R) = 1/R$. It can be shown that the metric on the moduli space $\mathbb{R}_{>0}$ is[17]

$$g = \frac{\beta^2}{R^2} dR^2, \tag{2.38}$$

where $\beta > 0$ is a constant.

[18]The way to derive this is to dimensionally reduce the Einstein-Hilbert action for the metric from five to four dimensions. This amounts to expanding the five-dimensional Ricci scalar in terms of the four-dimensional Ricci scalar and the radius R, and then reading off the action for the modulus R (see e.g. [8, p. 20]).

We see that there indeed exists an infinite distance limit, namely $R \to \infty$. Fixing a reference point $1 \in \mathbb{R}_{>0}$, the distance to a radius $R > 1$ is

$$\int_1^R \sqrt{\beta^2/r^2}\, dr = \beta \log R, \tag{2.39}$$

and the characteristic mass scale of the KK tower compares as

$$M(R) = \frac{M(1)}{R} = M(1)\, e^{-\beta\, d(1,R)}, \tag{2.40}$$

precisely the type of exponential behaviour stated in the Distance Conjecture. However, there is another infinite distance limit, namely $R \to 0$. Following the same argument as above, we find that the KK tower becomes exponentially heavy rather than exponentially light, seemingly contradicting the Distance Conjecture. The underlying reason is that we simply considered the compactification of a scalar field, whereas the Distance Conjecture is about quantum gravity or string theory. If we compactify *string theory* on a circle, one finds that strings are allowed to wind around the circle, leading to *winding modes*. Like the tower of KK modes, these modes are labelled by an integer and come with a characteristic mass scale, only this time the mass scale is proportional to a positive power of R. Therefore, in the infinite distance limit $R \to 0$, the winding tower is the infinite tower of states that becomes exponentially light, in agreement with the Distance Conjecture.

A setting in which the Distance Conjecture is well-understood is the complex structure moduli space $\mathcal{M}$ in Calabi-Yau compactifications [50, 51]. We have seen earlier in this chapter that the geometry of $\mathcal{M}$ is governed by the Weil-Petersson metric $\mathcal{K}_{i\bar{\jmath}}$. In this geometry there are indeed points that lie at an infinite distance from a generic point in $\mathcal{M}$, and these points all lie on the singular locus $\mathcal{M}_{\text{sing}}$ on which the Calabi-Yau degenerates. Using variations of Hodge structure, and in particular the orbit theorems, it is in many cases possible to identify the infinite towers of states that become light, providing much mathematical evidence for the Distance Conjecture. This has inspired the use of Hodge-theoretic techniques in the complex structure moduli space, leading to many new insights and results in string theory compactifications [38, 42, 52, 53].

There are many more important swampland conjectures, such as the Weak Gravity Conjecture and the de Sitter Conjecture, but since these will not play a role in this thesis we end our review of the conjectures here. One of the strengths of the swampland program is that the conjectures often do not stand on their own, but form a web of interrelated ideas. The consistency and connectivity of these ideas may then deliver a clear picture of what is allowed and what is not allowed in quantum gravity.

One of the focal points of this thesis is a new swampland conjecture, recently proposed in [1]. This new conjecture is called the Tameness Conjecture, and it is based on the notion of tame geometry mentioned in the introduction. In essence, the conjecture states that geometric quantities in effective field theories are 'tame'. These geometric quantities include the scalar field space, any type of function of the scalars that couples the fields, but also spaces parametrizing the effective field theory, such as the locus of flux vacua. Before we can make these statements precise, we have to learn the basics of tame geometry, which brings us to the next chapter.

Chapter 3

Tame Geometry and O-minimal Structures

The aim of this chapter is to provide a detailed introduction to tame geometry, intended for both physicists and mathematicians. Tame geometry is a type of geometry in which the underlying topology is moderate, or *tame*, in a sense that we will explain precisely below. The chapter begins with a discussion of the original motivation for tame topology, after which we proceed to discuss a more contemporary motivation for the topic. Equipped with the motivation to study tame geometry, we begin to explore how to make this idea precise. The central definition is that of an *o-minimal structure*. The resulting geometry based on o-minimal structures has various tameness properties, and we give an overview of the most important such properties. We then proceed to discuss how o-minimality interacts with other types of geometry, such as algebraic geometry and differential geometry. Some proofs will be provided, and otherwise we give references to appropriate literature.

3.1 Motivation for Tame Geometry

The idea of tame topology and tame geometry originates from a vision of Grothendieck for long-term research directions in mathematics [54]. Among these directions is a proposal for a new framework of general topology. He argued that topology had been created for the purposes of analysis, and that it is not the most natural setting for studying geometry. Instead he stated that, if used as a foundation for geometry, classical topology should be replaced by *tame* or *moderate* topology. One of the objectives that tame topology should accomplish is that 'pathological' objects are discarded. Deciding exactly which phenomena are pathological is subjective, but there are some basic undesirable properties that many of the standard examples of pathological spaces share. As an example, let us revisit the topologist's sine curve encountered in the introduction.

Example 3.1.1. Let A be the graph of the real function $x \mapsto \sin(1/x)$, defined for $x > 0$, as illustrated in Figure 3.1. There are at least three reasons why A is not considered to be a tame set. Firstly, the closure $\mathrm{cl}(A)$ consists of the union of A with a vertical line segment at $x = 0$, and we have

$$\dim\left(\mathrm{cl}(A)\backslash A\right) = \dim(A) = 1.$$

For geometric purposes this is a bad property, and in a framework of tame topology we expect that boundaries of spaces are of strictly lower dimension. This is what we meant when we vaguely stated that the dimension of this space becomes obscure near $x = 0$ in the introduction. Secondly, the closure $\mathrm{cl}(A)$ is connected but not path-connected, which is unsatisfactory since curves are vital to geometry. As a final reason, the intersection of A with the horizontal real axis is infinite and discrete. Although this is not necessarily pathological, we will later see that finiteness is a crucial ingredient for tame topology, and that infinite discrete objects should be excluded.

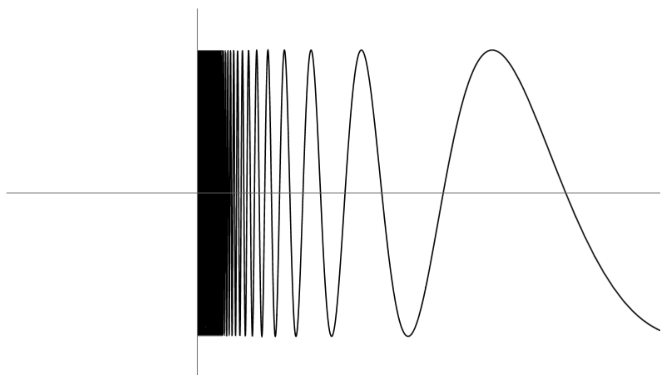


Figure 3.1: The graph of the function $x \mapsto \sin(1/x)$ for $x > 0$. This space shows pathological behavior, and is certainly not tame.

Over the years, it has become widely agreed that the right candidate for tame topology is *o-minimality*. The concept of o-minimality originates from logic and model theory, but more recently it has become clear that o-minimal geometry can be understood and applied without reference to model theory. The types of spaces that can be constructed in this framework are axiomatically restricted, and in this way pathologies are avoided.

The original motivation for tame topology is somewhat historic, and indeed the classic framework of topology is definitely still the standard for the purposes of studying geometry. Nonetheless, tame topology is still valuable. The category of o-minimal spaces lies somewhere in between the algebraic and analytic category, and consequently tame geometry turns out to give a very appropriate description to spaces and maps that naturally lie in between these categories. On the mathematical side, we will see that this is the case for many objects in Hodge theory (Chapter 4), and on the physical side this happens for many geometric objects in effective field theories, made precise by the Tameness Conjecture (Chapter 5).

The interaction between algebraic and analytic geometry is captured by a number of famous theorems, most notably the Chow theorem and Serre's GAGA[1] theorem [16, 55]. Interestingly, both of these theorems admit generalizations to o-minimal geometry. These generalizations will be discussed briefly later in this chapter.

Aside from this intrinsic interest, techniques from o-minimality have recently driven significant progress on several long-standing mathematical conjectures, most notably on the Ax-Schanuel conjecture, the André-Oort conjecture and the Griffiths conjecture [3, 56, 57]. Even for the famous Hodge conjecture, new insights have been obtained from o-minimality [2]. However, aside from the Hodge conjecture, the statements of these conjectures and a discussion of this recent progress is beyond the scope of this thesis.

Arguably, the points made in the last few paragraphs should be the main motivation to study tame geometry and o-minimal structures. Equipped with this motivation, we are now ready to set up the main definitions, following [58, 59].

3.2 O-minimal Structures

Our starting point will be the notion of a *structure*. This concept originates from model theory, but can be defined and studied entirely within set theory.

Definition 3.2.1. A *structure* on the real numbers $\mathbb{R}$ is a collection $\mathcal{S} = (\mathcal{S}_n)_{n\geq 1}$, with $\mathcal{S}_n$ consisting of subsets of $\mathbb{R}^n$, satisfying the following conditions.

(i) Each $\mathcal{S}_n$ is closed under finite unions, finite intersections, and complements.

(ii) If $A \in \mathcal{S}_m$ and $B \in \mathcal{S}_n$, then $A \times B \in \mathcal{S}_{m+n}$.

(iii) If $A \in \mathcal{S}_n$ and $p : \mathbb{R}^n \to \mathbb{R}^{n-1}$ is a linear projection, then $p(A) \in \mathcal{S}_{n-1}$.

(iv) Each $\mathcal{S}_n$ contains all the algebraic subsets of $\mathbb{R}^n$.

In other words, a structure specifies the sets that we wish to allow in our model of topology, with some reasonable conditions. The elements in the $\mathcal{S}_n$ are called $\mathcal{S}$-*definable*, or simply *definable* if the choice of structure is clear from the context. If A and B are definable sets, then a map $f : A \to B$ is definable if the graph

$$\Gamma(f) = \{(x, y) \,|\, f(x) = y\} \tag{3.1}$$

is a definable subset of the product $A \times B$.

Since 3.2.1(iv) tells us that every structure on $\mathbb{R}$ must at least contain the algebraic sets, it is natural to ask whether the collection of algebraic sets forms a structure. The answer is no, but by being slightly more general one arrives at the most basic example of a structure.

[1]The acronym 'GAGA' stands for *géométrie algébrique et géométrie analytique*.

Example 3.2.2. A *semi-algebraic set* is a set obtained by taking intersections, unions, and complements of sets defined by finitely many polynomial equalities and inequalities. For example, if P and Q are real polynomials in n variables, then the set

$$\{(x_1,\dots,x_n)\in\mathbb{R}^n \mid P(x_1,\dots,x_n)=0,\ Q(x_1,\dots,x_n)>0\}$$

is semi-algebraic. The collection of semi-algebraic subsets of $\mathbb{R}^n$ for $n\ge 1$ forms a structure denoted by $\mathbb{R}_{\mathrm{alg}}$. This turns out to be the smallest structure. To see that the algebraic sets by themselves are not large enough to form a structure, take for instance the linear projection $(x,y)\mapsto y$ of the parabola

$$\{(x,y)\in\mathbb{R}^2 \mid x^2=y\}\subseteq\mathbb{R}^2.$$

The resulting image is the set $[0,\infty)$, which is semi-algebraic (it is defined by the equation $y\ge 0$) but not algebraic. The non-trivial part in showing that $\mathbb{R}_{\mathrm{alg}}$ is a structure is the closure under linear projections, and this is a famous result in model theory[1] due to Tarski and Seidenberg [60].

The following result summarizes some basic properties of structures, and shows that structures interact well with point-set topology.

Proposition 3.2.3. *Fix a structure $\mathcal{S}$, and endow each $\mathbb{R}^n$ with the Euclidean topology.*

(i) Preimages and images of definable sets under definable maps are definable;

(ii) The composition of two definable maps is definable;

(iii) Closures, interiors, and boundaries of definable sets are definable.

Proof. (i) Let $f: A\to B$ be a definable map, and suppose that $U\subseteq A$ and $V\subseteq B$ are definable sets. Definability of $f(U)$ and $f^{-1}(V)$ then follows from writing

$$f(U)=p_B\big((U\times B)\cap\Gamma(f)\big),\quad f^{-1}(V)=p_A\big((A\times V)\cap\Gamma(f)\big). \tag{3.2}$$

Here p_A and p_B are the linear projections from $A\times B$ to A and B, respectively. Note that the images of these projections are definable by using 3.2.1(i)-(iii) iteratively. For (ii), let $g: B\to C$ be definable. We can express the graph of $g\circ f$ using the projection $p_{A\times C}: A\times B\times C\to A\times C$ and the definable graphs $\Gamma(f)$ and $\Gamma(g)$ as

$$\Gamma(g\circ f)=p_{A\times C}\big(\Gamma(f)\times C\cap A\times\Gamma(g)\big), \tag{3.3}$$

from which it follows that $g\circ f$ is definable. To see that (iii) holds, observe that the closure of a definable set A can be written as

$$\begin{aligned}\mathrm{cl}(A)&=\left\{x\in\mathbb{R}^n \,\middle|\, \forall\epsilon>0\ \exists y\in A \text{ s.t. } \|x-y\|^2<\epsilon\right\} \\ &=\mathbb{R}^n\setminus p_{\mathbb{R}^n}\left(\mathbb{R}^{n+1}\setminus p_{\mathbb{R}^n\times\mathbb{R}}\left(\left\{(x,\epsilon,y)\in\mathbb{R}^n\times\mathbb{R}\times A \,\middle|\, \|x-y\|^2<\epsilon\right\}\right)\right).\end{aligned} \tag{3.4}$$

In the second line we have re-expressed the conditions determined by the quantifiers as projections and complements of the semi-algebraic set

$$\left\{(x,\epsilon,y)\in\mathbb{R}^n\times\mathbb{R}\times A \,\middle|\, \|x-y\|^2<\epsilon\right\}.$$

[2]In model theory, it is often referred to as 'quantifier elimination' for the real numbers $\mathbb{R}$.

Note that this set is indeed semi-algebraic since $\|x-y\|^2$ is a polynomial in the coefficients of x and y. The definability of $\operatorname{int}(A)$ and ∂A now follow from basic identities in point-set topology. $\square$

The strategy of this proof was to construct a set defined in terms of certain logical conditions using the basic set-theoretic operations under which a structure $\mathcal{S}$ is closed. The reason that this works, is that the axioms that define a structure actually reflect logical operations, as illustrated in the proof of the proposition below. Identifying precisely which logical conditions lead to definable sets and functions is therefore of great value to streamline arguments in o-minimal geometry.

Definition 3.2.4. A *first-order formula* is a logical statement $\big[\dots\big]$ defined according to the following rules.

(i) If P is an n-variable polynomial with real coefficients, then

$$\big[P(x_1,\dots,x_n)=0\big] \quad \text{and} \quad \big[P(x_1,\dots,x_n)>0\big]$$

are first-order formulas.

(ii) If $A\subseteq\mathbb{R}^n$ is definable, then $\big[x\in A\big]$ is a first-order formula.

(iii) If $\big[\Phi(x)\big]$ and $\big[\Psi(x)\big]$ are first-order formulas, then

$$\big[\Phi(x) \text{ and } \Psi(x)\big], \quad \big[\Phi(x) \text{ or } \Psi(x)\big], \quad \big[\text{not } \Phi(x)\big], \quad \text{and} \quad \big[\Phi(x) \text{ implies } \Psi(x)\big]$$

are first-order formulas.

(iv) If $\big[\Phi(x,y)\big]$ is a first-order formula with $x\in\mathbb{R}^m$ and $y\in\mathbb{R}^n$, and $A\subseteq\mathbb{R}^n$ is definable, then

$$\big[\exists y\in A \text{ such that } \Phi(x,y)\big] \quad \text{and} \quad \big[\forall y\in A\ \Phi(x,y)\big]$$

are first-order formulas.

The use of this definition becomes apparent in the following lemma [61].

Proposition 3.2.5. *Let $\big[\Phi(x)\big]$ be a first-order formula. Then the set $\{x\in\mathbb{R}^n \mid \Phi(x) \text{ is true}\}$ is definable.*

Proof. We prove the result inductively, constructing sets using the rules that define first-order formulas. The sets constructed from rule (i) are semi-algebraic, which are definable in any structure. Rule (ii) tautologically gives definable sets. Let $\{\Phi(x)\}\subseteq\mathbb{R}^n$ denote the set defined by a first-order formula $\big[\Phi(x)\big]$. For rule (iii), observe that

$$\begin{aligned}
\{\Phi(x) \text{ and } \Psi(x)\} &= \{\Phi(x)\}\cap\{\Phi(x)\},\\
\{\Phi(x) \text{ or } \Psi(x)\} &= \{\Phi(x)\}\cup\{\Phi(x)\},\\
\{\text{not } \Phi(x)\} &= \mathbb{R}^n\setminus\{\Phi(x)\},\\
\{\Phi(x) \text{ implies } \Psi(x)\} &= \mathbb{R}^n\setminus\{\Phi(x) \text{ and not } \Psi(x)\},
\end{aligned}$$

which are all definable by definition of a structure. Finally, we can use the fact that linear projections preserve definability to show that rule (iv) leads to definable sets. Indeed, letting $p : \mathbb{R}^{m+n} \to \mathbb{R}^n$ denote the linear projection onto the first m coordinates,

$$\{x \mid \exists y \in A \text{ such that } \Phi(x,y)\} = p\big((\mathbb{R}^m \times A) \cap \{\Phi(x,y)\}\big),$$
$$\{x \mid \forall y \in A\, \Phi(x,y)\} = \mathbb{R}^m \setminus p\big((\mathbb{R}^m \times A) \cap (\mathbb{R}^{m+n} \setminus \{\Phi(x,y)\})\big),$$

we see that the sets arising from rule (iv) are definable. This shows that all sets arising from first-order formulas are definable, and completes the proof. $\square$

In practice, we do not work with first-order formulas. Instead, we use this proposition implicitly to quickly argue that sets are definable. Note that this result also explains the usage of the word 'definable': the definable sets are precisely those that can be defined using first-order formulas.

It is reasonably straightforward to show that the intersection of a set of structures is again a structure. Hence, we could start with a selection of sets that we wish to be definable and consider the smallest structure containing this selection, namely the intersection of all structures in which our selection is definable. The resulting structure is the structure *generated* by our selection of sets. Typically, this selection will consist of graphs of functions that we would like to be definable. This construction gives rise to the following important examples of structures.

Example 3.2.6. (i) The real exponential function $\exp : \mathbb{R} \to \mathbb{R}$ generates a structure denoted by $\mathbb{R}_{\text{exp}}$.

(ii) A *restricted analytic function* is a restriction to a closed ball B of a real analytic function defined on an open set containing B. The collection of restricted analytic functions generates a structure denoted by $\mathbb{R}_{\text{an}}$.

(iii) The real exponential function and the restricted analytic functions together generate the structure $\mathbb{R}_{\text{an,exp}}$. As we will see later, this structure plays a substantial role in tame geometry, from the perspective of both mathematics and physics.

So far we have considered structures in full generality, without considering any conditions on how large the structures can get. To obtain a framework for tame topology we now impose crucial finiteness condition, leading to the notion of an o-minimal structure.

Definition 3.2.7. An *o-minimal structure* is a structure $\mathcal{S}$ for which $\mathcal{S}_1$ consists precisely of the semi-algebraic subsets of $\mathbb{R}$.

In other words, in an o-minimal structure the definable subsets of the real line are finite unions of intervals and points. It turns out that this one-dimensional condition suffices to render the geometry in higher dimensions tame; the product and linear projection condition in the definition of a structure ensure that the finiteness condition permeates to any dimension. O-minimality rules out many analytic functions such as sine and cosine, as their zero sets are infinite and discrete and therefore not definable in any o-minimal structure. This shows that demanding o-minimality is quite strong. At the tamest end of the spectrum, the structure $\mathbb{R}_{\text{alg}}$ is evidently o-minimal, as the condition is trivially satisfied. With the following non-trivial result, it becomes clearer that o-minimal lies somewhere in between algebraic and analytic.

Theorem 3.2.8. *The structures* $\mathbb{R}_{\text{exp}}$, $\mathbb{R}_{\text{an}}$, *and* $\mathbb{R}_{\text{an,exp}}$ *are o-minimal.*

These results are due to Wilkie [62], Gabrielov [63], and Van den Dries-Miller [64], respectively. Although the theorem is easy to state, that fact that these structures are o-minimal is a deep result in model theory, and proving it is outside the scope of this thesis. Thanks to this result, it is possible to do tame geometry with the real exponential and restricted analytic functions available. However, other than the o-minimality condition itself, we have not yet discussed what makes this geometry tame. This is the topic of the next section.

3.3 Tameness of O-minimal Structures

Having set up the definition of an o-minimal structure, we are ready to discuss some of the universal tame features of definable objects in an o-minimal structure. Though there are many of these features, we only treat a few essential ones that illustrate the tameness of definable sets and maps. The most notable result is the one-dimensional *monotonicity theorem*, and its generalization to higher dimensions called the *cell decomposition theorem*. Throughout this section we work with a fixed o-minimal structure, and use the word 'definable' to refer to this o-minimal structure. This section is mostly based on the extensive book by Van den Dries [58].

3.3.1 Monotonicity and Cell Decomposition

Theorem 3.3.1 (Monotonicity Theorem)**.** *Let* $f : (a, b) \to \mathbb{R}$ *be a definable function. Then there is a finite number of points*

$$a = a_0 < a_1 < \cdots < a_n < a_{n+1} = b$$

such that the restrictions $f|_{(a_k,a_{k+1})}$ *are either constant or strictly monotonic and continuous.*

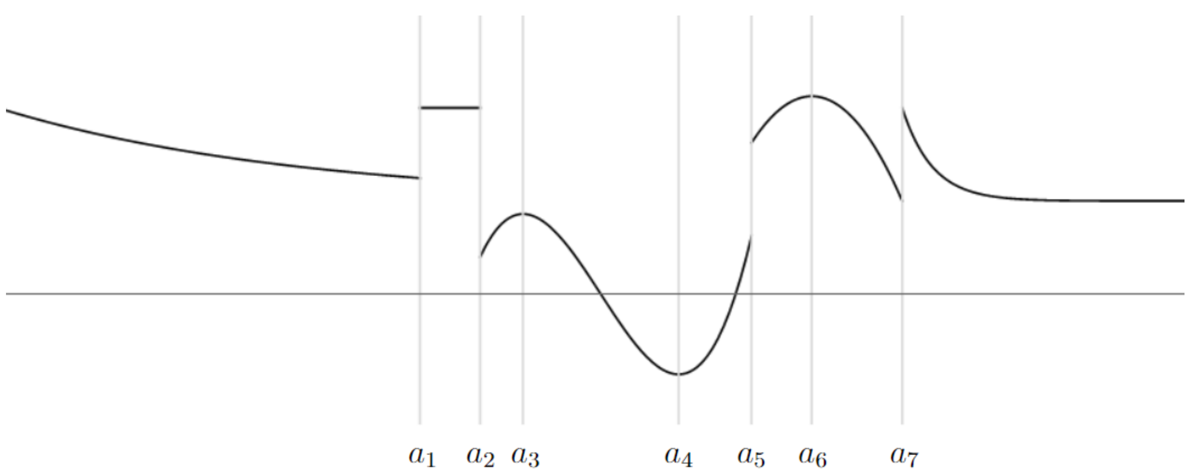


Figure 3.2: An example of a tame function on the real line. As a consequence of the monotonicity theorem, there are only finitely many discontinuities and isolated critical points. The function is either constant or strictly monotonic and continuous on each of the intervals (a_k, a_{k+1}).

This theorem precisely captures the consequences of the o-minimal finiteness condition to functions of one variable, making it one of the key results in the theory. The theorem is illustrated in Figure 3.2. Note that continuity is not part of the assumptions of the theorem, and that o-minimality alone constrains the function to have finitely many discontinuities. The reason is that the set of discontinuities is definable, and therefore constrained to be a finite union of intervals and points. This is the essential idea behind the proof of the theorem, which we present now.

Proof of Theorem 3.3.1. We follow the proof given in [58]. We structure the argument into small steps. **Step 1**. Our first step will be to prove that there exists a subinterval $I \subseteq (a,b)$ on which f is either constant or injective. If there exists an interval $I \subseteq (a,b)$ such that $f|_I$ is constant then we are done, so we now assume that there are no such subintervals. This assumption implies that the fibers[3] $f^{-1}(y)$ of f are discrete, and hence by o-minimality they are finite. To show that f is injective on some interval I we define a function

$$\begin{aligned} g : f((a,b)) &\to (a,b) \\ y &\mapsto \min f^{-1}(y). \end{aligned}$$

By construction, g is injective. As argued above, f has finite fibers, and combined with the fact that (a,b) contains infinitely many points we see that its image $f((a,b))$ is infinite. Since g is injective this in turn implies that the image of g is infinite, and hence by o-minimality it must contain an interval $I \subseteq f((a,b))$. On this interval we have $g \circ f = \mathrm{id}$, which can only hold if $f|_I$ is injective.

Step 2. Next, assuming that f is injective on (a,b), we show that the sets

$$\begin{aligned} A_\vee &= \{x \in (a,b) \,|\, \exists \epsilon > 0 \text{ s.t. } f|_{(x-\epsilon,x)} > f(x) < f|_{(x+\epsilon,x)}\}, \\ A_\diagdown &= \{x \in (a,b) \,|\, \exists \epsilon > 0 \text{ s.t. } f|_{(x-\epsilon,x)} > f(x) > f|_{(x+\epsilon,x)}\}, \\ A_\diagup &= \{x \in (a,b) \,|\, \exists \epsilon > 0 \text{ s.t. } f|_{(x-\epsilon,x)} < f(x) < f|_{(x+\epsilon,x)}\}, \\ A_\wedge &= \{x \in (a,b) \,|\, \exists \epsilon > 0 \text{ s.t. } f|_{(x-\epsilon,x)} < f(x) > f|_{(x+\epsilon,x)}\}, \end{aligned}$$

form a definable partition of (a,b). In other words, we claim that these sets are definable and that each $x \in (a,b)$ is contained in exactly one of them. The definability follows from the fact that these sets are constructed from first-order formulas (cf. Proposition 3.2.5). For the other part of the claim, note that the injectivity of f implies that that the interval (a,x) can be partitioned into the two definable sets

$$\{y \in (a,x) \,|\, f(y) < f(x)\} \quad \text{and} \quad \{y \in (a,x) \,|\, f(y) > f(x)\}.$$

Therefore, one of them contains an interval $(x-\epsilon,x)$, which determines whether $f|_{(x-\epsilon,x)} < f(x)$ or $f|_{(x-\epsilon,x)} > f(x)$ for some ϵ. Applying the same argument to the interval (x,b), we indeed see that x is contained in exactly one of the sets $A_\vee$, $A_\diagdown$, $A_\diagup$, $A_\wedge$, proving the claim.

Step 3. The result of the previous step allows us to show that there is a subinterval $I \subseteq (a,b)$ on which f is either constant or strictly monotonic. Again, we may assume that f is not constant on any subinterval of (a,b), and by Step 1 we may in addition assume that f is injective. As above, we now have a partition

$$(a,b) = A_\vee \cup A_\diagdown \cup A_\diagup \cup A_\wedge.$$

[3]By 'fiber' we mean the preimage of a point, even if there is no fiber bundle structure.

We claim that $A_\vee$ is finite. If this would not be true, then by o-minimality $A_\vee$ must contain an interval J. By the construction of $A_\vee$, every point in J is a local minimum. We can write the interval J as a union of definable sets

$$J = \bigcup_{n\geq 1} J_n, \quad \text{where } J_n = \{x \in J \,|\, x \text{ is a minimum on } (x - 1/n, x + 1/n)\}$$

(here the J_n are definable because they are constructed from a first-order formula). There must now be some J_n containing an interval, but this contradicts the construction of J_n. We thus conclude that $A_\vee$ is finite, and the same argument shows that $A_\wedge$ is finite as well. The union $A_\diagdown \cup A_\diagup$ is therefore infinite, and hence $A_\diagdown$ or $A_\diagup$ must contain an interval I by o-minimality. On this interval, f is strictly monotonic by definition of $A_\diagdown$ and $A_\diagup$. This completes Step 3.

Step 4. Next, we show that the statement of the previous step can be refined, and that there exists an interval I on which f is either constant or strictly monotonic and continuous. By the previous step, we may already assume that f is strictly monotonic on (a, b). Since f is injective, its image contains an interval J. Let $I = f^{-1}(J)$ be the preimage of this interval. We now have a strictly monotonic bijection

$$f : I \to J$$

between two intervals, which implies that f must be continuous.

Step 5. Using the statement of Step 4, we can now prove the theorem. We claim that the definable set

$$A = (a, b) \setminus \{x \,|\, \exists \epsilon > 0 \text{ s.t. } f \text{ is constant or strictly monotonic and continuous on } (x - \epsilon, x + \epsilon)\}$$

is finite. If it would be infinite, it would contain an interval. By restricting to this interval, Step 4 tells us that there must be some subinterval on which f is constant or strictly monotonic and continuous. This clearly contradicts the definition of A, so A indeed consists of a finite number of points, say $A = \{a_0, \ldots, a_{n+1}\}$. Now on each of the intervals (a_k, a_{k+1}), f is either strictly monotonic and continuous or constant since at most one of these two properties can hold at every point in (a_k, a_{k+1}). This concludes the proof of the theorem. $\square$

One major consequence of the monotonicity theorem is that asymptotic limits of definable functions are well-behaved. Consider a definable function f on the real line. The theorem asserts that there exists some $a_n \in \mathbb{R}$ such that f is constant or strictly monotonic and continuous on (a_n, ∞), and therefore that the limit $\lim_{x\to\infty} f(x)$ exists in $\mathbb{R} \cup \{\pm\infty\}$. This alludes to the fact that o-minimality does not only interact well with point-set topology, but also with differentiability. This manifests itself for instance in a generalization of the monotonicity theorem, according to which for any integer $k \geq 1$ the partition $a_0, \ldots, a_n$ can be refined so that f is $\mathcal{C}^k$-differentiable on each subinterval of the refined partition [61]. We revisit the interplay between o-minimality and differentiability later in this chapter.

We now move on to higher dimensions and discuss the cell decomposition theorem. First we introduce some notation and definitions. If A is a definable set, we denote

$$\mathcal{F}(A) = \{f : A \to \mathbb{R} \,|\, f \text{ is continuous and definable}\}, \quad \text{and} \quad \mathcal{F}_\infty(A) = \mathcal{F}(A) \cup \{\pm\infty\},$$

where $\pm\infty$ is interpreted as the constant function on A with value $\pm\infty$. For $f, g \in \mathcal{F}_\infty(A)$, we write

$$(f, g)_A = \{(a, r) \in A \times \mathbb{R} \mid f(a) < r < g(a)\}.$$

Cells are now defined as follows [58].

Definition 3.3.2. Let $(i_1, \ldots, i_n)$ be a sequence of zeros and ones. A *definable* $(i_1, \ldots, i_n)$*-cell* is a subset of $\mathbb{R}^n$ that is inductively defined in the following way.

(i) A (0)-cell of $\mathbb{R}$ is a point $\{r\}$, and a (1)-cell of $\mathbb{R}$ is an open interval (a, b).

(ii) Given a $(i_1, \ldots, i_n)$-cell A, a definable $(i_1, \ldots, i_n, 0)$-cell is the graph $\Gamma(f)$ of a definable function $f \in \mathcal{F}(A)$, and a definable $(i_1, \ldots, i_n, 1)$-cell is a set of the form $(f, g)_A$ with $f, g \in \mathcal{F}_\infty(A)$ and $f < g$.

Some examples of cells are illustrated in Figure 3.3. Intuitively, the definition implies that if A is a $(i_1, \ldots, i_n)$-cell, then $i_k \neq 0$ if and only if A is extended in the kth dimension, and A is bounded by definable and continuous functions. Every cell A has a well-defined dimension, namely $\dim(A) = \sum_{k=1}^n i_k$.

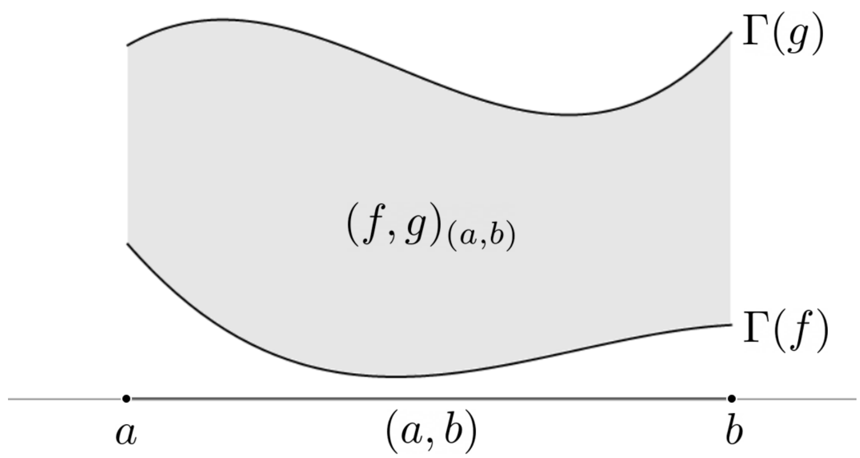


Figure 3.3: A number of cells in $\mathbb{R}^2$. The points indicated by a and b are $(0, 0)$-cells, the interval (a, b) as well as the graphs $\Gamma(f)$ and $\Gamma(g)$ are $(1, 0)$-cells, and the shaded region $(f, g)_{(a,b)}$ is a $(1, 1)$-cell.

From a topological point of view, cells are very simple.

Proposition 3.3.3. *Let A be a cell of dimension d. Then there is a definable homeomorphism $A \to \mathbb{R}^d$.*

Proof. The proof is by induction on n, where $\mathbb{R}^n$ is the ambient space of the cell A. For $n = 1$, the result is trivial. Assume now that cells in $\mathbb{R}^{n-1}$ are definable. If A is a graph-type cell of the form $\Gamma(f)$ with $f \in \mathcal{F}(B)$ for some definable cell $B \subseteq \mathbb{R}^{n-1}$, then A is homeomorphic to B via the linear projection map $\Gamma(f) \to B$. If A is a band-type cell of the form $(f, g)_B$, then A is homeomorphic to $B \times \mathbb{R}$ via the map

$$\begin{aligned} (f,g)_B &\to B \times \mathbb{R} \\ (x,y) &\mapsto \left(x, \frac{1}{f(x)-y} + y + \frac{1}{g(x)-y} \right). \end{aligned}$$

Both of these homeomorphisms are clearly definable[4]. Since B is a cell in $\mathbb{R}^{n-1}$, it is homeomorphic to $\mathbb{R}^{\dim(B)}$, and hence the result follows. □

The next essential definition is as follows.

Definition 3.3.4. A *definable cell decomposition* of $\mathbb{R}^n$ is a partition of $\mathbb{R}^n$ defined inductively as follows.

(i) A definable cell decomposition of $\mathbb{R}$ is a finite partition of $\mathbb{R}$ into (0)- and (1)-cells.

(ii) A definable cell decomposition of $\mathbb{R}^{n+1}$ is a finite partition of $\mathbb{R}^{n+1}$ into cells, such that the linear projections of the cells onto $\mathbb{R}^n$ form a definable cell decomposition of $\mathbb{R}^n$.

This definition is illustrated in Figure 3.4. Note that in the preceding two definitions, an ordering of the coordinates has implicitly been chosen. We can now state an essential tameness theorem for definable sets in o-minimal structures.

Theorem 3.3.5 (Cell Decomposition Theorem)**.** *Let $A_1, \dots, A_m \subseteq \mathbb{R}^n$ be definable sets. Then there exists a definable cell decomposition on $\mathbb{R}^n$ such that each A_k is precisely a finite union of cells. Additionally, if $f : A \to \mathbb{R}$ is definable, there exists a definable cell decomposition of $\mathbb{R}^n$ for which A is a union of cells and the restriction of f to each cell is continuous.*

For the proof we refer to Chapter 3 of [58]. This result shows that cells may be regarded as the building blocks of definable sets, and that any definable set may be built using finitely many such building blocks. Like the monotonicity theorem, the cell decomposition theorem admits a generalization in which f is $\mathcal{C}^k$-differentiable.

[4]Note that we also use the fact that division $x \mapsto 1/x$ is definable for $x \neq 0$. The reason for this is that the graph of division is precisely the algebraic set $\{(x,y) \,|\, 1 - xy = 0\}$.

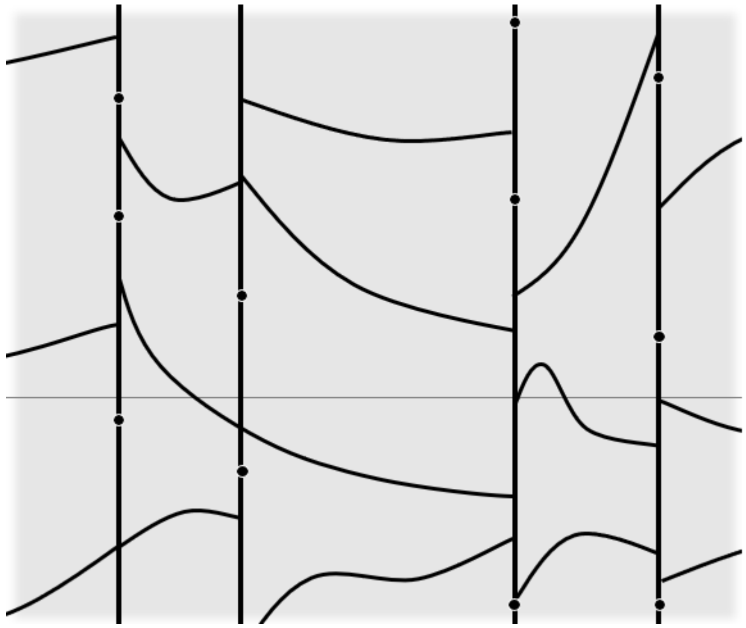


Figure 3.4: An example of a definable cell decomposition of the plane $\mathbb{R}^2$. The cells at the boundary of this figure extend infinitely far.

3.3.2 Definable Invariants and Triangulation

The existence of a cell decomposition for any definable set A allows us to assign several interesting definable invariants to A. For instance, we may define the dimension of A to be the dimension of the largest cell in the cell decomposition. It can be shown that this does not depend on the choice of cell decomposition, so that $\dim(A)$ is indeed well-defined. The following proposition, whose proof can be found in [58], shows that this definition has the right properties.

Proposition 3.3.6. *Let A and B be definable sets.*

(i) If $A \subseteq B$, then $\dim(A) \leq \dim(B)$*;*

(ii) if there is a definable bijection $A \to B$, then $\dim(A) = \dim(B)$*;*

(iii) $\dim(A \cup B) = \max(\dim(A), \dim(B))$*;*

(iv) $\dim(A \times B) = \dim(A) + \dim(B)$.

This allows us to revisit an observation from the introduction to this chapter, namely that a good notion of tame geometry should have the property that boundaries of spaces are of strictly lower dimension. Definability in an o-minimal structure indeed guarantees that this holds.

Proposition 3.3.7. *Let A be a non-empty definable set. Then*

$$\dim(\partial A) < \dim(A).$$

Let us now discuss another invariant which can be defined using the cell decomposition. Let A be a definable set, and write A as a partition of cells $A = \coprod_{j=1}^{k} A_j$. It can be shown that the integer

$$\chi(A) = \sum_{j=1}^{k} (-1)^{\dim(A_j)} \tag{3.5}$$

does not depend on the choice of cell decomposition of A; it is called the Euler characteristic. Like the dimension, it is a definable invariant in the sense that it is invariant under definable bijections. In fact, a converse also holds.

Theorem 3.3.8. *Let A and B be definable sets. Then there is a definable bijection between A and B if and only if*

$$\dim(A) = \dim(B) \quad \textit{and} \quad \chi(A) = \chi(B).$$

This definition of Euler characteristic is reminiscent of the Euler characteristic for simplicial complexes, where a similar weighted sum of the number of simplices is taken. This is no coincidence, and there is in fact a strong relation between definable sets and simplicial complexes. Let us first give a definition.

Definition 3.3.9. Let $a_0, \ldots, a_k \in \mathbb{R}^m$ be a number of points, and assume that these points are affine independent in the sense that the smallest affine subspace of $\mathbb{R}^m$ containing $a_0, \ldots, a_k$ is k-dimensional. A *simplex* in $\mathbb{R}^m$ is a set of the form

$$(a_0, \ldots, a_k) = \left\{ \sum_j t_j a_j \,|\, t_0, \ldots, t_k > 0 \text{ and } \sum_j t_j = 1 \right\},$$

and the latter is called the simplex spanned by $a_0, \ldots, a_k$. A *face* of $(a_0, \ldots, a_k)$ is a simplex spanned by a non-empty subset of $\{a_0, \ldots, a_k\}$. A *simplicial complex* in $\mathbb{R}^m$ is a finite collection $\mathcal{K}$ of simplices in $\mathbb{R}^m$ such that for any simplices $\sigma, \sigma' \in \mathcal{K}$, the intersection $\mathrm{cl}(\sigma) \cap \mathrm{cl}(\sigma')$ is either empty or equal to the closure $\mathrm{cl}(\tau)$ of a common face τ of σ and σ'.

Note that this is slightly more general than the standard definition of a simplicial complex, in which the face of every simplex in $\mathcal{K}$ is required to be in $\mathcal{K}$. The union of all simplices in a simplicial complex is denoted by $|\mathcal{K}|$ and called the polyhedron associated to $\mathcal{K}$. We can now state the *triangulation theorem*, proven in Chapter 8 of [58].

Theorem 3.3.10 (Triangulation Theorem)**.** *Let $A \in \mathbb{R}^m$ be a definable set. Then there is a definable homeomorphism between A and a polyhedron $|\mathcal{K}|$ for some simplicial complex $\mathcal{K}$.*

This result captures the idea that definable objects somehow have a finite geometric complexity, as a consequence of the o-minimality assumption. It allows one to express the topology of a definable set or function in a finite set of combinatorial data. The Euler characteristic defined above now coincides with the usual Euler characteristic of a simplicial complex,

$$\chi(|\mathcal{K}|) = \sum_{\sigma \in \mathcal{K}} (-1)^{\dim(\sigma)}. \tag{3.6}$$

An example of a triangulation is shown in Figure 3.5 below.

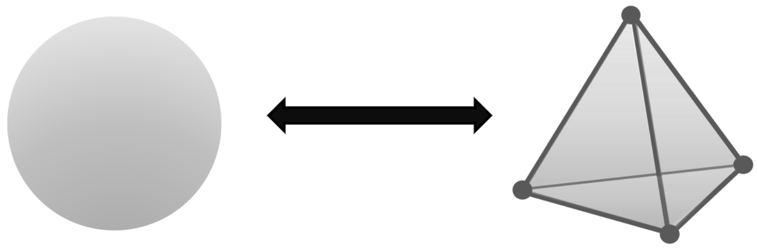


Figure 3.5: A triangulation of the 2-sphere S^2. The sphere is homeomorphic to a tetrahedron, corresponding to a simplicial complex with four 0-simplices, six 1-simplices, and four 2-simplices. The Euler characteristic is $\chi(S^2) = 4 - 6 + 4 = 2$.

3.3.3 Definable Families and Trivialization

In the first chapter we encountered families of complex manifolds, and in a similar spirit we can consider families of definable sets [58]. To do so, we let $A \subseteq \mathbb{R}^{m+n}$ be any definable set, and we consider the sets

$$A_y = \{x \in \mathbb{R}^m \mid (x, y) \in A\} \tag{3.7}$$

for $y \in \mathbb{R}^n$. In this way, A can be viewed as a collection of fibers A_y parametrized by $y \in \mathbb{R}^n$, and we regard $A = \{A_y\}_{y \in \mathbb{R}^n}$ as a *definable family.*

Example 3.3.11. Consider the definable subset of $\mathbb{R}^8 = \mathbb{R}^2 \times \mathbb{R}^6$ defined by

$$\big\{(x, y, a, b, c, d, e, f) \in \mathbb{R}^8 \mid ax^2 + bxy + cy^2 + dx + fy + e = 0\big\}. \tag{3.8}$$

This set constitutes a definable family of subset of $\mathbb{R}^2$, and topologically its fibers come in seven types: a circle, two intersecting lines, two non-intersecting lines, a single line, a single point, the whole plane, or the empty set.

The tameness of definable sets manifests itself in definable families in several ways, the most notable of which is the *trivialization theorem* [58].

Theorem 3.3.12 (Trivialization Theorem)**.** *Let $A \subseteq \mathbb{R}^{m+n}$ be a definable set and let $A = \{A_y\}_{y \in \mathbb{R}^n}$ be the corresponding definable family. Then there exists a finite partition of $\mathbb{R}^n$ into definable sets $\mathbb{R}^n = Y_1 \cup \cdots \cup Y_k$, such that for each set*

$$A_{Y_i} = \bigcup_{y \in Y_i} A_y$$

there exists a definable homeomorphism of the form

$$\begin{aligned} A_{Y_i} &\to X_i \times Y_i \\ (x, y) &\mapsto (f_i(x), y) \end{aligned}$$

for some definable set $X_i \subseteq \mathbb{R}^m$.

In other words, the fibers of a definable family always come in finitely many definable homeomorphism types, and the family can be trivialized over finitely many definable regions of the 'base space' $\mathbb{R}^n$. The proof of this theorem is complicated, but it is perhaps no surprise that the main technical tool is the cell decomposition theorem.

An interesting special case which is worth noting is the following.

Corollary 3.3.13. *Let $A \subseteq \mathbb{R}^{m+n}$ be a definable set whose definable family $A = \{A_y\}_{y\in\mathbb{R}^n}$ has finite fibers. Then the size of the fibers is uniformly bounded, i.e. there exists a constant M_0 such that $|A_y| \leq M_0$ for all $y \in \mathbb{R}^n$.*

3.3.4 Definability and Derivatives

As promised, we now revisit the idea of differentation in the context of o-minimal geometry, following [61]. There are two distinct aspects that are of interest to us: (i) how differentiability follows from o-minimality; and (ii) how we can work with o-minimality in a context where differentiability is assumed. The first aspect is mostly intended as an interesting observation and a further illustration of how the o-minimality assumption constrains definable objects in tame geometry. The second aspect serves as preparation for Section 3.6.

Let us first discuss aspect (i). For the following lemma it is convenient to allow functions to take values in the extended real numbers $\mathbb{R}\cup\{\pm\infty\}$. We say that a function $f : A \to \mathbb{R}\cup\{\pm\infty\}$ is definable if the preimage $f^{-1}(\{\pm\infty\}) \subseteq A$ is definable and the restriction $f|_{f^{-1}(\mathbb{R})}$ is a definable function. Recall that the left and right derivatives of a function $f : (a, b) \to \mathbb{R}$ are defined by the limits

$$f'_\ell(x) = \lim_{h\uparrow 0} \frac{1}{h}\big(f(x+h) - f(x)\big), \quad f'_r(x) = \lim_{h\downarrow 0} \frac{1}{h}\big(f(x+h) - f(x)\big),$$

if the limits exist.

Lemma 3.3.14. *Let $f : (a, b) \to \mathbb{R}$ be definable and continuous. The left and right derivatives $f'_\ell(x)$ and $f'_r(x)$ exist in $\mathbb{R}\cup\{\pm\infty\}$ for all $x \in (a, b)$. Moreover, the functions*

$$f'_\ell, f'_r : (a, b) \to \mathbb{R}\cup\{\pm\infty\} \tag{3.9}$$

are definable.

Proof. For the first part of the lemma, fix $x \in (a, b)$ and consider the definable function

$$h \mapsto \frac{1}{h}\big(f(x+h) - f(x)\big)$$

for h in a sufficiently small interval $(-\delta, 0)$. Applying the monotonicity theorem, it follows that this function is strictly monotonic or constant for h approaching 0, and therefore the limit $f'_\ell(x)$ exists in $\mathbb{R}\cup\{\pm\infty\}$. The same argument shows that $f'_r(x)$ exists. For the second part of the claim, note first that the sets

$$A_\ell = \{x \in (a, b) \mid f'_\ell(x) \in \mathbb{R}\} \quad \text{and} \quad A_r = \{x \in (a, b) \mid f'_r(x) \in \mathbb{R}\}$$

are definable, since they can be described in terms of first-order formulas[5]. The sets $(f'_\ell)^{-1}(\{\pm\infty\}) = (a,b)\setminus A_\ell$ and $(f'_r)^{-1}(\{\pm\infty\}) = (a,b)\setminus A_r$ are therefore also definable, as required. The definability of the restricted functions $f'_\ell|_{A_\ell}$ and $f'_r|_{A_r}$ follows from writing the graphs $\Gamma(f'_\ell)$ and $\Gamma(f'_r)$ using similar first-order formulas. $\square$

With this lemma, we can state and prove the following theorem. Recall that f is differentiable at a point x if $f'_\ell(x) = f'_r(x)$.

Theorem 3.3.15. *Let $f : (a,b) \to \mathbb{R}$ be definable and continuous. Then f is differentiable at all but finitely many points of (a,b).*

Proof. By o-minimality it suffices to show that there is no subinterval $I \subseteq (a,b)$ on which $f'_\ell \neq f'_r$. First, observe that there is no subinterval I on which $f'_\ell = \pm\infty$ or $f'_r = \pm\infty$. Hence, to prove the theorem by contradiction, suppose that there is a subinterval I on which f'_ℓ and f'_r are finite and satisfy, without loss of generality, $f'_\ell < f_r$. By the monotonicity theorem, we may pass to a smaller subinterval and assume that f'_ℓ and f'_r are continuous. Shrinking the interval I further we may in addition assume that $f'_\ell < c < f'_r$ for some $c \in \mathbb{R}$. The function $x \mapsto f(x) - cx$ is then simultaneously strictly increasing and strictly decreasing, which is a contradiction. $\square$

Note that it also follows that f' is a definable function. By applying this theorem iteratively, we see that f is in fact $\mathcal{C}^k$-differentiable outside a finite subset, for arbitrary $k \geq 1$. The generalizations of the cell decomposition to $\mathcal{C}^k$-differentiable functions mentioned earlier uses Theorem 3.3.15 extensively in its proof. Interestingly, many foundational theorems in real analysis admit definable versions, such as the inverse function theorem and the implicit function theorem. Discussing these would divert us too much from the main purpose of this chapter, and we refer to Chapter 7 of [58] for more details.

Although differentiability works in a natural way in the o-minimal setting, it is worth noting that the same is not true for integration. Defining integrals in real analysis involves infinite sums, and the infinite discrete nature of such sums signals that the o-minimal condition will generically be violated.

We now shift our focus to point (ii), and assume that we are dealing with smooth definable functions. An interesting observation is that with this assumption, the monotonicity theorem becomes simpler to understand. If $f : (a,b) \to \mathbb{R}$ is a smooth definable function, then the set

$$\{x \in (a,b) \,|\, f'(x) = 0\} \tag{3.10}$$

is definable, since it is the zero set of the definable function f'. As it is a definable subset of the real line, it is a finite union of intervals and points. The isolated points in this set correspond to points where f may change from increasing to decreasing or vice versa, and the intervals correspond to intervals on which f is constant. In other words, (a,b) can be subdivided into finitely many intervals on which f is either strictly monotonic or constant, which is precisely the statement of the monotonicity theorem.

[5] For instance, we have

$$A_\ell = \left\{x \in (a,b) \,|\, \exists f'_\ell(x) \in \mathbb{R} \text{ s.t. } \forall \epsilon > 0\, \exists \delta > 0 \text{ s.t. } \forall h \in (-\delta, 0),\, \left\| f'_\ell(x) - h^{-1}\big(f(x+h) - f(x)\big)\right\| < \epsilon \right\}.$$

The definability of derivatives of smooth definable functions of a single variable generalizes naturally to functions of multiple variables. If $f : \mathbb{R}^m \to \mathbb{R}^n$ is a definable function, it is straightforward to show using the results of this section that the partial dervatives $\partial_j f$, as well as the total derivative $Tf : \mathbb{R}^m \to \mathrm{Hom}(\mathbb{R}^m, \mathbb{R}^n)$ are definable. Later in this chapter we will see that this can be generalized even further to the case of functions on manifolds.

To close this section, we emphasize how remarkable it is that a single one-dimensional finiteness assumption gives rise to such constrained behavior for definable sets, as exemplified by the tameness results discussed in this section.

3.4 Definable Spaces

So far we have studied definable objects within an ambient Euclidean space. There is much more to say about such objects (see e.g. the extensive treatment in [58]), but we will now move on to take these objects as a local model and consider more general definable geometric objects. Most of the definitions and results that follow can be found in [2] and Chapter 10 and 11 of [58]. The first basic notion is that of a definable space.

Definition 3.4.1. An $\mathcal{S}$-*definable space* is a topological space X equipped with a finite open cover $\{U_i\}$ together with homeomorphisms $\phi_i : U_i \to A_i \subseteq \mathbb{R}^{n_i}$ such that the subsets A_i, the intersections $A_{ij} = \phi_i(U_i \cap U_j)$, and the transition functions $\phi_{ij} = \phi_j \circ \phi_i^{-1} : A_{ij} \to A_{ji}$ are $\mathcal{S}$-definable. Such a cover is called a *definable atlas*, and the maps ϕ_i are the charts of the definable atlas.

The definition is analogous to that of a manifold or variety, but a crucial distinction is that a definable space is required to have a *finite* atlas. This condition is necessary to preserve the essential finiteness of o-minimal structures. As done in the earlier sections, we usually omit the specification of the o-minimal structure $\mathcal{S}$.

Example 3.4.2. Let X be a real algebraic variety. Then X is an $\mathbb{R}_{\mathrm{alg}}$-definable space, since X can be covered by finitely many affine varieties which serve as an $\mathbb{R}_{\mathrm{alg}}$-definable atlas. Due to the compatibility of different affine covers, the definable structure on X is canonically determined by the algebraic variety structure. The construction also works if X is a complex algebraic variety, since we may then view X as a real algebraic variety of twice the dimension.

The finiteness of the cover in the definition of a definable space is reminiscent of compactness, since for compact spaces one may always extract a finite subcover from any open cover. Compactness and definability are indeed closely related.

Example 3.4.3. Let X be a compact real analytic manifold. Then X has a canonical structure of an $\mathbb{R}_{\mathrm{an}}$-definable space. To see this, cover X by real analytic coordinate charts whose chart domains are pre-compact. By compactness of X, we may take this cover to be finite. The coordinate transition functions are real analytic, and since it is assumed that the chart domains are pre-compact, they are in fact restricted real analytic. Hence, this atlas is an $\mathbb{R}_{\mathrm{an}}$-definable atlas. The manifold structure ensures that this definable structure is independent of the choice of charts.

If X is a definable space, a subset $A \subseteq X$ is definable if the image $\phi_i(A \cap U_i) \subseteq \mathbb{R}^{n_i}$ of every intersection $A \cap U_i$ is a definable set. A continuous map of definable spaces $f : X \to Y$ is definable if it is definable when viewed through the definable charts on X and Y. More precisely, we ask that each composition

$$\psi_j \circ f \circ \phi_i^{-1} : \phi_i(U_i \cap f^{-1}(V_j)) \to \psi_j(V_j)$$

is definable as a map between subsets of $\mathbb{R}^{n_i}$ and $\mathbb{R}^{m_j}$. In the following we denote compositions of this form by f_{ij}.

Many of the basic facts on definable sets and definable functions on Euclidean space extend to the global setting of definable spaces. To illustrate this, we discuss a few examples of such extensions. Firstly, we have an equivalent characterization of definable functions in terms of their graph.

Proposition 3.4.4. *A continuous map of definable spaces $f : X \to Y$ is definable if and only if its graph $\Gamma(f)$ is a definable subset of the product definable space $X \times Y$.*

Proof. Note that the definable structure on $X \times Y$ follows in a simple way from the definable structures on X and Y by taking products of charts. To be precise, given definable atlases $\{(U_i, \phi_i)\}$ for X and $\{(V_j, \psi_j)\}$ for Y, the product $X \times Y$ has a definable atlas $\{(U_i \times V_j, \phi_i \times \psi_j)\}$. First, suppose that f is definable. Then we have to show that each set

$$(\phi_i \times \psi_j)\big(\Gamma(f) \cap (U_i \times V_j)\big) \subseteq \mathbb{R}^{n_i + m_j}$$

is definable. Unwinding the definitions, we have

$$\begin{aligned}(\phi_i \times \psi_j)\big(\Gamma(f) \cap (U_i \times V_j)\big) &= (\phi_i \times \psi_j)\big(\{(x, f(x)) \,|\, x \in X\} \cap (U_i \times V_j)\big) \\ &= \{(\phi_i(x), \psi_j(y)) \,|\, x \in U_i,\, y \in V_j,\, y = f(x)\}.\end{aligned}$$

Observe that this is precisely the graph $\Gamma(f_{ij}) \subseteq \mathbb{R}^{n_i + m_j}$, which is definable by assumption. We thus conclude that $\Gamma(f)$ is definable. The converse follows directly from reversing the steps in this argument. $\square$

In the two classes of examples of definable spaces discussed above, the definable structures are canonical. However, spaces may be equipped with several inequivalent definable structures. This will be illustrated later in Example 3.4.14.

It is now natural to ask to what extent the tameness results of the previous section generalize to the setting of definable spaces. A simple way to address this question is to ask under what conditions definable spaces can be definably embedded in a Euclidean space. An appropriate sufficient condition is the following.

Definition 3.4.5. A topological space X is *regular* if for each point $x \in X$ and open $U \subseteq X$ containing x, there is an open $V \subseteq X$ with $x \in V$ and $\mathrm{cl}(V) \subseteq U$.

For definable spaces satisfying this mild topological condition, we have the following theorem, whose proof can be found in [58, p. 159].

Theorem 3.4.6. *Let X be a regular definable space. Then X is definably isomorphic to a definable subset of $\mathbb{R}^m$ for some $m \geq 1$.*

From this we observe that the tameness results from the previous section, such as cell decomposition and triangulation, continue to hold for regular definable spaces. Fortunately, every space that we will encounter in this thesis will be regular, so that we always have these tools at our disposal.

3.4.1 Quotients of Definable Spaces

Many spaces of interest in mathematics and physics arise as quotients of other spaces. In the context of this thesis it is therefore worthwhile to discuss under what conditions a quotient of a definable space is definable.

Definable equivalence relations. In general, quotient spaces arise by specifying an equivalence relation $\mathcal{R} \subseteq X \times X$ on a space X, where the resulting quotient space $X/\mathcal{R}$ is the set of $\mathcal{R}$-equivalence classes in X. If X is definable and $\mathcal{R}$ is a definable subset, it is not guaranteed that the quotient space $X/\mathcal{R}$ is definable, and further conditions on $\mathcal{R}$ must be specified to ensure definability. The situation is understood best if the following condition is assumed (Section 10.2 of [58]).

Definition 3.4.7. An equivalence relation $\mathcal{R}$ on a definable space X is *definably proper* if $\mathcal{R}$ is a definable subset and the two natural projection maps $X \times X \supseteq \mathcal{R} \to X$ are definable and proper.

Recall that a map is proper if preimages of compact sets under this map are compact.

Theorem 3.4.8. *Let X be a definable space and let $\mathcal{R}$ be a closed definably proper equivalence relation. Then the quotient $X/\mathcal{R}$ is a definable space and the natural projection $X \to X/\mathcal{R}$ is definable.*

Though the result is relatively simple to state, the proof is surprisingly technical and we refer to [58] for details. The proof is by induction on the dimension of X, and the main idea is to use a cell decomposition and construct $X/\mathcal{R}$ explicitly by a certain gluing procedure.

As an important application of this result we consider the following example [2].

Example 3.4.9. Let G be a connected semi-simple linear algebraic group and let H be a connected compact subgroup. These assumptions imply in particular that G and H are semi-algebraic and hence definable in $\mathbb{R}_{\text{alg}}$. The quotient G/H is associated to the equivalence relation

$$\mathcal{R} = \{(g, gh) \in G \times G \,|\, h \in H\},$$

which is definable since the group multiplication is algebraic. The compactness of H implies that $\mathcal{R}$ is in fact definably proper, and it follows that the quotient $G \to G/H$ is $\mathbb{R}_{\text{alg}}$-definable.

Note that the period domains $\mathcal{D}$ introduced in Chapter 1 are precisely of this form, and thus this example already allows us to conclude the following.

Corollary 3.4.10. *Period domains are definable in the o-minimal structure* $\mathbb{R}_{\mathrm{alg}}$.

This observation can be regarded as a prelude to the next chapter, where we show how tame geometry is a natural setting for Hodge theory.

Discrete group actions and fundamental sets. In the previous subsection we briefly discussed a general setting in which definable quotients exist, but for our purposes it is necessary to consider quotients that do not fall under this class. Let X be a definable space and suppose that Γ is a discrete group acting on X by definable homeomorphisms. We now wish to understand how we can make sense of $\Gamma\backslash X$ as a definable quotient[6].

An immediate problem is that if Γ is an infinite discrete group acting non-trivially, the natural projection map $X \to \Gamma\backslash X$ can never be definable. Indeed, if Γ acts non-trivially then the preimage under this projection of any point in $\Gamma\backslash X$ is given by an infinite and discrete Γ-orbit which cannot be definable in any o-minimal structure. This is an unfortunate observation, and it suggests that a weaker notion of definable quotient is required. Our next best hope is then to look for a smaller definable set $F \subseteq X$ on which the projection $F \to \Gamma\backslash X$ *is* definable. Such a subset F must be large enough to cover $\Gamma\backslash X$ but at the same time small enough to be finite under the action of Γ. This leads us to the definition of a fundamental set [65, 66].

Definition 3.4.11. Let X be a locally compact Hausdorff definable topological space and let Γ be a group acting on X by definable homeomorphisms. A *fundamental set* for the action of Γ on X is an open definable subset $F \subseteq X$ such that

(i) $\Gamma \cdot F = X$;

(ii) $\Gamma_F := \{\gamma \in \Gamma \,|\, \gamma \cdot F \cap F \neq \emptyset\}$ is a finite set.

The assumption that X is locally compact Hausdorff is a technical condition which is always satisfied in our case. Condition (i) ensures that the projection map $F \to \Gamma\backslash X$ is surjective, and condition (ii) avoids the problem that the preimage of a point in $\Gamma\backslash X$ under this quotient is an infinite and discrete set. The existence of a fundamental set guarantees that the quotient $\Gamma\backslash X$ is definable, as indicated by the following result from [65].

Proposition 3.4.12. *Let X and Γ be as in Definition 3.4.11 and suppose that F is a fundamental set for the action of Γ. Then there exists a unique definable structure on the quotient $\Gamma\backslash X$ such that the natural projection $F \to \Gamma\backslash X$ is definable. Moreover, the quotient $\Gamma\backslash X$ is locally compact Hausdorff.*

The proof uses the following lemma.

[6]In we write the quotient by Γ as $\Gamma\backslash X$ instead of X/Γ in anticipation of the next chapter.

Lemma 3.4.13. *In the setting of Proposition 3.4.12, the action of* Γ *on* X *is proper.*

Proof. By the action of Γ being proper we mean that the map

$$\begin{aligned} s : \Gamma \times X &\to X \times X \\ (\gamma, x) &\mapsto (x, \gamma \cdot x) \end{aligned}$$

is proper, i.e. the preimage $s^{-1}(L)$ of any compact subset $L \subseteq X \times X$ is compact. To show this, we first claim that for $K \subseteq X$ compact, the set

$$\Gamma_K := \{\gamma \in \Gamma \,|\, \gamma \cdot K \cap K \neq \emptyset\}$$

is finite. If Γ is a finite group then this is immediate, so assume that Γ is infinite. There are only finitely many γ for which the translate of the fundamental set $\gamma \cdot F$ has non-empty intersection with K. If this would not be true, then $\{\gamma \cdot F \cap K\}_{\gamma\in\Gamma}$ would be an open cover of K with no finite subcover, contradicting the compactness of K. Hence, denote these elements by $\gamma_1, \ldots, \gamma_k$. We now have

$$\Gamma_K \subseteq \bigcup_{j=1}^{k} \gamma_k \Gamma_F \gamma_k^{-1}, \tag{3.11}$$

because if there is an element $x \in \gamma \cdot K \cap K$, then $x \in \gamma_j \cdot F$ for some γ_j and consequently $\gamma_j^{-1} \cdot x \in \gamma_j^{-1}\gamma\gamma_j \cdot F \cap F$ so that $\gamma \in \gamma_j \Gamma_F \gamma_j^{-1}$. The set on the right-hand side of (3.11) is finite, so this shows that Γ_K is finite.

Now let $L \subseteq X \times X$ be compact. We may assume that L is of the form $K \times K$ with $K \subseteq X$ compact. The preimage $s^{-1}(K \times K)$ is contained in the set $\Gamma_K \times K \subseteq \Gamma \times X$, which is compact by what we have shown above. Since $s^{-1}(K \times K)$ is closed it follows that it is compact, and this concludes the proof that s is proper. □

Proof of Proposition 3.4.12. We follow [65]. Let

$$\mathcal{R} = \{(x, \gamma \cdot x) \,|\, x \in G, \gamma \in \Gamma\} \subseteq X \times X$$

denote the equivalence relation on X determined by the action of Γ. Observe that $\mathcal{R}$ is precisely the image of the map $s : \Gamma \times X \to X \times X$ from the previous lemma. We have shown that s is proper, and a fact from point-set topology is that a proper map into a locally compact space has closed image. It follows that $\mathcal{R}$ is a closed subset of $X \times X$. Hence, the induced equivalence relation $\mathcal{R}_F = \mathcal{R} \cap (F \times F)$ on the fundamental set F is closed as well. We have

$$\mathcal{R}_F = \bigcup_{\gamma\in\Gamma_F} \{(x, \gamma \cdot x) \in F \times F\},$$

which shows that $\mathcal{R}_F$ is a finite union of definable sets and hence definable. From the finiteness of Γ_F it follows that $\mathcal{R}_F$ is in fact definably proper, and hence by Theorem 3.4.8 the quotient $F \to F/\mathcal{R}_F$ is definable. As topological spaces, the fact that F is a fundamental set means that $F/\mathcal{R}_F = \Gamma\backslash X$, and we conclude that $\Gamma\backslash X$ has a definable structure for which the projection $F \to \Gamma\backslash X$ is definable. The uniqueness follows from the uniqueness of the definable structure on $F/\mathcal{R}_F$. □

Though the definable structure is unique for a given fundamental set $F \subseteq X$, different fundamental sets may lead to inequivalent definable structures. This is illustrated in the following example.

Example 3.4.14. The punctured disk Δ^* may be realized as a quotient $\mathbb{H}/\mathbb{Z}$ where $\mathbb{Z}$ acts as $m \cdot t = t + m$. The corresponding projection map is the usual covering map $t \mapsto e^{2\pi i t}$. Consider the semi-algebraic sets

$$F = \{x + iy \in \mathbb{H} \,|\, -1 < x < 1\}, \quad F' = \{x + iy \in \mathbb{H} \,|\, y - 1 < x < y + 1\},$$

which are evidently fundamental sets for the $\mathbb{Z}$-action on $\mathbb{H}$.

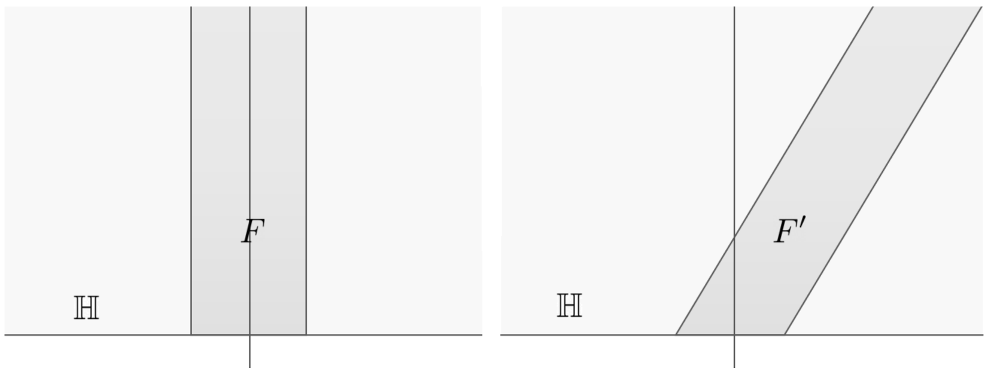


Figure 3.6: Two fundamental sets F and F' for the action of $\mathbb{Z}$ on $\mathbb{H}$.

Through Proposition 3.4.12, Δ^* can be endowed with two $\mathbb{R}_{\mathrm{alg}}$-definable structures using the fundamental sets F and F'. The definable structure determined by F coincides with the natural semi-algebraic structure from viewing Δ^* as a semi-algebraic subset of $\mathbb{C}$, as in Example 3.4.2. On the other hand, the definable structure determined by F' is inequivalent to the natural one. One way to see this is that the curve $\{e^{t+it} \in \Delta^* \,|\, t > 0\}$, which looks like an infinite spiral, is definable if the fundamental set is taken to be F', but not definable in the natural semi-algebraic structure. This is illustrated in Figure 3.7.

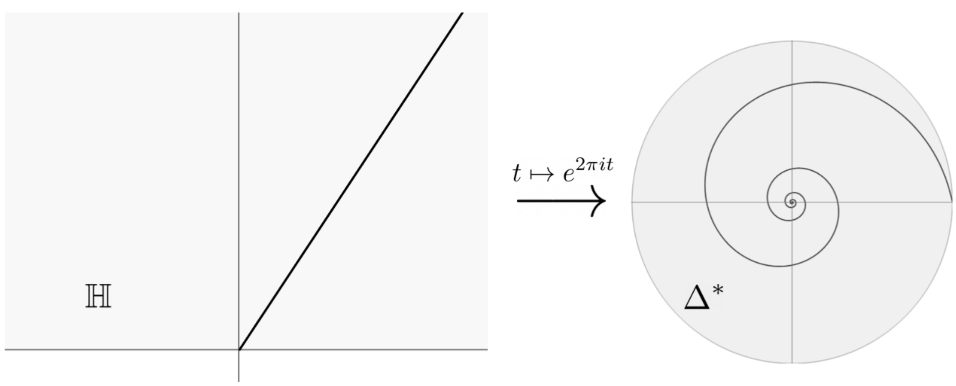


Figure 3.7: The covering map $\mathbb{H} \to \Delta^*$ sends a diagonal line to an infinite spiral in the punctured disk. This spiral is not definable in the natural semi-algebraic structure on Δ^*, but from the point of view of F' it *is* definable.

The construction of fundamental sets for certain group actions will be a central theme in the next chapter.

From this point onward there are several directions into which the theory can be developed further. One could use tame geometry as a category lying in between algebraic and analytic geometry to further study the interplay between algebraic and analytic. This direction has proven to be quite fertile, and many powerful generalizations of deep algebraic-analytic comparison theorems such as the GAGA theorem and the Chow theorem have been obtained [57] [67]. We will briefly discuss this in the following subsection.

Alternatively, one can combine tame geometry with differential geometry and study definable smooth manifolds. This direction is less explored in the literature, but nonetheless has led to some interesting developments such as an o-minimal version of de Rham cohomology [68]. For our purposes, we are mainly interested in this direction because we want explore a notion of tameness for effective field theories.

3.5 Definable Algebraic Geometry

We first recall the original Chow theorem from algebraic geometry (see e.g. Chapter 1, Section 3 of [16]).

> **Theorem 3.5.1** (Chow). *Let $X \subseteq \mathbb{CP}^n$ be a closed complex analytic subvariety. Then X is algebraic.*

In other words, the ambient variety in which X is contained dictates that analyticity coincides with algebraicity. Crucially, the ambient variety $\mathbb{CP}^n$ is compact (in the Euclidean topology). As explained by example 3.4.3 in the previous section, tameness and compactness are closely related, and one could therefore wonder if the Chow theorem admits any generalization where definability is imposed instead of compactness. It turns out that such a generalization indeed exists, as shown by Peterzil and Starchenko [67].

> **Theorem 3.5.2** ('Definable Chow'). *Let $X \subseteq \mathbb{C}^n$ be a closed complex analytic subvariety. If X is a definable set, then X is algebraic.*

Note that definability is now imposed on the variety X itself, as opposed to compactness being imposed on the ambient variety. The proof, though interesting, would require a digression which is too long. We refer to [67] for the original proof and [59] for two alternative proofs, one using mostly techniques from complex analysis and one using mostly techniques from o-minimal geometry. To give a small glimpse of how analyticity and definability together imply algebraicity, we do prove the following proposition [59].

> **Proposition 3.5.3.** *Let $f : \mathbb{C} \to \mathbb{C}$ be a holomorphic function. If f is definable, then f is algebraic.*

Proof. We first claim that f is meromorphic at infinity. To argue by contradiction, suppose instead that f has an essential singularity at infinity. Picard's great theorem then states that, near infinity, f takes any value in $\mathbb{C}$, except possibly one, infinitely often. Since f is holomorphic on $\mathbb{C}$, its fibers are discrete, and by o-minimality they must therefore be finite, leading to a contradiction. Since f is meromorphic at infinity, it is a rational function on $\mathbb{CP}^1$, and therefore algebraic. $\square$

The logical next step in this discussion would be to proceed with the o-minimal version of the GAGA theorem. The details of this discussion would however require a significant detour through sheaf theory, so we will not cover the statement of the theorem, and refer to [57, 59] for a complete discussion. Nonetheless, let us briefly comment on the result.

The idea behind the original GAGA theorem is as follows [55]. A complex algebraic variety has an interpretation as an algebraic space X (with the Zariski topology), and as an analytic space X^{an} (with the Euclidean topology). The two interpretations are related by an 'analytification functor'

$$X \mapsto X^{\mathrm{an}}. \tag{3.12}$$

For both X and X^{an}, one can consider a certain class of sheaves, namely the *coherent sheaves*. These sheaves may be viewed as a generalization of vector bundles, and capture a lot of the geometry of the underlying space. The statement of the GAGA theorem is, roughly speaking, that the analytification functor induces an equivalence between the category of coherent sheaves on X and on X^{an}, allowing one to pass back and forth between algebraic and analytic. For this reason, it is both an elegant and practical result.

In the definable setting, a new functor enters the game, which associates to a complex algebraic variety X its natural interpretation as a definable space X^{def} (as done in Example 3.4.2)[7]. The content of the definable GAGA theorem, proven in [57], is that this functor allows one to compare coherent sheaves on X and X^{def} in a similar, but slightly weaker way[8]. This becomes a powerful tool when studying the interplay between algebraic and definable geometry, and is used for instance in applications of o-minimality to Hodge theory. In the Hodge theoretic-applications studied in this thesis, we will not require these techniques.

[7]For the functor $X \mapsto X^{\mathrm{def}}$, Tsimerman coined the term 'definabilization', which he described as "... a word that I am very proud of, but no one else seems to like."

[8]The precise statement is that the functor $\mathrm{Coh}(X) \to \mathrm{Coh}(X^{\mathrm{def}})$ is fully faithful, but not essentially surjective (as in the analytic case). Here $\mathrm{Coh}(-)$ denotes the category of coherent sheaves.

3.6 Definable Differential Geometry

We now combine the notion of definable space with smoothness and arrive at the definition of a definable manifold. These may be defined in terms of $\mathcal{C}^k$-differentiability for any k, but we restrict ourselves to the $\mathcal{C}^\infty$ case. The interaction of definable and differential geometry is not explored much in the literature. Most of the results in this section can be found in [68], and in some cases we provide a new point of view or result.

> **Definition 3.6.1.** Let $\mathcal{S}$ be an o-minimal structure. An *n-dimensional $\mathcal{S}$-definable smooth manifold* is a topological space X equipped with a definable atlas for which the images of the charts are definable open subsets of $\mathbb{R}^n$ and for which the transition functions are smooth. Such an atlas is called a *definable smooth atlas*.

We will usually simply refer to such a space as a definable manifold. In the spirit of the triangulation theorem, definable manifolds can be thought of as manifolds with a finite geometric complexity. This includes (real analytic) compact manifolds, as seen in Example 3.4.3, but also manifolds which are non-compact which have a tame geometry at infinity.

A first natural question that emerges is whether it is possible to actually do differential geometry with such spaces. One would at least require the tangent bundle TX of a definable manifold X to be definable, and fortunately this is the case.

> **Proposition 3.6.2.** *Let X be a definable manifold. Then TX is a definable manifold.*

Proof. Let $\{(U_i, \phi_i)\}$ be a definable smooth atlas for X and assume that X is n-dimensional. In ordinary differential geometry, one shows that TX is a manifold by constructing an atlas $\{(TX|_{U_i}, \widetilde{\phi}_i)\}$, with $\widetilde{\phi}_i$ defined by

$$\begin{aligned} \widetilde{\phi}_i : TX|_{U_i} &\to \phi(U_i) \times \mathbb{R}^n \\ (x, v^\mu \partial_\mu) &\mapsto (\phi_i(x), v^\mu) = (x^\mu, v^\mu). \end{aligned}$$

Here the vector fields ∂_μ form the local frame associated to the coordinates x^μ provided by ϕ_i. In the current setting we have finitely many of these charts, and clearly the images of these charts are definable subsets of $\mathbb{R}^{2n}$. The transition functions are smooth and it is straightforward to verify that they are definable as well. $\square$

This result naturally extends to any bundle obtained from the tangent bundle via a linear-algebraic construction, such as the cotangent bundle, bundles of differential forms and tensor bundles. Note that the projection maps to X associated to these bundles are definable as well; this is a consequence of the fact that linear projections are always definable. Before we proceed with studying sections of such bundles, it will be worthwhile to be slightly more general.

3.6.1 Vector Bundles

Intuitively, a vector bundle is definable if it can be covered by finitely many definable trivializations. More precisely, we have the following definition [68].

Definition 3.6.3. A *real definable vector bundle* is a definable map $\pi : E \to X$ of definable manifolds such that

(i) for each $x \in X$, the fiber $E_x = \pi^{-1}(x)$ is a real vector space;

(ii) X has a finite definable open cover $\{U_i\}$ such that for each i there is a definable diffeomorphism $\psi_i : E|_{U_i} \to U_i \times \mathbb{R}^d$ such that the diagram

$$\begin{tikzcd} E|_{U_i} \arrow[rr, "\psi_i"] \arrow[dr, "\pi"'] & & U_i \times \mathbb{R}^d \arrow[dl] \\ & U_i & \end{tikzcd}$$

commutes.

As usual, E is the total space, X is the base space, and d is the rank. With this definition, the tangent bundle and related bundles are not just definable manifolds but also definable vector bundles. We define a *definable section* of a definable vector bundle $\pi : E \to X$ to be a definable[9] smooth map

$$\sigma : X \to E \quad \text{such that} \quad \pi \circ \sigma = \mathrm{id}. \tag{3.13}$$

We will denote the set of definable sections of E by $\Gamma^{\mathrm{def}}(E)$, or $\Gamma^{\mathrm{def}}(E, X)$ if we wish to emphasize the base space. Since definability of a map can be characterized in terms of its graph, one can view definable sections in the following way.

Observation 3.6.4. There is a one-to-one correspondence

$$\Gamma^{\mathrm{def}}(E) \xleftrightarrow{1:1} \{\text{definable submanifolds } \Sigma \subseteq E \text{ with } |E_x \cap \Sigma| = 1 \text{ for all } x \in X\}.$$

This correspondence sends a definable section $\sigma : X \to E$ to the projection $X \times E \to E$ of its graph $\Gamma(\sigma) \subseteq X \times E$.

In other words, we may interpret a definable section as a definable subset of a bundle that intersects each fiber exactly once. While formally this says nothing new, it provides a nice intuition for definable sections.

Locally, one can always make use of the existence of a frame of a vector bundle E. By this we mean that, around any point in X there exists an open set U together with a number of sections $e_1, \ldots, e_d : U \to E$ such that for any $x \in U$, the vectors $e_1(x), \ldots, e_d(x)$ form a basis for E_x. This idea continues to hold in the definable setting: any definable trivialization $\psi : E|_U \to U \times \mathbb{R}^d$ gives rise to a definable frame $e_1, \ldots, e_d \in \Gamma^{\mathrm{def}}(E, U)$ through the identity

$$\psi((x, v)) = \big(x, (e_1(v), \ldots, e_d(v))\big) \in U \times \mathbb{R}^d. \tag{3.14}$$

[9] We apologize for the heavy but almost unavoidable usage of the word 'define'.

As an example, the tangent bundle TX and cotangent bundle T^*X of a definable space have natural coordinate frames

$$(\partial_1, \dots, \partial_n) \quad \text{and} \quad (dx^1, \dots, dx^n) \tag{3.15}$$

on any definable chart U with coordinates x^μ.

The use of a frame comes from the fact that any smooth section $\sigma : U \to E$ can be expressed as

$$\sigma = \sum_{j=1}^{d} \sigma_j \, e_j \tag{3.16}$$

for a set of smooth functions $\sigma_1, \dots, \sigma_d \in \mathcal{C}^\infty(X)$. This idea gives an extremely useful characterization of definable sections [68].

Proposition 3.6.5. *Let $E \to X$ be a definable vector bundle with a definable frame $e_1, \dots, e_d \in \Gamma^{\text{def}}(X)$. Let $\sigma : X \to E$ be a smooth section, and write $\sigma = \sum_{j=1}^{d} \sigma_j \, e_j$. Then σ is definable if and only if each coefficient function $\sigma_j : X \to \mathbb{R}$ is a definable function.*

Proof. We may prove this locally in a finite cover, so we assume that E is of the form $E = X \times \mathbb{R}^d$ by means of a definable trivialization, and that the frame consists of the functions

$$e_j : X \to X \times \mathbb{R}^d, \quad x \mapsto \big(x, (0, \dots, 0, 1, 0, \dots, 0)\big)$$

with the 1 at the jth entry. In this setting, the functions σ_j are simply given by

$$\sigma : x \mapsto \big(x, (\sigma_1(x), \dots, \sigma_d(x))\big)$$

from which it follows that σ is definable if and only if each component function σ_j is definable. □

By applying this result to the bundles TX and T^*X, we find that definable sections $v \in \Gamma^{\text{def}}(X)$ and $\omega \in \Gamma^{\text{def}}(T^*X)$, i.e. definable vector fields and definable 1-forms, are definable if and only if they can be expressed as

$$v = v^\mu \, \partial_\mu \quad \text{and} \quad \omega = \omega_\mu \, dx^\mu \tag{3.17}$$

for some definable functions v^μ and ω_μ in a definable coordinates x^μ on an open set U.

Earlier in this chapter we have seen that differentiability interacts well with definability. In the context of definable differential geometry, this manifests itself in several ways.

Proposition 3.6.6. *Let X and Y be definable manifolds, and let $f : X \to Y$ be a smooth definable map. Then the tangent map $Tf : TX \to TY$ is definable.*

Proof. As usual, we may prove this locally by working in a finite definable open cover. In definable coordinates, the tangent map is given by the total derivative, whose definability was argued in Section 3.3.4. It follows that Tf is definable as a map between definable manifolds. □

One could phrase this result as the statement that the tangent functor makes sense in the category of definable manifolds.

Proposition 3.6.7. *Let X be a definable manifold, and let $\omega \in \Gamma^{\mathrm{def}}(\bigwedge^k T^*X)$ be a definable k-form. Then the $(k+1)$-form $d\omega$ is definable.*

Proof. As usual, we may prove this locally by working in a finite definable open cover. Hence, work on an open set U with definable coordinates x^μ. We can now write ω as

$$\omega = \omega_{\mu_1 \cdots \mu_k}\, dx^{\mu_1} \wedge \cdots \wedge dx^{\mu_k}.$$

On U, the form $d\omega$ can now be expressed as

$$d\omega = (\partial_\mu \omega_{\mu_1 \cdots \mu_k})\, dx^\mu \wedge dx^{\mu_1} \wedge \cdots \wedge dx^{\mu_k}.$$

Since the coefficient functions $\partial_\mu \omega_{\mu_1 \cdots \mu_k}$ are definable by Subsection 3.3.4, we conclude via Proposition 3.6.5 that $d\omega$ is definable. Note that we have used the fact that the $dx^{\mu_1} \wedge \cdots \wedge dx^{\mu_k}$ form a definable frame of $\bigwedge^k T^*X$, and likewise for $\bigwedge^{k+1} T^*X$. □

In other words, the exterior derivative d preserves definability. Another important operation on forms is the pullback. In a similar fashion to the arguments above, it is straightforward to prove that:

Proposition 3.6.8. *Let X and Y be definable manifolds, and let $f : X \to Y$ be a smooth definable map. If ω is a definable k-form on Y, then the pullback $f^*\omega$ is a definable k-form on X.*

Note that throughout this section we have been checking whether definability holds locally, and that this only works because we work in a *finite* definable open cover.

3.6.2 Riemannian Metrics

In this short final section we specialize to a particular kind of tensor, which will be relevant in the following chapters.

Definition 3.6.9. A *definable Riemannian metric* on a definable manifold X is a Riemannian metric g on X which is definable as a section of $T^{(0,2)}X$.

Definable manifolds equipped with such a metric are called *definable Riemannian manifolds.* If the context is clear, we usually call g a definable metric. In view of Proposition 3.6.5, a definable metric can be understood locally as a metric for which the coefficient functions g_{ij} in $g = g_{ij}\, dx^i \otimes dx^j$ are definable in any definable coordinate chart.

At present manifolds with a definable metric are not well-established in the literature. Our main reason for introducing them here is because later we will encounter manifolds that naturally have a definable metric. The definability of the metric allows one to deduce several basic properties of definable Riemannian manifolds. Calculations in local definable coordinates show, for instance, that on such a manifold the various curvature tensors and scalars are definable.

Chapter 4

Tame Geometry in Hodge Theory

Now that we have introduced the three main subjects of this thesis, we are finally in a position to begin to explore how they are connected. In this chapter we will study the connection between tame geometry and Hodge theory. In particular we will see that, in a certain sense, tame geometry seems to be the perfect description for a number of objects that appear naturally in Hodge theory. The central result of this chapter is that the period map is definable in the o-minimal structure $\mathbb{R}_{\text{an,exp}}$, recently proved in [2]. After stating the theorem more precisely below, we will explain the main steps of the proof. This will first require us to understand the tame geometry of the target space of period maps, whose proof will take us on a detour through the theory of algebraic groups. We conclude the chapter with an outlook of how the tameness of the period map provides new insights into Hodge theory.

4.1 Fundamental Sets for Arithmetic Quotients

In the first chapter we have seen that a variation of Hodge structure over a space $\mathcal{M}$ can be encoded in a period map Φ taking values in a quotient $\Gamma\backslash\mathcal{D}$, where $\mathcal{D}$ is a period domain and Γ is the discrete subgroup containing the monodromy group. We have learned that the period map is holomorphic on $\mathcal{M}$, and that it becomes singular when approaching the singular locus $\overline{\mathcal{M}} \setminus \mathcal{M}$. However, the singularities are rather mild and can be described in detail due to the tools of asymptotic Hodge theory discussed in Chapter 1. Later in this chapter we will be able to capture this behavior by the non-trivial statement that the period map is definable. In order for this statement to even begin to make sense, it is first necessary to show that the target space $\Gamma\backslash\mathcal{D}$ of the period map is definable, and showing that this is true is the focus of this section.

As we have seen, the target spaces of period maps fall under a class of spaces called *arithmetic quotients*. These are spaces of the form $\Gamma\backslash G/H$, where G is a connected semi-simple Lie group, $K \subseteq G$ is a connected compact subgroup and $\Gamma \subseteq G$ is an arithmetic subgroup. In this thesis, we mean by arithmetic[1] that Γ is a finite index subgroup of $G_{\mathbb{Z}}$, where $G_{\mathbb{Z}} = G \cap \mathrm{GL}(n, \mathbb{Z})$ is defined with respect to some embedding $G \hookrightarrow \mathrm{GL}(n, \mathbb{R})$. In the Hodge-theoretic setting, we have $G = \mathrm{Aut}(H_{\mathbb{R}}, q_{\mathbb{R}})$ and we may take the group $G_{\mathbb{Z}}$ to be defined by $\mathrm{Aut}(H_{\mathbb{Z}}, q_{\mathbb{Z}})$.

[1] There exists a slightly more general definition of arithmetic which allows for Γ to be larger than $G_{\mathbb{Z}}$, but this situation will not be encountered in this thesis.

The tame geometry of arithmetic quotients is one of the main results of [2]. The first step of the proof has already been taken in the previous chapter, where we showed that the quotient $G \to G/H$ is definable. With the knowledge of Section 3.4.1, we know that to show that the arithmetic quotient $\Gamma\backslash G/H$ is definable, it is necessary to construct a definable fundamental set $F \subseteq G/H$ for the action of Γ. Fortunately, the construction of fundamental sets for actions of arithmetic groups has been studied extensively in the classical branch of mathematics called arithmetic group theory [69]. Unfortunately, however, the constructions are quite complicated and will require us to review some concepts from the theory of algebraic and arithmetic groups. It will be worth the effort though, since understanding the tame geometry of arithmetic quotients will not only allow us to see how tameness manifests itself in Hodge theory, but also to learn how tameness emerges in higher supergravity theories, to be discussed in Chapter 6.

4.1.1 Algebraic Groups

> **Definition 4.1.1.** A complex *algebraic group* is a complex algebraic variety $G_{\mathbb{C}}$ equipped with a group operation for which the group multiplication and inversion map are algebraic.

In this definition, the algebraicity of the group multiplication and inversion maps means that they are morphisms of algebraic varieties. An algebraic group $G_{\mathbb{C}}$ is *linear algebraic* if it can be embedded in $\mathrm{GL}(n,\mathbb{C})$ for some $n \geq 0$, and in what follows we will only consider linear algebraic groups unless explicitly stated otherwise. The existence of such an embedding also allows us to define the groups of real, rational, and integer points as

$$G_{\mathbb{R}} = G_{\mathbb{C}} \cap \mathrm{GL}(n,\mathbb{R}), \quad G_{\mathbb{Q}} = G_{\mathbb{C}} \cap \mathrm{GL}(n,\mathbb{Q}), \quad G_{\mathbb{Z}} = G_{\mathbb{C}} \cap \mathrm{GL}(n,\mathbb{Z}) \tag{4.1}$$

respectively. We often write G for the group of real points $G_{\mathbb{R}}$. These concepts were already used loosely in Chapter 1, where we found $\mathcal{D} = G_{\mathbb{R}}/H$ and $\check{\mathcal{D}} = G_{\mathbb{C}}/B$.

As suggested by the example of period domains, it is often relevant to be explicit about the field over which a subgroup is defined when studying subgroups of an algebraic group. Let us make this more precise. Suppose that X is a complex algebraic subvariety of $\mathbb{C}^n$ defined by the vanishing of a number of polynomials. We then say that X is *defined over* k, where k is subfield of $\mathbb{C}$, if X can be defined by the vanishing of polynomials with coefficients in k. In the context of algebraic groups, we will frequently encounter certain subgroups which are required to be defined over the field of rational numbers $\mathbb{Q}$ as an algebraic variety.

To construct fundamental sets for arithmetic quotients we introduce the notion of a Siegel set, which is a type of set that has various finiteness properties with respect to the action of Γ [69, 70]. The main idea of the construction of Siegel sets is to first decompose the group G as a product of certain subgroups. Siegel sets are then defined as products of specific subsets of the subgroups in this decomposition. To make this precise, we have to introduce the *Langlands decomposition* of an algebraic group. In order to be able to state the Langlands decomposition, we first require the following list of terminology from algebraic group theory [71].

Definition 4.1.2. Let $G \subseteq \mathrm{GL}(n, \mathbb{C})$ be a complex linear algebraic group.

(i) an element $g \in G$ is *unipotent* if every eigenvalue of g is equal to 1, and the group $G_{\mathbb{C}}$ is unipotent if each of its elements is unipotent;

(ii) G is *nilpotent* if its lower central series, inductively defined[1] by $G_{(j+1)} = [G_{(j)}, G]$ and $G_{(0)} = G$, terminates to the trivial group after a finite number of steps;

(iii) G is *solvable* if its derived series, inductively defined by $G^{(j+1)} = [G^{(j)}, G^{(j)}]$ and $G^{(0)} = G$, terminates to the trivial group after a finite number of steps;

(iv) the *unipotent radical* of G is the maximal normal unipotent subgroup of G, and denoted by $\mathrm{R_u}(G)$;

(v) G is *reductive* if its unipotent radical $\mathrm{R_u}(G)$ is trivial.

This is a lot of new terminology, and the reason for us to introduce it is to make the construction of Siegel sets precise. For our purposes, we note that we do not require a detailed understanding of this terminology. Nonetheless, let us illustrate these concepts by some examples.

Consider the group $\mathrm{GL}(n, \mathbb{C})$. An example of a unipotent group is given by the group of upper unitriangular matrices, i.e. the subgroup

$$\mathrm{U}_n = \left\{ g = \begin{pmatrix} 1 & * & \cdots & * & * \\ 0 & 1 & \cdots & * & * \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & * \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix} \middle| \det(g) \neq 0 \right\}. \tag{4.2}$$

Here the symbol $*$ indicates that the entry is unspecified. In a certain sense, this is the most important unipotent group, since it can be shown that any unipotent group can be embedded in U_n for some $n \geq 1$ [71].

The unipotent group U_n has the special property of being solvable. This can be shown directly by computing its derived series, as in the definition above. To illustrate the pattern of this series, we consider U_5 as an example. By starting with $\mathrm{U}_5^{(0)} = \mathrm{U}_5$ and successively computing the commutator subgroups $\mathrm{U}_5^{(j+1)} = [\mathrm{U}_5^{(j)}, \mathrm{U}_5^{(j)}]$, we find that its derived series $\mathrm{U}_5^{(1)}, \mathrm{U}_5^{(2)}, \dots$ is

$$\left\{\begin{pmatrix} 1 & 0 & * & * & * \\ 0 & 1 & 0 & * & * \\ 0 & 0 & 1 & 0 & * \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}\right\}, \quad \left\{\begin{pmatrix} 1 & 0 & 0 & * & * \\ 0 & 1 & 0 & 0 & * \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}\right\}, \quad \left\{\begin{pmatrix} 1 & 0 & 0 & 0 & * \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}\right\}, \quad \left\{\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}\right\}.$$

It terminates to the trivial group at the 4th step in the sequence, and hence it is solvable. In this manner one can show that every U_n is solvable.

[2] Here $[A, B]$ denotes the commutator subgroup of the groups A and B, generated by elements of the form $a\,b\,a^{-1}\,b^{-1}$ with $a \in A$, $b \in B$.

The group U_n is also nilpotent. This follows from the fact any solvable group is nilpotent, because the terms in the derived series are always contained in those of the lower central series. Hence, if the derived series terminates to the trivial group, then so does the lower central series. We remark that this notion of 'nilpotent' is not the same as nilpotency of an operator or matrix, as for instance seen for the log-monodromy operators from Chapter 1. The relation between the two is that for a nilpotent group G, the maps

$$\phi_g : G \to G, \quad h \mapsto g^{-1}h^{-1}gh \tag{4.3}$$

are nilpotent in the sense that the there is an integer $n \geq 1$ such that the nth power ϕ_g^n sends every element to the identity in G.

We should be careful not to confuse U_n, the largest unipotent subgroup of $\mathrm{GL}(n, \mathbb{C})$, with the unipotent radical $\mathrm{R_u}(\mathrm{GL}(n, \mathbb{C}))$, the largest *normal* unipotent subgroup of $\mathrm{GL}(n, \mathbb{C})$. The reason is that the group U_n is not normal. For example, consider two generic elements

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}(n, \mathbb{C}), \quad \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \in \mathrm{U}_2. \tag{4.4}$$

We then compute the conjugate element

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} -bc + ad - acs & a^2 s \\ -c^2 s & -bc + ad + acs \end{pmatrix}, \tag{4.5}$$

and see that this is generically not an element of U_2, implying that U_2 is not normal.

In general, it can be shown that the unipotent radical $\mathrm{R_u}(\mathrm{GL}(n, \mathbb{C}))$ is the trivial group, which means that $\mathrm{GL}(n, \mathbb{C})$ is reductive. Other examples of reductive groups include many of the standard linear algebraic groups, such as

$$\mathrm{SL}(n, \mathbb{C}), \quad \mathrm{SO}(n), \quad \mathrm{O}(n), \quad \text{and} \quad \mathrm{Sp}(n).$$

Reductive groups are mainly of interest due to their rich and well-understood representation theory [71]. For our purposes, they appear in the construction of Siegel sets.

4.1.2 Parabolic Subgroups and the Langlands Decomposition

Having briefly reviewed some essential concepts in algebraic group theory, we are closer to setting up the construction of fundamental sets for arithmetic quotients. The next important notion that we need is that of a parabolic subgroup. The so-called Langlands decomposition that parabolic subgroups possess will provide us with the building blocks of Siegel sets. The results shown in this section can be found in [66, 70], and the textbook by Borel [69] in which many foundations of the subject were laid.

Definition 4.1.3. Let $G_{\mathbb{C}}$ be a connected linear algebraic group. A *parabolic subgroup* of $G_{\mathbb{C}}$ is a closed subgroup $P_{\mathbb{C}} \subseteq G_{\mathbb{C}}$ such that the quotient space $G_{\mathbb{C}}/P_{\mathbb{C}}$ is compact.

Intuitively, this means that $P_{\mathbb{C}}$ contains the unbounded directions in the group $G_{\mathbb{C}}$. There are several equivalent definitions of parabolic subgroups, and this is the one that requires the least amount of technical background to state and understand. The terminology of the previous subsection gives us a more abstract, but useful alternative characterization [71].

Proposition 4.1.4. *Let $G_{\mathbb{C}}$ be a connected linear algebraic group. A closed subgroup $P_{\mathbb{C}} \subseteq G_{\mathbb{C}}$ is parabolic if and only if it contains a Borel subgroup, i.e. a maximal connected solvable group.*

This alternative characterization is particularly useful in cases where the Borel subgroups are known. Let us discuss some examples of parabolic subgroups.

Example 4.1.5. Consider the algebraic group $G_{\mathbb{C}} = \mathrm{SL}(2, \mathbb{C})$. Then it can be shown that the subgroup

$$P_{\mathbb{C}} = \left\{ \begin{pmatrix} t & s \\ 0 & t^{-1} \end{pmatrix} \,\middle|\, t \in \mathbb{C}^*,\, s \in \mathbb{C} \right\} \tag{4.6}$$

is parabolic. We will frequently use this parabolic subgroup as an example in the discussion that follows.

Example 4.1.6. For the group $\mathrm{GL}(n, \mathbb{C})$, consider the upper triangular subgroup

$$\mathrm{B}_n = \left\{ g = \begin{pmatrix} * & * & \cdots & * & * \\ 0 & * & \cdots & * & * \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & * & * \\ 0 & 0 & \cdots & 0 & * \end{pmatrix} \,\middle|\, \det(g) \neq 0 \right\}. \tag{4.7}$$

By computing its derived series, it can be shown that B_n is solvable (in fact, we have $\mathrm{B}_n^{(1)} = \mathrm{U}_n$). It turns out that B_n is the maximal connected subgroup with this property, so that it defines a Borel subgroup. Therefore, by Proposition 4.1.4, B_n is a parabolic subgroup, and so is any closed subgroup containing B_n.

As alluded to above, parabolic subgroups come with a special decomposition into various smaller subgroups, and this decomposition will be of significant importance in the construction of Siegel sets.

Proposition 4.1.7. *Let $G_{\mathbb{C}}$ be a semisimple algebraic group and let $P_{\mathbb{C}} \subseteq G_{\mathbb{C}}$ be a parabolic subgroup. Then the group of real points $P = P_{\mathbb{R}}$ can be decomposed*[2] *as*

$$P = N_P \cdot A_P \cdot M_P, \tag{4.8}$$

where N_P is a nilpotent subgroup, A_P is a connected abelian subgroup, and $M_P \subseteq P$ is a reductive subgroup.

This decomposition is called the *Langlands decomposition* [70]. Let us illustrate it the parabolic subgroup of $\mathrm{SL}(2, \mathbb{C})$.

Example 4.1.8. The group of real points of the parabolic subgroup of $\mathrm{SL}(2, \mathbb{C})$ from Example 4.1.5 is the subgroup

$$P = \left\{ \begin{pmatrix} t & s \\ 0 & t^{-1} \end{pmatrix} \,\middle|\, t \in \mathbb{R}^*,\, s \in \mathbb{R} \right\} \subseteq \mathrm{SL}(2, \mathbb{Z}). \tag{4.9}$$

[3]To be precise, this notation means that the map $M_P \times A_P \times N_P \to P$ is a diffeomorphism.

The Langlands decomposition for P is given by $P = N_P \cdot A_P \cdot M_P$, where

$$N_P = \left\{ \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \,\middle|\, s \in \mathbb{R} \right\}, \quad A_P = \left\{ \begin{pmatrix} t & 0 \\ 0 & t^{-1} \end{pmatrix} \,\middle|\, t > 0 \right\}, \quad M_P = \left\{ \pm \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\}. \tag{4.10}$$

Indeed, M_P is (in a trivial way) reductive, A_P is an abelian group isomorphic to the real multiplicative group $\mathbb{R}_{>0}$, and N_P is nilpotent; it coincides with the group of real points of the upper unitriangular group U_2.

In general, there is an explicit recipe for obtaining the subgroups N_P, A_P and M_P, which we now describe. Let $G_{\mathbb{C}}$ be a semi-simple linear algebraic group, and let $P_{\mathbb{C}} \subseteq G_{\mathbb{C}}$ be a parabolic subgroup which is defined over $\mathbb{Q}$. The group N_P in the Langlands decomposition can be obtained by letting $(N_P)_{\mathbb{C}}$ be the unipotent radical of $P_{\mathbb{C}}$, and taking N_P to be the group of real points of $(N_P)_{\mathbb{C}}$.

Next, let $(L_P)_{\mathbb{C}}$ be the quotient group $(N_P)_{\mathbb{C}} \backslash P_{\mathbb{C}}$, and let L_P denote the group of real points. The group A_P can then by identified as the connected component of center $Z(L_P)$. The resulting group A_P is a connected abelian group isomorphic to a power of the multiplicative group, i.e. $A_P \cong (\mathbb{R}_{>0})^r$ for some $r \geq 0$. We can now consider the subgroup $M_P \subseteq L_P$ defined by the formula

$$M_P = \bigcap_{\chi : L_P \to \mathbb{R}^\times} \ker(\chi^2) \tag{4.11}$$

In words, we take M_P to be the group of elements which are sent to the identity by the squares of all homomorphisms $\chi : L_P \to \mathbb{R}^\times$. It can be shown that M_P is reductive, and that it is complementary to A_P in the sense that $L_P = A_P \cdot M_P$ [69].

Currently, the groups M_P and A_P are defined as subgroups of the quotient group L_P, but in order to make sense of the Langlands decomposition we are required to 'lift' them to subgroups of the parabolic group P. This can be done by fixing a maximal compact subgroup $K \subseteq G$. In the situation of interest, namely arithmetic quotients of period domains, K will be the maximal compact subgroup containing the compact stabilizer H of a reference Hodge structure, which appears in the description of the period domain as $\mathcal{D} = G/H$. The maximal subgroup K comes with a Cartan involution $\theta_K : G \to G$, characterized by the fact that θ_K restricts to the identity on K. The key result is now that there is a unique subgroup of P which is isomorphic to L_P, and has the property that it is stable under the map θ_K [72]. The upshot of this procedure is that we may uniquely identify A_P and M_P with subgroups of P, and that P indeed decomposes as $P = N_P \cdot A_P \cdot M_P$. As a bonus, we obtain a decomposition of the whole group G as

$$G = N_P \cdot A_P \cdot M_P \cdot K, \tag{4.12}$$

known as the *horospherical decomposition.*

To illustrate this procedure, let us reproduce the Langlands decomposition of Example 4.1.8.

Example 4.1.9. The first step is to find the unipotent radical of $P_{\mathbb{C}}$. As we have learned earlier in this section, these can be found by looking inside the group of upper unitriangular matrices. We concluded that this subgroup is not a normal subgroup in $\mathrm{GL}(2,\mathbb{C})$, but inside the smaller group $P_{\mathbb{C}}$ it is in fact normal. We thus recover

$$(N_P)_{\mathbb{C}} = \left\{ \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \;\middle|\; s \in \mathbb{C} \right\}.$$

By selecting the maximal compact subgroup $\mathrm{SO}(2) \subseteq \mathrm{SL}(2,\mathbb{R})$, we may identify the quotient L_P with the subgroup

$$L_P = \left\{ \begin{pmatrix} t & 0 \\ 0 & t^{-1} \end{pmatrix} \;\middle|\; t \in \mathbb{R}^{\times} \right\}$$

Since L_P is abelian, the procedure tells us that A_P is given by the connected component of L_P, which coincides with the subgroup A_P found in the previous example. Finally, the homomorphisms $\chi : L_P \to \mathbb{R}^{\times}$ are given by

$$\begin{pmatrix} t & 0 \\ 0 & t^{-1} \end{pmatrix} \mapsto t^n,$$

for $n \in \mathbb{Z}$. The formula of Equation (4.11) then yields

$$M_P = \bigcap_{n \in \mathbb{Z}} \ker \left(\begin{pmatrix} t & 0 \\ 0 & t^{-1} \end{pmatrix} \mapsto t^{2n} \right) = \left\{ \pm \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\},$$

which is the same result as stated earlier.

4.1.3 Roots and Characters

With the Langlands decomposition at hand, we are almost ready to define Siegel sets. As briefly indicated earlier in this chapter, Siegel sets will be defined as a product of sets with respect to the Langlands decomposition, which are designed to have certain finiteness properties with respect to the action of the arithmetic group Γ. Making this precise requires an understanding of the roots and characters of the group A_P.

First, we note that the group of real points underlying a linear algebraic group is always a Lie group; the smoothness follows from the algebraic structure. This means that the tools from Lie theory are available to study linear algebraic groups. The subgroup A_P acts on P by conjugation, and this gives rise to the adjoint representation of the Lie algebra $\mathfrak{a}_P$ of A_P on the Lie algebra $\mathfrak{n}_P$ of N_P, where an element $\alpha \in \mathfrak{a}_P$ acts on $\nu \in \mathfrak{n}_P$ as

$$\nu \mapsto [\alpha, \nu].$$

Let $\alpha_1, \ldots, \alpha_r$ be a basis of $\mathfrak{a}_P$. Since these elements commute, it is possible to simultaneously diagonalize the α_i and choose a basis $\nu_1, \ldots, \nu_k$ of $\mathfrak{n}_P$ such that

$$[\alpha_i, \nu_j] = \phi_{ji}\, \nu_j, \quad \text{with} \quad \phi_{ji} = \phi_j(\alpha_i) \in \mathbb{R}. \tag{4.13}$$

The linear maps $\phi_j : \mathfrak{a}_P^* \to \mathbb{R}$ which send a basis element α_i to the coefficient $\phi_{ji} = \phi_j(\alpha_i)$ form the *roots* of the action of $\mathfrak{a}_P$ on $\mathfrak{n}_P$. We denote the set of roots by $\Phi(\mathfrak{a}_P, \mathfrak{n}_P)$.

To each root we can associate a *character* of the group A_P, i.e. a group homomorphism $A_P \to \mathbb{R}_{>0}$, as follows. Given a root ϕ_j we define the character χ_j by sending a group element $a \in A_P$ to

$$\chi_j(a) = e^{\phi_j(\log a)}. \tag{4.14}$$

Here $\log : A_P \to \mathfrak{a}_P$ is the inverse of the exponential map. The set of such characters is denoted by $\Phi(A_P, N_P)$. We note that these characters can equivalently be characterized by

$$\chi_j(a)\nu_j = a\nu_j a^{-1}, \tag{4.15}$$

which follows from a simple computation. It can be shown that there exists a unique 'linearly independent' subset of characters $\chi_1, \dots, \chi_r$ such that any character is a positive linear combination of the χ_j [70]. The linear independence of these characters is a common abuse of language, and refers to the linear independence of the corresponding roots $\phi_1, \dots, \phi_r$. These characters are the *simple characters* of P with respect to A_P, and they are denoted by $\Delta(A_P, N_P) \subseteq \Phi(A_P, N_P)$.

Example 4.1.10. By inspecting the groups A_P and N_P appearing in the Langlands decomposition of the parabolic subgroup $P \subseteq \mathrm{SL}(2, \mathbb{R})$ from Example 4.1.8, we see that the Lie algebras $\mathfrak{a}_P$ and $\mathfrak{n}_P$ are spanned by

$$\alpha = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \nu = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \tag{4.16}$$

respectively. We have $[\alpha, \nu] = 2\nu$, so there is only one root ϕ, which is given by $\phi(\alpha) = 2$. The corresponding character is then computed to be

$$\chi : \begin{pmatrix} t & 0 \\ 0 & t^{-1} \end{pmatrix} \mapsto t^2. \tag{4.17}$$

Since this is the only character, the set of simple characters $\Delta(A_P, N_P)$ consists of just χ.

4.1.4 Siegel Sets and Tameness of Arithmetic Quotients

We are finally ready to state the definition that we were after.

Definition 4.1.11. Let $G_{\mathbb{C}}$ be a semi-simple linear algebraic group, let $P_{\mathbb{C}} \subseteq G_{\mathbb{C}}$ be a parabolic subgroup with Langlands decomposition $P = M_P \cdot A_P \cdot N_P$. Let $K \subseteq G$ be a maximal compact subgroup such that G has a horospherical decomposition $G = M_P \cdot A_P \cdot N_P \cdot K$. A *Siegel set* of G associated to P and K is a subset of G of the form

$$\mathfrak{S} = U \cdot A_{P,\lambda} \cdot V, \tag{4.18}$$

where $U \subseteq N_P$ and $V \subseteq M_P \cdot K$ are any bounded sets, and for $\lambda > 0$, the subset $A_{P,\lambda} \subseteq A_P$ is defined by

$$A_{P,\lambda} = \{a \in A_P \,|\, \chi(a) > \lambda \text{ for all simple characters } \chi \in \Delta(A_P, N_P)\}. \tag{4.19}$$

If $\mathfrak{S} \subseteq G$ is a Siegel set, then its image in the quotient space G/H is also called a Siegel set[4]. Siegel sets satisfy the following crucial finiteness properties [2, 70].

[4]Here it is required that the maximal compact subgroup K contains a $G_{\mathbb{Q}}$-conjugate of H.

Theorem 4.1.12 (Siegel Sets)**.** *Let $G_{\mathbb{C}}$ be a semi-simple linear algebraic group, let $H \subseteq G$ be a connected compact subgroup and let $\Gamma \subseteq G_{\mathbb{Z}}$ be an arithmetic subgroup.*

(i) There are finitely many Γ-conjugacy classes of parabolic subgroups defined over $\mathbb{Q}$. Letting $P_1, \ldots, P_k$ denote a set of representatives, there exist Siegel sets $\mathfrak{S}_i \subseteq G/H$ associated to P_i, such that

$$\Gamma \cdot \bigcup_{i=1}^{k} \mathfrak{S}_i = G/H. \tag{4.20}$$

(ii) For any two Siegel sets $\mathfrak{S}_i$, $\mathfrak{S}_j$ associated to parabolic subgroups P_i and P_j, we have

$$\big|\{\gamma \in \Gamma \,|\, \gamma \cdot \mathfrak{S}_i \cap \mathfrak{S}_j \neq \emptyset\}\big| < \infty, \tag{4.21}$$

i.e. there are only finitely many group elements $\gamma \in \Gamma$ for which the translate $\gamma \cdot \mathfrak{S}_i$ intersects $\mathfrak{S}_j$. This includes the case $i = j$.

(iii) If P_i is not Γ-conjugate to P_j, and $\mathfrak{S}_i = U_i \cdot A_{P_i,\lambda_i} \cdot V_i$ and $\mathfrak{S}_j = U_j \cdot A_{P_j,\lambda_j} \cdot V_j$ are two Siegel sets, then

$$\gamma \mathfrak{S}_i \cap \mathfrak{S}_j \quad \textit{for all } \gamma \in \Gamma \tag{4.22}$$

for λ_i and λ_j sufficiently large.

The modern version of these results stated here can be found in [70], and their original proofs are given in the classic textbook by Borel [69]. Discussing these proofs is beyond the scope of this thesis.

Comparing to Section 3.4.1, we see that these are precisely the sort of properties that we need in the construction of fundamental sets. Indeed, we can choose the $U_i \subseteq N_{P_i}$ and $V_i \subseteq M_{P_i} \cdot K$ to be semi-algebraic, and the A_{P_i,λ_i} are semi-algebraic by construction. The group multiplication is algebraic, and in Example 3.4.9 we found that the projection map $G \to G/H$ is $\mathbb{R}_{\mathrm{alg}}$-definable. The resulting Siegel sets $\mathfrak{S}_i \subseteq G/H$ are thus definable in $\mathbb{R}_{\mathrm{alg}}$. By the theorem above, the union

$$F = \bigcup_{i=1}^{k} \mathfrak{S}_i \tag{4.23}$$

then satisfies $\Gamma \cdot F = G/H$ and $|\Gamma_F| = \big|\{\gamma \in \Gamma \,|\, \gamma \cdot F \cap F \neq \emptyset\}\big| < \infty$. In other words, the projection map $F \to \Gamma \backslash G/H$ is surjective and has finite fibers, which are the defining properties of a fundamental set F. Through Proposition 3.4.12, we thus find the following recent result of Bakker, Klingler and Tsimerman on the tame geometry of arithmetic quotients [2].

Theorem 4.1.13 (Definability of Arithmetic Quotients)**.** *Let $G_{\mathbb{C}}$ be a connected semi-simple linear algebraic group defined over $\mathbb{Q}$, and let G be the connected component of its group of real points. Furthermore, let $H \subseteq G$ be a connected compact subgroup and let $\Gamma \subseteq G_{\mathbb{Z}}$ be an arithmetic subgroup. Then the arithmetic quotient $\Gamma\backslash G/H$ has a natural $\mathbb{R}_{\text{alg}}$-definable structure, and there is a fundamental set $F \subseteq G/H$ such that the projection map $F \to \Gamma\backslash G/H$ is definable in $\mathbb{R}_{\text{alg}}$. Moreover, for any Siegel set $\mathfrak{S} \subseteq G/H$, the projection map $\mathfrak{S} \to \Gamma\backslash G/H$ is $\mathbb{R}_{\text{alg}}$-definable.*

In addition to proving that arithmetic quotients have a natural definable structure, it is shown in [2] that that this structure is 'functorial', in the sense that maps between arithmetic quotients are also $\mathbb{R}_{\text{alg}}$-definable. More precisely, a *morphism of arithmetic quotients* is a real analytic map

$$\Gamma\backslash G/H \to \tilde{\Gamma}\backslash\tilde{G}/\tilde{H} \tag{4.24}$$

which sends a double coset $\Gamma h H$ to a double coset of the form $\tilde{\Gamma}\phi(h)g\tilde{H}$, where $\phi : G_{\mathbb{C}} \to \tilde{G}_{\mathbb{C}}$ is a group homomorphism and $g \in G$ is an element such that $\phi(K) \subseteq g\tilde{H}g^{-1}$ and $\phi(\Gamma) \subseteq \tilde{\Gamma}$. It is then shown in [2] that any morphism of arithmetic quotients is also definable in $\mathbb{R}_{\text{alg}}$. We will make use of this fact in Chapter 5 and Chapter 6.

For our purposes, the lesson that we extract from this theorem is that the target space of the period map of a variation of Hodge structure is $\mathbb{R}_{\text{alg}}$-definable. This is our initial motivation to study arithmetic quotients, but in Chapter 6 we will find that arithmetic quotients also appear as moduli spaces in higher supergravity theories, which provides an additional application of this theorem.

Before we move on to discuss period maps, let us discuss an example of a Siegel set.

Example 4.1.14. Consider again the group $\mathrm{SL}(2,\mathbb{R})$, with maximal compact subgroup $\mathrm{SO}(2)$. Then we have the horospherical decomposition

$$\mathrm{SL}(2,\mathbb{R}) = N_P \cdot A_P \cdot M_P \cdot \mathrm{SO}(2), \tag{4.25}$$

with N_P, A_P, and M_P as above. As an arithmetic subgroup we take $\mathrm{SL}(2,\mathbb{Z})$. In the previous example we saw that A_P has only one simple character, so we have

$$A_{P,\lambda} = \left\{ \begin{pmatrix} t & 0 \\ 0 & t^{-1} \end{pmatrix} \,\middle|\, t > 0,\, t^2 > \lambda \right\}. \tag{4.26}$$

For the bounded sets U and V we may then select

$$U = \left\{ \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \,\middle|\, -1 < s < 1 \right\} \subseteq N_P, \quad V = \mathrm{SO}(2) \subseteq M_P \cdot \mathrm{SO}(2). \tag{4.27}$$

Fixing a number $\lambda > 0$, the resulting Siegel set in the upper-half plane $\mathbb{H} = \mathrm{SL}(2,\mathbb{R})/\mathrm{SO}(2)$ is the set

$$\mathfrak{S} = \{x + iy \in \mathbb{H} \,|\, -1 < x < 1,\, y > \lambda\}. \tag{4.28}$$

To compare, the standard fundamental domain for the action of $\mathrm{SL}(2,\mathbb{Z})$ on $\mathbb{H}$ is given by

$$\mathcal{F} = \left\{x + iy \in \mathbb{H} \,|\, -\tfrac{1}{2} < x < \tfrac{1}{2},\, |x+iy| > 1\right\}. \tag{4.29}$$

For a sufficiently small choice of λ, the Siegel set $\mathfrak{S}$ contains the fundamental domain $\mathcal{F}$. These two sets are illustrated in Figure 4.1. In this figure we see that the Siegel set $\mathfrak{S}$ is coarser than the fundamental domain $\mathcal{F}$, but with the advantage of having a relatively simple shape.

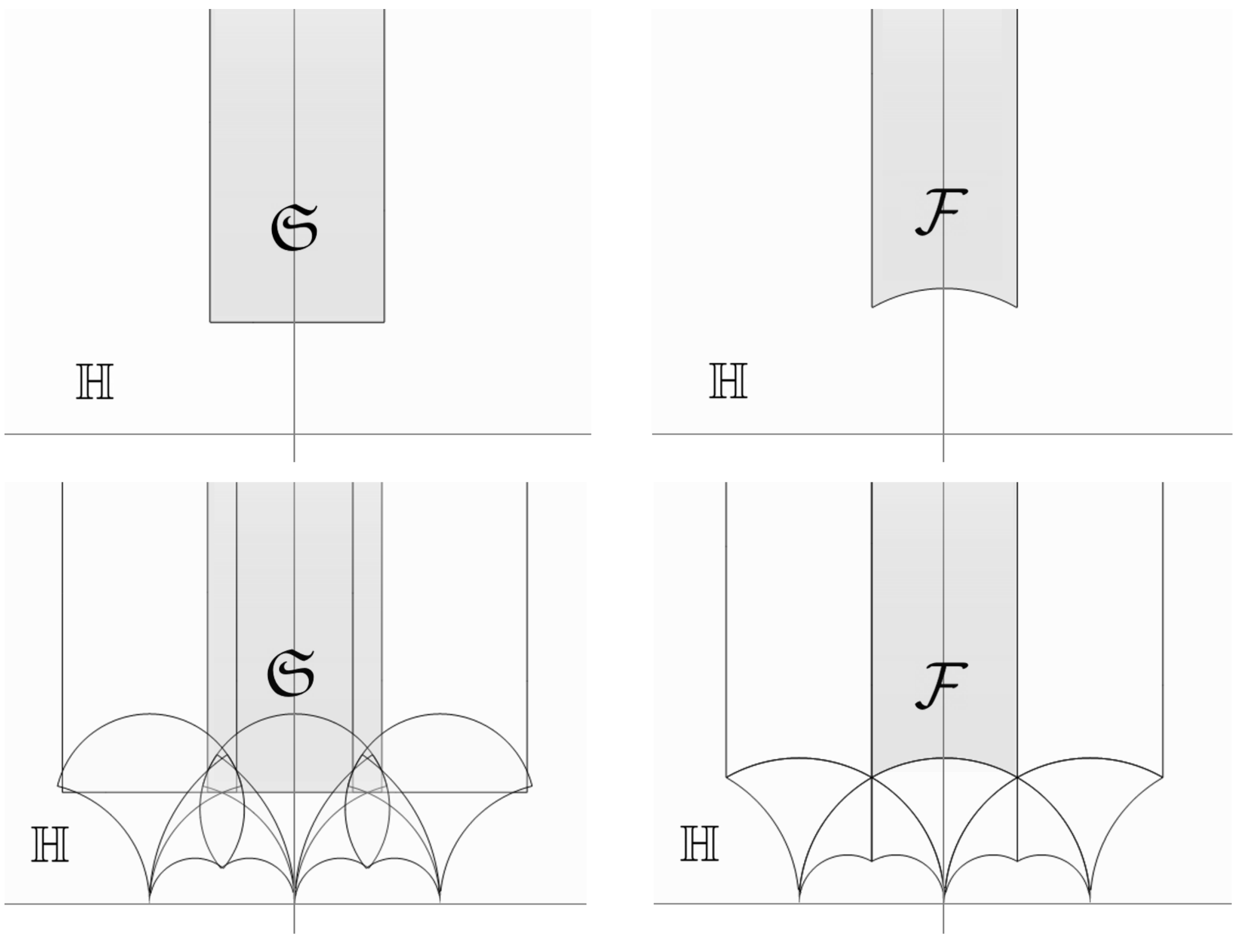


Figure 4.1: Comparison of a Siegel set $\mathfrak{S}$ and the standard fundamental domain $\mathcal{F}$. For illustrative purposes, the width of $\mathfrak{S}$ is taken to be close to the width of the fundamental domain. The upper figures show $\mathfrak{S}$ and $\mathcal{F}$, and the lower figures show $\mathfrak{S}$ and $\mathcal{F}$ together with their neighbouring translates under the action of $\mathrm{SL}(2,\mathbb{Z})$. The fundamental domain $\mathcal{F}$ does not intersect its neighbours, and $\mathfrak{S}$ intersects its neighbours in a complicated way.

Although the construction of Siegel sets is rather technically involved, the idea that Siegel sets have a relatively simple shape should be kept in mind in general. Compared to the possibly complicated structure of the algebraic group G, Siegel sets have the relatively simple structure as the product of a cone $A_{P,\lambda}$ and two bounded sets U and V.

4.2 Tameness of Period Maps

We have now seen that the target space of a period map $\Phi : \mathcal{M} \to \Gamma\backslash G/H$ has a natural $\mathbb{R}_{\text{alg}}$-definable structure. This brings us closer to the main result of this chapter, proven by Bakker, Klingler and Tsimerman [2]:

> **Theorem 4.2.1.** *Let $(\mathcal{H}_{\mathbb{Z}}, q, \mathcal{F}^\bullet)$ be a variation of Hodge structure over a smooth complex quasi-projective variety $\mathcal{M}$, and let $\Gamma \subseteq G$ be an arithmetic subgroup containing the monodromy group. Then the associated period map $\Phi : \mathcal{M} \to \Gamma\backslash G/H$ is definable in the o-minimal structure $\mathbb{R}_{\text{an,exp}}$.*

Through the orbit theorems of asymptotic Hodge theory discussed in Section 1.5, we already saw that the singularities of the period map Φ are 'tame', and this is the result that makes the notion of tameness precise. Remarkably, this result distills much of the complicated machinery of the orbit theorems to a single statement. Although the proof is long and technical, the essence of the proof can be understood intuitively. In summary, the idea behind the proof is as follows.

Summary of proof of Theorem 4.2.1.

(i) **Reduction to a local statement**. The period map Φ is holomorphic, so its restriction to any compact subset of $\mathcal{M}$ is restricted analytic and hence definable in $\mathbb{R}_{\text{an,exp}}$. The difficulty therefore lies in the non-compact, asymptotic regions of $\mathcal{M}$, which are locally of the form $(\Delta^*)^l \times \Delta^{k-l}$.

(ii) **Definability of lifted period map**. Using the results of asymptotic Hodge theory, we can show that the lift of the period map $\widetilde{\Phi} : \mathbb{H}^l \times \Delta^{k-l} \to G/H$ is definable on a suitable subset of $\mathbb{H}^l \times \Delta^{k-l}$.

(iii) **Image of period map and Siegel sets**. To show that the definability of $\widetilde{\Phi}$ descends to Φ, we have to project back down to the arithmetic quotient $G/H \to \Gamma\backslash G/H$. Since the full projection map is not definable, we have to show that the image of the period map is contained in a finite number of Siegel sets, on which the projection to $\Gamma\backslash G/H$ is then definable. This is the most technical part of the proof.

The aim of the remainder of this section is to discuss the proof of the tameness of the period map in more detail. We focus on step (i) and (ii).

4.2.1 Reduction to a Local Statement

The first step of the proof is to look at the period map locally. As we did in the previous chapters, we view $\mathcal{M}$ as a subvariety of a larger variety $\overline{\mathcal{M}}$, and by Hironaka's theorem we may assume that the inclusion $\mathcal{M} \subseteq \overline{\mathcal{M}}$ is of the form $(\Delta^*)^l \times \Delta^{m-l} \subseteq \Delta^m$ [19]. Since $\mathcal{M}$ is algebraic, we can cover $\mathcal{M}$ by finitely many such subsets. In Chapter 3 we learned that definability of a map can be checked locally, as long as we have a *finite* cover of definable subsets. Hence, to prove the theorem, we may forget about the ambient space $\mathcal{M}$ and focus on the local period map defined on $(\Delta^*)^l \times \Delta^{m-l}$. The situation may be simplified further by taking $l = 0$; this does not change any of the arguments that follow. Thus, we consider a variation of Hodge structure on the punctured polydisk $(\Delta^*)^m$ and look at the period map

$$\Phi : (\Delta^*)^m \to \Gamma\backslash G/H. \tag{4.30}$$

In addition, we may assume that the monodromy around the punctures is unipotent by using a finite covering map $\Delta^* \to \Delta^*$, as done in Section 1.5.

4.2.2 Definability of Lifted Period Map

With our reduction to a local period map defined on a punctured polydisk $(\Delta^*)^m$, we are in the setting of the nilpotent orbit theorem. In what follows, we will use the same notation as established in Section 1.5. The period map lifts to a map on the universal cover

$$\widetilde{\Phi} : \mathbb{H}^m \to G/H. \tag{4.31}$$

By the nilpotent orbit theorem, $\widetilde{\Phi}$ takes the form

$$\widetilde{\Phi}(t_1, \ldots, t_m) = \big(g_N(t_1, \ldots, t_m)\big)^{-1} \cdot \widetilde{\Psi}(t_1, \ldots, t_m) = \big(g_N(t_1, \ldots, t_m)\big)^{-1} \cdot \Psi(e^{2\pi i t_1}, \ldots, e^{2\pi i t_m}), \tag{4.32}$$

where $g_N : \mathbb{H}^k \to G_{\mathbb{C}}$ was the nilpotent orbit factor defined by

$$g_N(t_1, \ldots, t_m) := \exp\left(-\sum_{j=1}^{m} t_j N_j\right), \tag{4.33}$$

and $\Psi : \Delta^m \to \check{\mathcal{D}}$ is a holomorphic map defined on the whole polydisk Δ^m. In equation (4.32), we traded the lifted map $\widetilde{\Psi}$ for Ψ, following the commutative diagram

$$\begin{tikzcd} \mathbb{H}^m \arrow[d, "p"] \arrow[dr, "\widetilde{\Psi}"] & \\ \Delta^m \arrow[r, "\Psi"] & \check{\mathcal{D}}. \end{tikzcd} \tag{4.34}$$

Here p is the covering map $(t_1, \ldots, t_n) \mapsto (e^{2\pi i t_1}, \ldots, e^{2\pi i t_m})$. A crucial observation now is that Ψ is holomorphic on the compact set Δ^m, meaning that it is restricted analytic on a disk of slightly smaller radius[5] and therefore definable in the o-minimal structure $\mathbb{R}_{\mathrm{an}}$. However, the map p is not definable in any o-minimal structure since it involves the complex exponential function. On the other hand, since we are using the upper half-plane $\mathbb{H}^m$ as a covering space for $(\Delta^*)^m$, we do not require the full space; a bounded strip large enough to cover $(\Delta^*)^m$ will suffice. To this end, we define the strip

$$\Sigma = \big\{x + iy \in \mathbb{H} \,\big|\, -1 < x < 1,\ y > 1\big\}. \tag{4.35}$$

We now claim that the restriction $p : \Sigma^m \to \Delta^m$ is definable in $\mathbb{R}_{\mathrm{an,exp}}$. To see this, write $t_j = x_j + iy_j$, and note that

$$x_j + iy_j \mapsto e^{2\pi i(x_j+iy_j)} = e^{2\pi i x_j} e^{-2\pi y_j}.$$

The factor $e^{-2\pi y_j}$ is the real exponential, which is definable in $\mathbb{R}_{\exp}$. On the other hand, the factor $e^{2\pi i x_j}$ is now only defined for $|x_j| < 1$, so that it is restricted analytic and thus definable in $\mathbb{R}_{\mathrm{an}}$. It thus follows that the map

$$(t_1, \ldots, t_m) \mapsto \Psi(e^{2\pi i t_1}, \ldots, e^{2\pi i t_m}) \tag{4.36}$$

is definable on the set $\Sigma^m \subseteq \mathbb{H}^m$, since it is the composition of two definable functions.

The nilpotent orbit factor $g_N : \mathbb{H}^m \to G_{\mathbb{C}}$ is definable, because it is a polynomial in the t_j since the log-monodromy operators N_j are nilpotent. Finally, the group action of g_N^{-1} on Ψ is definable, since the group $G_{\mathbb{C}}$ acts algebraically on $\check{\mathcal{D}} = G_{\mathbb{C}}/B$. Altogether, we obtain the following lemma, which is a major step in proving the definability of the period map.

[5] Recall that restricted analytic functions are required to be defined on an open set containing the compact set to which they are restricted. In our case, this means that we have to shrink the disk slightly to ensure that Ψ is restricted analytic. This does not matter for the argument, since the non-trivial part of the proof is to show that Φ is definable near the puncture.

Lemma 4.2.2. *The lifted period map* $\widetilde{\Phi} : \Sigma^m \to G/H$ *is definable in* $\mathbb{R}_{\mathrm{an,exp}}$.

4.2.3 Image of Period Map and Siegel Sets

The present situation is summarized by the commutative diagram

$$\begin{array}{ccc} \Sigma^m & \xrightarrow{\widetilde{\Phi}} & G/H \\ \downarrow p & & \downarrow \pi \\ (\Delta^*)^m & \xrightarrow{\Phi} & \Gamma\backslash G/H \end{array} \tag{4.37}$$

The next step is to pass from G/H to the arithmetic quotient $\Gamma\backslash G/H$. This is a highly non-trivial step, since as we have seen, the quotient map π is not definable due to the infinite discrete nature of Γ.

In the previous section we learned that π is definable when restricted to a Siegel set. Thus, in order to project down to the arithmetic quotient, we have to control the image of the period map. The starting point is the following result of Schmid, which is proven as a consequence of the single-variable SL(2)-orbit theorem [15].

Proposition 4.2.3. *Assume that* $m = 1$. *For given constants* $C > 0$, $\eta > 0$, *consider the subset*

$$\mathbb{H}_{C,\eta} := \left\{ t \in \mathbb{H} \,\middle|\, |\mathrm{Re}\, t| \leq C,\ \mathrm{Im}\, t \geq \eta \right\}. \tag{4.38}$$

Then there exists a Siegel set $\mathfrak{S} \subseteq G/H$ *such that* $\widetilde{\Phi}(\mathbb{H}_{C,\eta}) \subseteq \mathfrak{S}$.

This is precisely the sort of result that we need, except for the fact that it is formulated in a single-variable case. The generalization to multiple variables was proven[6] in [2].

Proposition 4.2.4. *For given constants* $C > 0$, $\eta > 0$, *consider the subset*

$$\mathbb{H}^m_{C,\eta} := \left\{ (t_1, \ldots, t_m) \in \mathbb{H}^n \,\middle|\, \max_j\{|\mathrm{Re}\, t_j|\} \leq C,\ \min_j\{\mathrm{Im}\, t_j\} \geq \eta \right\}. \tag{4.39}$$

Then there exists finitely many Siegel sets $\mathfrak{S}_i \subseteq G/H$, $i = 1, \ldots, l$ *such that*

$$\widetilde{\Phi}(\mathbb{H}^m_{C,\eta}) \subseteq \bigcup_{i=1}^{l} \mathfrak{S}_i. \tag{4.40}$$

Before we proceed, let us indicate how the definability of the period map follows from this proposition. The situation is now summarized by the commutative diagram

$$\begin{array}{ccc} \mathbb{H}^m_{C,\eta} & \xrightarrow{\widetilde{\Phi}} & \bigcup_{i=1}^{l} \mathfrak{S}_i \\ \downarrow p & & \downarrow \pi \\ (\Delta^*)^m & \xrightarrow{\Phi} & \Gamma\backslash G/H. \end{array} \tag{4.41}$$

[6] It is interesting to note that the generalization to multiple variables was stated and proven 45 years after the original version.

Since the projection map π is now restricted to a finite union of Siegel sets, it is definable by Theorem 4.1.13. The graph $\Gamma(\Phi) \subseteq (\Delta^*)^m \times \Gamma\backslash G/H$ now coincides[7] with the projection under

$$p \times \mathrm{id} : \mathbb{H}^m_{C,\eta} \times \Gamma\backslash G/H \to (\Delta^*)^m \times \Gamma\backslash G/H \tag{4.42}$$

of the graph $\Gamma(\pi \circ \widetilde{\Phi})$. The latter is definable, since it is the graph of the composition of the $\mathbb{R}_{\mathrm{an,exp}}$-definable maps. Thus, $\Gamma(\Phi)$ is definable, and we conclude that the period map Φ is definable in $\mathbb{R}_{\mathrm{an,exp}}$, assuming Proposition 4.2.4.

The generalization of Proposition 4.2.3 to a multi-variable version is the most technical part of the proof [2]. Explaining it in full detail would require a discussion which is too long for the purposes of this thesis, so instead we outline the main idea.

The first step is to consider the space $\mathcal{Q}$ of positive definite symmetric bilinear forms on $(H_0)_\mathbb{R}$. This space is a symmetric space which can be obtained as a quotient of $\mathrm{SL}((H_0)_\mathbb{R})$. Any polarized Hodge structure on H_0 naturally comes with such a symmetric bilinear form, namely the Hodge form h. Recall that this form is defined by

$$h(u, v) = q(u, Cv)), \tag{4.43}$$

where q is the polarization form and C is the Weil operator. For a variation of Hodge structure, every point in the period domain $\mathcal{D} = G/H$ defines such a form, so we obtain a natural map

$$G/H \to \mathcal{Q}.$$

Siegel sets for the space $\mathcal{Q}$ turn out to be well-understood; in fact, the classical reduction theory of positive definite bilinear forms, in which $\mathcal{Q}$ plays a central role, is the original motivation for introducing Siegel sets [73]. Roughly speaking, the Siegel sets of $\mathcal{Q}$ correspond to subsets of bilinear forms which are *reduced*. The Siegel sets of $\mathcal{Q}$ and G/H turn out to be related, and via this observation it is argued that we have to show that the image of the composition

$$\mathbb{H}^m_{C,\eta} \xrightarrow{\widetilde{\Phi}} G/H \longrightarrow \mathcal{Q}$$

is contained in a finite union of Siegel sets in $\mathcal{Q}$ [2]. This reduces to problem to a question about the variation of the Hodge form h_t as t ranges over $\mathbb{H}^m_{C,\eta}$. The asymptotic behavior of this variation is known, as a consequence of the single-variable SL(2)-orbit theorem [74]. In particular, the Hodge form turns out to be 'roughly polynomial', in a precise sense. If we let t vary over a curve[8], then together with Schmid's one-variable result stated in Proposition 4.2.3 it can be shown that the image of the Hodge form lies in a single Siegel set. Extending this to the case where t varies over the whole space is the final insight of [2], where it is proven that one can always reduce to curves.

[7] To be precise, we may have to shrink the disk $(\Delta^*)^m$, but this is an unimportant detail since the challenge is to prove that Φ is definable around the punctures.

[8] An important detail here is that it is also required to impose an ordering of the real coordinates $y_1, \ldots y_m$ on the upper-half plane. Since there are only finitely many such orderings, the finiteness result continues to hold.

4.3 The Hodge Conjecture Revisited

To conclude this chapter, we briefly discuss an application of the tameness of the period map. The details of this discussion are technical, and we only provide the main ideas, referring to [2, 59] for a complete explanation. At the end of Chapter 1 we introduced the Hodge Conjecture, and we mentioned that the strongest evidence for this conjecture is the statement that for any variation of Hodge structure, the Hodge locus

$$\mathrm{HL}(\mathcal{M}) = \{z \in \mathcal{M} \mid (\mathcal{H}_{\mathbb{C}})_z \text{ contains a non-zero Hodge class}\}.$$

is a countable union of algebraic varieties. This result was proven unconditionally by Cattani, Deligne and Kaplan using the powerful but complicated multi-variable SL(2)-orbit theorem [22].

In general, it can be shown that the Hodge locus has the structure of a countable union

$$\mathrm{HL}(\mathcal{M}) = \bigcup_{\text{special } \mathcal{S} \subseteq \Gamma\backslash G/H} \Phi^{-1}(\mathcal{S}), \tag{4.44}$$

where the union is taken over preimages of certain *special* subvarieties $\mathcal{S}$ of $\Gamma\backslash G/H$. We will not give their precise definition, but we note that they are subvarieties of the arithmetic quotient defined from a reference Hodge structure on H_0. The preimages $\Phi^{-1}(\mathcal{S})$ are known to be analytic subvarieties of $\mathcal{M}$. Since the period map is $\mathbb{R}_{\text{an,exp}}$-definable, each preimage $\Phi^{-1}(\mathcal{S})$ is in fact definable in $\mathbb{R}_{\text{an,exp}}$. The key observation made by Bakker, Klingler and Tsimerman is now that this implies that every $\Phi^{-1}(\mathcal{S})$ is *algebraic* by the definable Chow theorem, discussed in Section 3.5. Thus, the *algebraicity of Hodge loci* is recovered as a relatively simple consequence of the tameness of the period map [2].

This is no new result, but since it avoids the explicit use of the intricate multi-variable SL(2)-orbit theorem, it is regarded as a worthwhile simplification of the proof. Schematically, we may view the situation as follows:

$$\begin{tikzcd}
\text{asymptotic Hodge theory} \arrow[d, Rightarrow] \arrow[dr, Rightarrow] & \text{Hodge Conjecture} \arrow[d, Rightarrow] \\
\text{tameness of period map} \arrow[r, Rightarrow] & \text{algebraicity of Hodge loci.}
\end{tikzcd}$$

For this reason, it is sometimes said that, for certain applications, asymptotic Hodge theory can be 'replaced' by the tameness of the period map [75].

Earlier in this thesis we have learned that Hodge theory provides a useful tool in compactifications of string theory. The results of this chapter are therefore the first signals that tameness has applications to string theory as well. This is the topic of the next chapter.

Chapter 5

Tame Geometry in String Theory

So far we have seen how tame geometry has provided a natural description for two of the primary objects in Hodge theory, namely the arithmetic quotients of period domains and the period maps. In this chapter, we turn to physics again. Our aim is to explain how, quite remarkably, tame geometry is a natural description for certain objects in the string theory landscape as well. The objects considered in this chapter are closely related to the period map and based on recent developments [1,4].

The first object of interest will be the locus of flux vacua $\mathscr{F}_X(t_0)$ discussed in Chapter 2. Using the techniques and results from [2] presented in Chapter 4 together with a remarkable result in lattice theory, a new finiteness result for the so-called locus of self-dual classes in a variation of Hodge structure has been proven in [4]. This result may be regarded as a generalization of the algebraicity of Hodge loci. A special case of this result implies that the locus of flux vacua is definable in $\mathbb{R}_{\mathrm{an,exp}}$, which sheds new light on the structure of the flux landscape. This result, and most of the proof, will be described in this chapter.

Secondly, we have seen in Chapter 2 that in the setting of Calabi-Yau compactifactions, the theory of variations of Hodge structure naturally emerges when considering the complex structure moduli. It was also noted that string vacua give rise to effective field theories, whose description strongly depends on the geometry of the internal space. Since the period map encodes much of this geometry, many physical coupling functions that appear in the low-energy effective field theory can be expressed in terms the period map. This reveals that in a broad class of string vacua, these physical functions are tame. In the spirit of the swampland program, it is then natural to conjecture that tameness in the form of o-minimality is a more general feature of effective field theories arising in the landscape. This leads us to the recent proposal of a Tameness Conjecture [1], whose formulation we review. In this chapter we will discuss the top-down evidence for this conjecture given in [1] by proving that the physical functions of interest are indeed tame for many string vacua.

We conclude the chapter with an outlook on bottom-up evidence for the Tameness Conjecture, and suggest potentially promising paths to take for finding such evidence.

5.1 Definability of the Self-Dual Locus

5.1.1 Self-Dual Locus

First, let us recall that in Definition 2.4.4, we defined the locus of flux vacua as the set

$$\mathscr{F}_X(t_0) = \big\{(z, G_4) \in \mathcal{H}_{\mathbb{C}} \,\big|\, G_4 \text{ is integral, self-dual, and } q_z(G_4, G_4) = t_0\big\}, \tag{5.1}$$

where $\mathcal{H}_{\mathbb{C}}$ is the Hodge bundle over the complex structure moduli space $\mathcal{M}$ of a Calabi-Yau fourfold X. Points in this locus are global minima of the scalar potential of the complex structure moduli, and therefore represent string vacua in flux compactifications. Here t_0 was a rational number encoding the tadpole condition, depending on the number of D3-branes and the Euler characteristic of X (cf. equation 2.26). The self-duality of G_4 refers to the Weil operator C_z.

In this section we will be more general, and consider an arbitrary integral variation of Hodge structure over a smooth quasi-projective variety $\mathcal{M}$. In this more general context, we will consider the *self-dual locus*

$$\mathscr{S}_{\mathcal{H}}(t_0) := \big\{(z, v) \in \mathcal{H}_{\mathbb{C}} \,\big|\, v \text{ is integral, self-dual, and } q_z(v, v) = t_0\big\}, \tag{5.2}$$

where $t_0 \geq 1$ is a fixed positive integer. The locus of flux vacua is recovered as a special case of the self-dual locus[1].

Inspired by questions of finiteness for the locus of flux vacua, Bakker, Grimm, Schnell, and Tsimerman have proven the following theorem in [4].

> **Theorem 5.1.1.** *Let $(\mathcal{H}_{\mathbb{Z}}, q, \mathcal{F}^\bullet)$ be a variation of Hodge structure over a smooth complex algebraic variety $\mathcal{M}$. For any $t_0 \geq 1$, the self-dual locus $\mathscr{S}_{\mathcal{H}}(t_0)$ is definable in $\mathbb{R}_{\text{an,exp}}$.*

This is a remarkable result, since the integrality condition on the self-dual classes means that $\mathscr{S}_{\mathcal{H}}(t_0)$ is a definable subset of a lattice bundle $\mathcal{H}_{\mathbb{Z}}$. The fibers of this lattice bundle are infinite discrete sets and thus far from definable in any o-minimal structure. Since $\mathscr{S}_{\mathcal{H}}(t_0)$ is definable, for any $z \in \mathcal{M}$ the subset

$$(\mathscr{S}_{\mathcal{H}}(t_0))_z = \{v \in (\mathcal{H}_{\mathbb{C}})_z \,|\, v \text{ is integral, self-dual, and } q_z(v, v) = t_0\} \tag{5.3}$$

is a definable subset of $(\mathcal{H}_{\mathbb{C}})_z$, and because it is discrete it is in fact finite. In Section 3.3.3 we encountered definable families, and viewing the self-dual locus as a definable family we conclude by the trivialization theorem that this finiteness holds uniformly in $z \in \mathcal{M}$. That is, there exist a global upper bound M_0 on the amount of integral self-dual classes v with a fixed value of $q(v, v)$ in the fibers $(\mathcal{H}_{\mathbb{C}})_z$. This is precisely the statement that the scenario sketched in Figure 2.3 does not happen.

[1]Note that t_0 was allowed to be rational for the locus of flux vacua, but since the pairing q is integral, t_0 can be restricted to be an integer without loss of generality.

5.1.2 Finiteness of Flux Vacua

Specializing to the case of flux vacua in Calabi-Yau compactifications, we deduce the following.

> **Observation 5.1.2.** For a fixed Calabi-Yau fourfold X, the landscape of flux vacua $\mathscr{F}_X$ has a tame geometry. Moreover, there is a uniform bound M_0 on the number of flux vacua that lie at any fixed point z in the complex structure moduli space $\mathcal{M}$. This bound only depends on X.

This result, highlighted in [1], answers the long-standing question of whether there is a finite number of flux vacua [5, 6]. Over the years, there have been several approaches to address this problem. For instance, in [76] a statistical method for analyzing flux vacua is developed. The idea of this method is to introduce a density measure on the moduli space $\mathcal{M}$, which has led to estimates that provide evidence for the finiteness of flux vacua. The first step towards a mathematical proof was taken in [38], where the main idea was that the moduli space $\mathcal{M}$ has a certain holographic duality, inspired by the single-variable SL(2)-orbit theorem. This idea led to a proof in the restricted case where the singular locus $\mathcal{M}_{\text{sing}}$ has codimension one. The results of [1, 4] are based on new techniques coming from tame geometry and provide the first complete mathematical proof.

Let us comment on the interpretation of the finiteness of flux vacua. Theorem 5.1.1 does *not* say that the locus of flux vacua $\mathscr{F}_X$ is a finite set. The locus $\mathscr{F}_X$ may consist of several positive-dimensional connected analytic subvarieties of the Hodge bundle $\mathcal{H}_{\mathbb{C}}$. Physically, vacua lying along the same subvariety may be identified as a single vacuum, since the flatness of the scalar potential implies that they can be continuously connected without changing the energy. In other words, there is no energy barrier that enables us to distinguish points on the same subvariety. This is similar to the notion of finiteness considered in [77]. The definability of $\mathscr{F}_X$ implies that it has a finite number of connected components, and hence there is a finite number of such classes of vacua. Another interpretation of the finiteness comes from the uniform bound M_0 on the size of the sets $(\mathscr{F}_X)_z$. From this point of view, the theorem says that there is a bound on the number of physically permissible fluxes that we can turn on.

5.1.3 Outline of Proof

Let us now turn to the proof of Theorem 5.1.1. We begin by recalling some notation used earlier. We consider a variation of Hodge structure $(\mathcal{H}_{\mathbb{Z}}, q, \mathcal{F}^\bullet)$ on a non-singular complex algebraic variety $\mathcal{M}$, and we fix a basepoint $z_0 \in \mathcal{M}$. The fiber of the Hodge bundle $\mathcal{H}_{\mathbb{C}}$ is denoted by H_0, and the pairing q evaluated on H_0 is denoted by q_0. We regard H_0 as a reference vector space. The reference vector space H_0 comes with a reference Weil operator C_0, and the presence of a polarized Hodge structure on (H_0, q_0) says that the bilinear form

$$\langle u, v\rangle_0 = q_0(u, C_0 v) \tag{5.4}$$

is positive definite on $(H_0)_{\mathbb{R}}$. The period domain $\mathcal{D}$ was obtained as a quotient of the group

$$G = G_{\mathbb{R}} = \big\{ g \in \mathrm{GL}((H_0)_{\mathbb{R}}) \,|\, q_0(gu, gv) = q_0(u, v) \ \text{ for all } \ u, v \in (H_0)_{\mathbb{R}} \big\}$$

by the stabilizer H of a the reference Hodge structure on H_0. In the proof of this theorem, the Weil operator C_z is the central object, as self-duality is defined in terms of C_z. Instead of $\mathcal{D} = G/H$, we therefore consider the symmetric space G/K, where K is the subgroup

$$K = \big\{ g \in G \,|\, gC_0 = C_0 g \big\}. \tag{5.5}$$

Note that elements in K preserve the positive definite inner product $\langle -,-\rangle_0$, so it follows that K is compact. In [15] it is in fact shown that K the maximal compact subgroup containing H. The reason for considering the space G/K is as follows.

Lemma 5.1.3. *There is a one-to-one correspondence*

$$\begin{aligned} G/K &\xleftrightarrow{\;1:1\;} \{\text{Weil operators on } H_0\} \\ gK &\longmapsto \quad gC_0g^{-1} \end{aligned}$$

Here a Weil operator is any operator C on H_0 for which the pairing $q(-,C-)$ is positive definite on $(H_0)_{\mathbb{R}}$. We show that this correspondence is well-defined, and refer to [4] for the complete proof. Let gK be a coset in G/K and let $u,v\in(H_0)_{\mathbb{R}}$. Then gC_0g^{-1} is a Weil operator, since the pairing

$$q_0(u,gC_0g^{-1}v)=q_0g^{-1}u,C_0g^{-1}v=\left\langle g^{-1}u,g^{-1}v\right\rangle_0$$

is positive definite. If $\tilde{g}\in K$ is another representative of the same coset, then $\tilde{g}=gk$ for some $k\in K$, and

$$\tilde{g}C_0\tilde{g}^{-1}=gkC_0k^{-1}g^{-1}=gC_0g^{-1},$$

so both representatives define the same Weil operator.

Instead of letting the arithmetic subgroup $\Gamma\subseteq G_{\mathbb{Z}}=\mathrm{Aut}((H_0)_{\mathbb{Z}},q_0)$ be defined as an arbitrary arithmetic group containing the monodromy group associated to the fundamental group of $\mathcal{M}$, we take $\Gamma=G_{\mathbb{Z}}$. This choice will be crucial later. In the present setting, we have a 'quasi'-period map

$$\Phi:\mathcal{M}\to\Gamma\backslash G/K. \tag{5.6}$$

This map may be regarded as the composition of the usual period map and a morphism of arithmetic quotients. Using the identification of the lemma above, it associates to a point $z\in\mathcal{M}$ the Weil operator C_z on the fiber $(\mathcal{H}_{\mathbb{C}})_z$, which by parallel transport is interpreted as a Weil operator on the reference vector space H_0. By Theorem 4.2.1, the map Φ is definable in the o-minimal structure $\mathbb{R}_{\mathrm{an,exp}}$.

The idea behind the proof of the definability of the self-dual locus $\mathscr{S}_{\mathcal{H}}(t_0)$ is to move the problem to the arithmetic quotient $\Gamma\backslash G/K$. Here the self-dual locus has an explicit description, and using a result from lattice theory we can prove that this locus is definable. By pulling back along the map Φ, the definability of $\mathscr{S}_{\mathcal{H}}(t_0)$ follows. Let us now discuss the details.

Consider the universal cover $p:\widetilde{\mathcal{M}}\to\mathcal{M}$ of the base space $\mathcal{M}$. The period map then lifts to a map $\widetilde{\Phi}$ for which the diagram

$$\begin{tikzcd} \widetilde{\mathcal{M}} \arrow[r, "\widetilde{\Phi}"] \arrow[d, "p"] & G/K \arrow[d, "\pi"] \\ \mathcal{M} \arrow[r, "\Phi"] & \Gamma\backslash G/K \end{tikzcd} \tag{5.7}$$

commutes. The first step of the proof is to move the situation to $\Gamma\backslash G/K$, and this requires us to extend the definability of the period map Φ to a certain map of vector bundles.

To this end, we start with a trivial vector bundle $G/K \times H_0$, and then take a quotient by Γ to obtain a vector bundle

$$\Gamma\backslash(G/K \times H_0) \to \Gamma\backslash G/K \tag{5.8}$$

on the arithmetic quotient. Here Γ acts on $G/K \times H_0$ as

$$\gamma \cdot (gK, v) = (\gamma gK, \gamma v). \tag{5.9}$$

This bundle is 'universal', in the sense that a general Hodge bundle may be obtained by taking the pullback along a period map. In order to make this precise, we first consider the pullback bundle $p^*\mathcal{H}_\mathbb{C}$ of the Hodge bundle to the universal cover $\widetilde{\mathcal{M}}$. On the universal cover there is no monodromy, and recalling that the Hodge bundle is flat, we see that $p^*\mathcal{H}_\mathbb{C}$ is trivialized by flat sections. In other words, we have $p^*\mathcal{H}_\mathbb{C} \cong \widetilde{\mathcal{M}} \times H_0$. This allows us to look at the trivial morphism of vector bundles

$$\widetilde{\Phi} \times \mathrm{id} : \widetilde{M} \times H_0 \to G/K \times H_0. \tag{5.10}$$

Since the monodromy group is contained in Γ, this morphism descends to a well-defined map $\Phi_\mathcal{H}$ which makes the following diagram commute:

$$\begin{array}{ccc} \mathcal{H}_\mathbb{C} & \xrightarrow{\Phi_\mathcal{H}} & \Gamma\backslash(G/K \times H_0) \\ \downarrow & & \downarrow \\ \mathcal{M} & \xrightarrow{\quad\Phi\quad} & \Gamma\backslash G/K \end{array}$$

To be explicit, the map $\Phi_\mathcal{H}$ sends a point $(z, v) \in (\mathcal{H}_\mathbb{C})_z$ to the class $\Gamma(gK, v_0)$, where gK is characterized by $\Phi(z) = \Gamma gK$ and v_0 is obtained by parallel transporting v to the reference fiber H_0, which is well-defined up to the action of Γ. We will show that this map is definable, extending the definability of the period map Φ.

Lemma 5.1.4. *The complex vector bundles $\mathcal{H}_\mathbb{C} \to \mathcal{M}$ and $\Gamma\backslash(G/K \times H_0) \to \Gamma\backslash G/K$ are $\mathbb{R}_{\mathrm{alg}}$-definable.*

Proof. We follow the proof of [4]. The Hodge bundle $\mathcal{H}_\mathbb{C} \to \mathcal{M}$ can be given a definable structure as follows. The base space $\mathcal{M}$ is algebraic and hence definable, so it has a definable atlas $\{(U_i, \phi_i)\}$. By invoking the triangulation theorem, we may assume that each of the finitely many U_i is simply connected. On each U_i, we can then consider a frame of sections $\sigma_1, \ldots, \sigma_r$ which are flat with respect to the Gauss-Manin connection ∇. As explained in Section 3.6, this induces holomorphic trivializations $\psi_i : \mathcal{H}_\mathbb{C}|_{U_i} \to U_i \times \mathbb{C}^r$. Since the Hodge bundle comes from a local system $\mathcal{H}_\mathbb{Z}$, the transition functions of these trivializations are constant. An $\mathbb{R}_{\mathrm{alg}}$-definable atlas for $\mathcal{H}_\mathbb{C}$ is then given by $\{(\mathcal{H}_\mathbb{C}|_{U_i}, (\phi_i \times \mathrm{id}) \circ \psi_i)\}$.

The definable structure on $\Gamma\backslash(G/K \times H_0)$ is obtained as follows. Let $F \subseteq G/K$ be a fundamental set for the action of Γ; in the previous chapter we constructed this fundamental set as a finite union of Siegel sets $F = \bigcup_{i=1}^{l} \mathfrak{S}_i$. This naturally provides us with a fundamental set for the action of Γ on the trivial (and hence definable) vector bundle $G/K \times H_0$, namely $F \times H_0$. By Proposition 3.4.12 it follows that $\Gamma\backslash(G/K \times H_0)$ has an $\mathbb{R}_{\mathrm{alg}}$-definable structure. □

The definable structure on the Hodge bundle constructed here is the unique structure for which every flat section of $\mathcal{H}_\mathbb{C}$ is definable. It can be shown that this definable structure is equivalent to the one obtained from the algebraic structure coming from the underlying local system $\mathcal{H}_\mathbb{Z}$ [78].

Lemma 5.1.5. *The morphism of complex vector bundles $\Phi_{\mathcal{H}} : \mathcal{H}_{\mathbb{C}} \to \Gamma\backslash(G/K \times H_0)$ is $\mathbb{R}_{\text{an,exp}}$-definable.*

Proof. As in the proof of the definability of the period map, we may check this locally and assume that $\mathcal{M}$ is of the form $(\Delta^*)^m$. Recall from the previous chapter that then the period map and the lifted period map fit into a commutative diagram

$$\begin{tikzcd} \mathbb{H}^m_{C,\eta} \arrow[r, "\widetilde{\Phi}"] \arrow[d, "p"] & \bigcup_{i=1}^{l} \mathfrak{S}_i \arrow[d, "\pi"] \\ (\Delta^*)^m \arrow[r, "\Phi"] & \Gamma\backslash G/K, \end{tikzcd} \tag{5.11}$$

where the $\mathfrak{S}_i \subseteq G/K$ are Siegel sets. We claim that the vector bundle isomorphism

$$p^*\mathcal{H}_{\mathbb{C}} \to \mathbb{H}^m_{C,\eta} \times H_0,$$

whose global version was constructed above, is definable. Indeed, this follows from the lemma above and the observation that this global trivialization is constructed by means of a frame of flat sections. Since the lifted period map $\widetilde{\Phi}$ is definable, it follows that the composition

$$\begin{tikzcd} p^*\mathcal{H}_{\mathbb{C}} \arrow[r, "\cong"] \arrow[rr, bend right] & \mathbb{H}^m_{C,\eta} \times H_0 \arrow[r, "\widetilde{\Phi}\times\mathrm{id}"] & \bigcup_{i=1}^{l} \mathfrak{S}_i \times H_0, \end{tikzcd}$$

is definable in $\mathbb{R}_{\text{an,exp}}$. Taking the quotient by $\pi_1((\Delta^*)^m)$ on the left and by Γ on the right, which is definable since we are restricted to appropriate fundamental sets, we conclude that the map $\Phi_{\mathcal{H}}$ is $\mathbb{R}_{\text{an,exp}}$-definable. $\square$

With this result, we can begin to study the self-dual locus in the vector bundle $\Gamma\backslash(G/K \times H_0)$, so that eventually this locus can be pulled back along $\Phi_{\mathcal{H}}$ to conclude that $\mathscr{S}_{\mathcal{H}}(t_0)$ is definable. The first step is to prove a version of the theorem for a single Γ-orbit [4].

Proposition 5.1.6. *Let $a \in (H_0)_{\mathbb{Z}}$ be a non-zero vector with $C_0 a = a$. Then the set*

$$\mathscr{S}_a = \left\{\Gamma(gK, v) \in \Gamma\backslash(G/K \times H_0) \,\middle|\, v \in \Gamma a \text{ and } (gC_0g^{-1})v = v\right\} \tag{5.12}$$

is definable in $\mathbb{R}_{\text{alg}}$.

Proof. The idea behind the proof is to construct this set as the image of an embedding of an appropriately defined arithmetic quotient. Let

$$G_a = \{g \in G \,|\, ga = a\} \tag{5.13}$$

be the stabilizer of a, and set $K_a = K \cap G_a$, $\Gamma_a = \Gamma \cap G_a$. Then K_a is a maximal subgroup of G_a, and Γ_a is an arithmetic subgroup of G_a. By linear-algebraic arguments, one can show that if C is any Weil operator for which a is self-dual, i.e. $Ca = a$, then C is related to the reference Weil operator C_0 by

$$C = gC_0g^{-1}$$

for some element $g \in G_a$. We refer to Lemma 29.1 of [4] for the details of this argument. It then follows that the image of the embedding $G_a/K_a \hookrightarrow G/K$ consists of those gK for which the Weil operator gC_0g^{-1} satisfies $(gC_0g^{-1})a = a$. By bringing the arithmetic subgroups Γ and Γ_a into the picture, we obtain a morphism of arithmetic quotients

$$\iota : \Gamma_a \backslash G_a/K_a \to \Gamma\backslash G/K, \tag{5.14}$$

which is $\mathbb{R}_{\mathrm{alg}}$-definable by Theorem 4.1.13. By arguing with Siegel sets, it can be shown that the lifted map

$$\begin{aligned} \tilde{\iota} : \Gamma_a \backslash G_a/K_a &\to \Gamma\backslash (G/K \times H_0), \\ \Gamma_a g K_a &\mapsto \Gamma(gK, a) \end{aligned} \tag{5.15}$$

is also $\mathbb{R}_{\mathrm{alg}}$-definable; for details we refer to [4]. We now have an embedding of an arithmetic quotient into the universal vector bundle that we are interested in, and we claim that the self-dual locus $\mathscr{S}_a$ is precisely the image of $\tilde{\iota}$. To see this, suppose that a general element $\Gamma(gK, v)$ is given by $\Gamma(\tilde{g}K, a)$ for some $\tilde{g} \in G_a$, i.e. it lies in the image of $\tilde{\iota}$. Then the representatives of these elements are related by a group element $\gamma \in \Gamma$,

$$v = \gamma a \quad \text{and} \quad g = \gamma \tilde{g}.$$

It follows that v lies in the orbit Γa, and we have

$$(gC_0g^{-1})v = \gamma\tilde{g}C_0\tilde{g}^{-1}\gamma^{-1}v = \gamma\tilde{g}C_0\tilde{g}^{-1}a = \gamma a = v,$$

so that indeed $\Gamma(gK, v) \in \mathscr{S}_a$. Conversely, if $\Gamma(gK, a) \in \mathscr{S}_a$, then $gC_0g^{-1}a = a$ and as in the argument above there exists a $\tilde{g} \in G_a$ with $gC_0g^{-1} = \tilde{g}C_0\tilde{g}^{-1}$. This implies that $g^{-1}\tilde{g}$ commutes with C_0, which means that is is an element of K. In other words, $gK = \tilde{g}K$, so that $\Gamma(gK, a)$ lies in the image of $\tilde{\iota}$. Since the image of $\tilde{\iota}$ is definable, we conclude that set $\mathscr{S}_a$ is definable. □

This shows that the self-dual locus is definable when restricted to a *single* orbit for the action of Γ. In general, the infinite discrete nature of the group Γ suggests that there are infinitely many such orbits, and there appears little hope to prove that the self-dual locus is definable. However, there is one crucial ingredient that we have not yet used: the tadpole condition $q_0(v, v) = t_0$. Miraculously, with the assumption of the tadpole condition it turns out that there exists a classic result in lattice theory that solves the problem at hand. This result says that for $t_0 \geq 1$, Γ acts on the set

$$\{v \in (H_0)_{\mathbb{Z}} \,|\, q_0(v, v) = t_0\}$$

with *finitely many orbits*, and is proven in the book on quadratic forms by Kneser [79]. Here it is essential that $\Gamma = G_{\mathbb{Z}}$, which is the automorphism group of the lattice $((H_0)_{\mathbb{Z}}, q_0)$. If Γ was taken to be a smaller arithmetic group, this finiteness theorem would in general not be true.

Applied to the present setting, we find that there exist finitely many self-dual elements $a_1, \ldots, a_k \in (H_0)_{\mathbb{Z}}$ with $q_0(a_i, a_i) = t_0$ such that the set

$$\mathscr{S}(t_0) = \big\{\Gamma(gK, v) \in \Gamma\backslash(G/K \times H_0) \,\big|\, v \text{ is integral and } (gC_0g^{-1})v = v\big\} \tag{5.16}$$

is given by the finite union $\mathscr{S}(t_0) = \bigcup_{i=1}^{k} \mathscr{S}_{a_i}$. We thus conclude:

Proposition 5.1.7. *The set* $\mathscr{S}(t_0) \subseteq \Gamma\backslash(G/K \times H_0)$ *is definable in* $\mathbb{R}_{\text{alg}}$.

By using the map $\Phi_{\mathcal{H}} : \mathcal{H}_{\mathbb{C}} \to \Gamma\backslash(G/K \times H_0)$ to return to the Hodge bundle over $\mathcal{M}$, we arrive at the final step in the proof.

Proof of Theorem 5.1.1. We claim that the self-dual locus that we are interested in is given by the preimage

$$\mathscr{S}_{\mathcal{H}}(t_0) = \Phi_{\mathcal{H}}^{-1}\big(\mathscr{S}(t_0)\big). \tag{5.17}$$

To see this, note that a point $(z, v) \in \mathscr{S}_{\mathcal{H}}(t_0)$ is such that v is integral and satisfies $q_z(v, v) = t_0$, $C_z v = v$. Let $v_0 \in H_0$ denote a parallel transport of v_0 to the reference fiber H_0. The preceding conditions then translate to $gC_0g^{-1}v_0 = v_0$ and $q_0(v_0, v_0) = t_0$, where g is determined by $\Phi(z) = \Gamma gK$. This is precisely the condition that $\Phi_{\mathcal{H}}$ maps (z, v) to a point in $\mathscr{S}(t_0)$, which proves the claim. Since the map $\Phi_{\mathcal{H}}$ is $\mathbb{R}_{\text{an,exp}}$-definable, we conclude that the self-dual locus $\mathscr{S}_{\mathcal{H}}(t_0) \subseteq \mathcal{H}_{\mathbb{H}}$ is definable in $\mathbb{R}_{\text{an,exp}}$. □

This completes the proof of the definability of the self-dual locus, and therefore of the finiteness of flux vacua. It is beautiful and quite surprising to see that certain physical conditions, most notably the tadpole bound and self-duality, conspire to produce this mathematical result. The proof brings together many of the ideas and techniques that have been discussed so far in this thesis: the analysis of flux vacua for the physical conditions, Hodge theory to formulate the setting mathematically, and the tame geometry of arithmetic quotients and the period map for the proof. We now continue to discuss how tameness manifests itself in effective field theories arising from string theory.

5.2 Definability in Type IIB/F-Theory Compactifications

In Chapter 2 we have discussed that, after compactifying and taking a low-energy limit, string vacua give rise to four-dimensional effective field theories whose structure is determined by the geometry of the internal space. We now make this more precise, focusing on a Calabi-Yau compactification of Type IIB or F-Theory. Various types of fields emerge in the effective theory, and in what follows we look at the scalars, abelian 1-form gauge fields, and the spacetime metric, thus leaving the fermions and p-form gauge fields for $p > 1$ out of the discussion. For this sector, a generic action takes the form

$$S_{\text{EFT}} = \int_{M_4} \frac{1}{2}\mathcal{R} * 1 - \frac{1}{2}\mathcal{G}_{ij} D\phi^i \wedge *D\phi^j - \frac{1}{2} f_{ab} F^a \wedge *F^b + \frac{1}{2}\tilde{f}_{ab} F^a \wedge F^b - V * 1. \tag{5.18}$$

In this action, the following fields appear:

(i) the spacetime metric $g_{\mu\nu}$, which determines the Hodge star $*$ and whose Ricci scalar $\mathcal{R}$ appears in the Einstein-Hilbert term;

(ii) the scalar fields ϕ^i which take values in a field space $\mathcal{M}_{\text{s}}$, coupled kinetically by a field space metric $\mathcal{G}_{ij}$ on $\mathcal{M}_{\text{s}}$ and by a scalar potential $V(\phi)$;

(iii) the abelian 1-form gauge fields A^a_μ whose field strength $F^a = dA^a_\mu$ are coupled kinetically by the gauge coupling functions f_{ab} and $\tilde{f}_{ab}$ which depend on the scalars, and further coupled to the scalars through the gauge covariant derivative D. The gauge coupling functions f_{ab} and $\tilde{f}_{ab}$ form symmetric matrices.

In this action we are ignoring the presence of constant prefactors, such as numerical factors and powers of the Planck mass. Writing this action down in coordinates using spacetime indices, it becomes

$$S_{\mathrm{EFT}} = \int_{M_4} d^4x\sqrt{-g}\Big(\frac{1}{2}\mathcal{R} - \frac{1}{2}\mathcal{G}_{ij}D_\mu\phi^i D^\mu\phi^j - \frac{1}{2}f_{ab}F^a_{\mu\nu}F^{b,\mu\nu} + \frac{1}{4}\tilde{f}_{ab}\epsilon^{\mu\nu\rho\sigma}F^a_{\mu\nu}F^b_{\rho\sigma} - V(\phi)\Big). \tag{5.19}$$

In the following sections we will see that this effective action, when arising as an effective description of Type IIB or F-theory, has a certain tame structure. More precisely, we will see that the various coupling functions that appear in this action are definable in an o-minimal structure. This is remarkable, because as an effective field theory nothing appears to demand a priori that these functions should be definable, as one could for instance imagine that the scalar potential takes the form

$$V(\phi) \sim \sin(\phi), \tag{5.20}$$

with ϕ taking values on the real line $\mathbb{R}$. To see how tameness arises in the action S_{EFT}, we will now discuss the form that these coupling functions take for compactifications of Type IIB and F-theory, starting with the scalar sector.

5.2.1 Scalars and Field Space Geometry

For two of the terms in the action S_{EFT}, we have already seen how they emerge from string theory. The complex structure moduli z^i, $\bar{z}^{\bar{\imath}}$ take values in the complex structure moduli space $\mathcal{M}$, whose geometry is given by the Weil-Petersson metric $\mathcal{K}_{i\bar{\jmath}}$. Recall that the component functions $\mathcal{K}_{i\bar{\jmath}}$ are given by $\mathcal{K}_{i\bar{\jmath}} = \partial_i\bar{\partial}_{\bar{\jmath}}\mathcal{K}$, where the Kähler potential $\mathcal{K}$ is given by

$$\mathcal{K}(z,\bar{z}) = -\log i^n \int_X \Omega(z)\wedge\overline{\Omega(z)}. \tag{5.21}$$

Here $\Omega(z)$ defines a non-vanishing holomorphic section of the line bundle $\mathcal{F}^n = \mathcal{H}^{n,0}$ over $\mathcal{M}$, and n is the complex dimension of the Calabi-Yau X that we are compactifying on.

This leads us to the first manifestation of tameness in an effective field theory.

Proposition 5.2.1. *The Weil-Petersson metric $\mathcal{K}_{i\bar{\jmath}}$ on the complex structure moduli space $\mathcal{M}$ of a Calabi-Yau threefold or fourfold X is definable in $\mathbb{R}_{\mathrm{an,exp}}$.*

The proof will reflect the philosophy of the proof of the definability of the self-dual locus, where we use the period map to transport the problem to a setting where definability is more clear.

Proof. As explained in Section 3.6, we prove this locally by considering a definable coordinate system on a definable open set $U \subseteq \mathcal{M}$ and checking the definability of the component functions $\mathcal{K}_{i\bar{\jmath}}$. By definable triangulation we may assume that U is simply connected, and this will be useful since we may then disregard the presence of monodromy. Since derivatives preserve definability, the problem reduces to proving that the Kähler potential $\mathcal{K}$ is definable on U. To do so, we will rewrite $\mathcal{K}$ to a form that makes its dependence on the period map manifest.

Since U is simply connected, the period map Φ lifts to an $\mathbb{R}_{\mathrm{an,exp}}$-definable map $\widetilde{\Phi} : U \to G/K$. Fix a reference point $z_0 \in U$ and let H_0 be the reference fiber of the Hodge bundle. Inside the subspace $H_0^{n,0}$ we choose a reference vector Ω_0, which corresponds to the cohomology class of a holomorphic $(n,0)$-form on X_0. The Kähler potential can now be obtained as the composition

$$\begin{tikzcd} U \arrow[r, "\widetilde{\Phi}"] \arrow[rr, bend right, "\mathcal{K}"'] & G/H \arrow[r, "\widetilde{\mathcal{K}}"] & \mathbb{R}, \end{tikzcd} \tag{5.22}$$

where $\widetilde{\mathcal{K}}$ is the map

$$gH \mapsto -\log\left(i^n\, q_0(g\,\Omega_0, \overline{g\,\Omega_0})\right). \tag{5.23}$$

This map is well-defined, since the group H the leaves the Hodge decomposition, and hence the subspace $H_0^{n,0} = \mathbb{C}\Omega_0$, invariant. Since H is compact, it therefore acts on Ω_0 by a $\mathrm{U}(1)$ phase factor, which gets cancelled in the map $\widetilde{\mathcal{K}}$. To see that this composition indeed yields the Kähler potential, recall that the lifted period map sends a point z to the double coset $g(z)H$, where $g(z)$ satisfies $\Omega(z) = g(z)\,\Omega_0$. This equality holds up to parallel transport to the reference fiber H_0, which is unique since there is no monodromy on the simply connected set U.

We claim that the map $\widetilde{\mathcal{K}}$ is $\mathbb{R}_{\mathrm{an,exp}}$-definable. First, note that the logarithm is $\mathbb{R}_{\mathrm{exp}}$-definable on $\mathbb{R}_{>0}$ since it is the inverse of the real exponential. Secondly, the group G acts algebraically on H_0, so that the action of the group element g on Ω_0 is definable. More formally, we have to show that the action of G/H on H_0 is definable, but this follows from the definability of the projection map $G \to G/H$ as shown in Example 3.4.9. Finally, the pairing q_0 is bilinear and hence definable. Assembling the pieces, we find that $\widetilde{\mathcal{K}}$ is definable in $\mathbb{R}_{\mathrm{an,exp}}$, and by the definability of the lifted period map we conclude that the Kähler $\mathcal{K}$ potential is $\mathbb{R}_{\mathrm{an,exp}}$-definable. □

We thus find that the complex structure moduli z^i, $\bar{z}^{\bar{\jmath}}$ in the effective field theory arising from a Type IIB or F-theory compactification are coupled by a definable function, owing to the tameness of the period map. The dynamics of these scalars is further determined by the scalar potential V discussed in Section 2.4. In F-theory, we saw that this potential takes the form

$$V(z, G_4) = \int_X G_4 \wedge *G_4 - \int_X G_4 \wedge G_4. \tag{5.24}$$

This potential depends on the scalar fields z^i through the Hodge star $*$, but also on the flux G_4 which may be regarded as a parameter of the compactification. In the language of Hodge theory, we viewed this potential as the following map, defined on the Hodge bundle $\mathcal{H}_{\mathbb{C}}$:

$$(z, v) \mapsto q_z(v, C_z v) - q_z(v, v).$$

With this definition of the scalar potential, we can prove the following.

Proposition 5.2.2. *The scalar potential $V : \mathcal{H}_{\mathbb{C}} \to \mathbb{R}$ is definable in* $\mathbb{R}_{\text{an,exp}}$.

Proof. As in the proof of the previous proposition, we move the problem to a simpler setting by using the period map. More precisely, we use the map $\Phi_{\mathcal{H}} : \mathcal{H}_{\mathbb{C}} \to \Gamma\backslash(G/H \times H_0)$ from the proof of the definability of the self-dual locus. In the notation of that proof, we define the map

$$\begin{aligned} \widetilde{V} : \Gamma\backslash(G/H \times H_0) &\to \mathbb{R} \\ \Gamma(gH, v) &\mapsto q_0(v, gC_0g^{-1}v) - q_0(v, v) \end{aligned}$$

Analogous to the proof of the definability of the Weil-Petersson metric, we recover the scalar potential V as the composition

$$\begin{tikzcd} \mathcal{H}_{\mathbb{C}} \arrow[r, "\Phi_{\mathcal{H}}"] \arrow[rr, bend right, "V"'] & \Gamma\backslash(G/H \times H_0) \arrow[r, "\widetilde{V}"] & \mathbb{R}. \end{tikzcd} \tag{5.25}$$

To see that this composition indeed yields V, recall that the map $\Phi_{\mathcal{H}}$ sends a point (z, v) in the Hodge bundle to $\Gamma(gH, v_0)$, where $\Phi(z) = \Gamma gH$ and v_0 is a parallel transport of v to the reference fiber. The map $\widetilde{V}$ then evaluates to the scalar potential, where the dependence of the Weil operator C_z is encoded in the Weil operator gC_0g^{-1} on the reference fiber, with g determined by the period map.

We thus have to argue that the map $\widetilde{V}$ is definable. On the space $G/H \times H_0$, the definability of the map

$$(gK, v) \mapsto q_0(v, gC_0g^{-1}v) - q_0(v, v)$$

follows from the algebraicity of the group action and the pairing q_0, as in the previous proof. By using a fundamental set $F \subseteq G/H$ consisting of a finite union of Siegel sets, we can definably descend to the quotient $\Gamma\backslash(G/H \times H_0)$, and the resulting map is precisely $\widetilde{V}$. By composing with the $\mathbb{R}_{\text{an,exp}}$-definable map $\Phi_{\mathcal{H}}$, we conclude that the scalar potential V is definable in $\mathbb{R}_{\text{an,exp}}$. □

Note that this result required us to let the flux v take values in the Hodge bundle $\mathcal{H}_{\mathbb{C}}$, while in Section 2.4 we have seen that quantization of fluxes requires v to take values in the lattice subbundle $\mathcal{H}_{\mathbb{Z}}$. The fibers of this bundle are infinite discrete and therefore never definable in any o-minimal structure. Nonetheless, if we restrict to fluxes satisfying the tadpole condition, the zero set of the potential $V : \mathcal{H}_{\mathbb{Z}} \to \mathbb{R}$ *is* definable, since it is precisely the locus of flux vacua.

5.2.2 Abelian Gauge Fields and Duality Structure

Next, we consider the gauge field sector of the action S_{EFT}. Before focusing on how this sector is realized in string theory compactifications, it is necessary to briefly review some generalities on abelian gauge fields in field theories, following [25]. This discussion will also be relevant for Chapter 6, where we revisit gauge fields in a different context. We are interested in Lagrangian containing the gauge kinetic terms,

$$\mathscr{L} = -\frac{1}{2} f_{ab} F^a \wedge *F^b + \frac{1}{2} \tilde{f}_{ab} F^a \wedge F^b, \tag{5.26}$$

where $F^a = dA^a$ are the 2-form field strengths, and $a = 1, \ldots, n$. The gauge couplings f_{ab} and $\tilde{f}_{ab}$ are functions on the scalar field space $\mathcal{M}$. In the following we only consider the case that the A^a are 1-forms in four dimensions, but the discussion holds more generally for $(m-1)$-forms in a $2m$-dimensional spacetime. Viewing the Lagrangian as a function of the fields, the classical equation of motion is given by

$$d\left(*\frac{\partial\mathscr{L}}{\partial F^a}\right) = 0. \tag{5.27}$$

Introducting the dual field strength

$$G_a = *\frac{\partial\mathscr{L}}{\partial F^a}, \tag{5.28}$$

the equation of motion is simply the statement that G_a is closed. Since F^a is a field strength defined by dA^a, it is also closed, and F^a is said to satisfy the Bianchi identity. In summary, we have

$$dF^a = 0, \quad dG_a = 0. \tag{5.29}$$

The symmetry between these two equations reveals a duality symmetry of the theory. Collecting the $2n$ 2-forms F^a and G_a into a single vector, we can perform linear transformations on this vector, thereby rotating the equations of motion and Bianchi identities into each other. Through a careful analysis, it can be shown that the group of linear transformations that preserves the structure of the Lagrangian is the symplectic group $\mathrm{Sp}(2n,\mathbb{R})$ [25]. In this setting, this group is thought of as a group of dualities, encoding a symmetry of the theory at the level of the equations of motion[2].

For this reason, the gauge coupling functions f_{ab} and $\tilde{f}_{ab}$ are actually a frame-dependent quantity, requiring us to first disentangle the field strengths from their duals. In other words, given a collection of $2n$ closed 2-forms, we have to select which ones are to be interpreted as the field strengths appearing in the Lagrangian. Such a selection amounts to a choice of *duality frame*. An obstacle in formulating the gauge coupling functions is that duality frames experience monodromy. In other words, if the scalars traverse a non-trivial loop in $\mathcal{M}$, the symplectic frame is rotated by a $\mathrm{Sp}(2n,\mathbb{R})$-transformation.

We now turn to effective field theories arising from string compactifications. More precisely, we focus on a $\mathcal{N}=2$ supergravity theory arising from a Calabi-Yau compactification of Type IIB. In this setting, there are $n = h^{2,1}+1$ abelian gauge fields, one of which belongs to the gravity multiplet, and the remaining of which appear in the $h^{2,1}$ vector multiplets. Because of the monodromy experienced by the duality frame, we will only look at the gauge coupling functions locally, on a finite simply connected open cover of $\mathcal{M}$. Hence, let U be an open definable subset in this cover. A choice of duality frame can be conveniently determined in terms of a symplectic homology basis for $H_3(X_z;\mathbb{Z})$ for $z\in U$. This is a basis $(\alpha_1,\ldots,\alpha_{h^{2,1}+1},\beta^1,\ldots,\beta^{h^{2,1}+1})$ whose intersection pairing is given by

$$\alpha_a\cdot\beta^b = \delta_a^b, \quad \alpha_a\cdot\alpha_b = \beta^a\cdot\beta^b = 0. \tag{5.30}$$

Such a basis can be consistently defined over U, since the simply connectedness ensures that there is no monodromy. The distinction between the field strengths F^a and G_b is then determined by this choice of homology basis through the compactification procedure[3] [46]. The gauge coupling function f_{ab} may then be expressed as [1]

$$f_{ab} = (\Pi_a, \overline{D_i\Pi_a})(\Pi^b, \overline{D_i\Pi^b})^{-1}, \tag{5.31}$$

[2]This duality is often called electric-magnetic duality, in analogy with Maxwell theory.

[3]More precisely, the 5-form field strength $\tilde{F}_5$ of the ten-dimensional theory decomposes as $\tilde{F}_5 = F^a\wedge\beta_a^* + G_b\wedge\alpha^{b*}$.

where Π_a and Π^b are the periods of Ω with respect to the chosen symplectic homology basis:

$$\Pi^a = \int_{\alpha_a} \Omega, \quad \Pi_b = \int_{\beta^b} \Omega. \tag{5.32}$$

Here $D_i = \partial_i + \partial_i \mathcal{K}$ again denotes the Kähler covariant derivative with respect to the complex structure moduli z^i. The dependence of f_{ab} on these moduli is determined by the dependence of the $\Omega(z)$. Completely analogous arguments to the ones given in the previous two propositions, invoking the definability of the period map, show that:

Proposition 5.2.3. *The gauge coupling function f_{ab} is $\mathbb{R}_{\text{an,exp}}$-definable on U.*

The gauge coupling $\tilde{f}_{ab}$ appearing in the action S_{EFT} has a similar construction in terms of the periods, and its (local) definability follows similarly. Globally, one has to take care of monodromy to consistently define the gauge coupling functions over the whole field space $\mathcal{M}$. Though we will not do so here, we will do this in the next chapter for gauge coupling functions in a different setting.

5.3 The Tameness Conjecture

In the previous section we have seen that in effective field theories arising from Type IIB or F-theory, there are several manifestations of tameness. The scalar field space, at least for the complex structure moduli, is an algebraic variety and hence definable, and moreover the physically relevant Weil-Petersson metric is definable. Secondly, the coupling functions for the U(1)-gauge fields was also seen to be definable in a suitable way. In addition, these effective field theories were determined by a choice of flux on the internal space. The parameter space for these fluxes is the locus of flux vacua, which was shown to be definable.

This shows that in a wide class of string vacua, tameness appears to be a general phenomenon. This is precisely the sort of observation that motivates swampland conjectures, and it leads us to the proposal of a Tameness Conjecture [1].

Conjecture 5.3.1 (Tameness Conjecture)**.** *Consider an effective field theory which is coupled to gravity and admits a UV-completion to quantum gravity. Then the scalar field space and the coupling functions are definable in an o-minimal structure. In addition, for all such effective field theories below a fixed energy cut-off scale, the parameter space is definable.*

In what follows, we will say that an effective field theory is $\mathcal{S}$-definable if it satisfies the above conditions for the structure $\mathcal{S}$. In all the cases arising from string theory that we discussed, the definability held in the o-minimal structure $\mathbb{R}_{\text{an,exp}}$. Therefore, a strong version of the Tameness Conjecture proposes that the right o-minimal structure for effective field theories is $\mathbb{R}_{\text{an,exp}}$ [1]. Since the restricted analytic functions and real exponential are ubiquitous in physics, the o-minimal structure in this conjecture can certainly not be smaller[4] than $\mathbb{R}_{\text{an,exp}}$.

[4]Though the structure cannot be smaller, the coupling functions may in fact be more tame then merely definable in $\mathbb{R}_{\text{an,exp}}$; in [80] it was proposed that in a large class of effective theories in the landscape, the couplings are *polynomially tamed* in addition to being definable.

Ever since the birth of the swampland programme, finiteness has been one of the essential principles in distinguishing the landscape from the swampland [7]. This finiteness comes in various forms, such as finiteness of field space volumes, finiteness of field content and finiteness of vacua. More recently, an additional notion of finiteness was proposed in the context of quantum gravity amplitudes [77]. The beauty, and perhaps also the power, of the Tameness Conjecture comes from the fact that finiteness is implemented in a structural mathematical way. The o-minimality condition ensures that, essentially, all countable data in an effective field theory is finite.

Therefore, the Tameness Conjecture has the possibility of being more than a single conjecture on its own, and o-minimality could serve as a unifying principle of finiteness in the swampland program. For this reason, exploring the connection between the Tameness Conjecture and other swampland conjectures is a promising avenue for further work. Recently, the first step was taken by investigating the connection between the Tameness Conjecture and the Distance Conjecture [80].

5.3.1 Definable Structures and Definable Coordinates

There is an important subtlety regarding the formulation of the Tameness Conjecture concerning the definable structure and definable coordinates that are used. According to the conjecture, the field space $\mathcal{M}$ of an effective theory in the landscape is endowed with a definable structure, which provides an atlas of definable coordinates. However, the conjecture does not specify the choice of definable structure[5]. This observation is relevant, because topological spaces may be equipped with several inequivalent definable structures, as illustrated in Example 3.4.14. We have now seen three cases in which spaces have a natural definable structure, namely

(i) algebraic varieties;

(ii) compact real-analytic manifolds;

(iii) arithmetic quotients.

For (i) and (ii) the $\mathbb{R}_{\text{alg}}$- and $\mathbb{R}_{\text{an}}$-definable structure followed naturally as the one consistent with the algebraic or analytic structure, respectively. In case (iii), the $\mathbb{R}_{\text{alg}}$-definable structure is functorial with respect to the construction of arithmetic quotients, and is independent of the choice of Siegel sets used for a fundamental set. Because of this, the definable structure on arithmetic quotients may also be regarded as natural. As a slight extension of the Tameness Conjecture, we then propose that if the field space carries a *natural* definable structure (as in (i)-(iii)), then this should be consistent with the one used in the realization of the Tameness Conjecture. By consistent we mean that, for example, if we wish to equip $\mathcal{M}$ with an $\mathbb{R}_{\text{an,exp}}$-definable structure and $\mathcal{M}$ happens to be an algebraic variety, then the natural $\mathbb{R}_{\text{alg}}$-definable structure should be compatible (in the sense that the two structures agree on what it means for a subset to be $\mathbb{R}_{\text{alg}}$-definable).

As an example, consider the two inequivalent definable structures on the punctured disk shown in Example 3.4.14. In cases that the punctured disk arises as a subset of an algebraic variety $\mathcal{M}$, the proposal says that the 'correct' definable structure is the one consistent with the algebraic structure.

[5]Note that we mean the choice of 'structure of a definable space', and not a choice of o-minimal structure $\mathcal{S}$.

A closely related subtlety is that the tameness of coupling functions, which are functions on the field space $\mathcal{M}$, is only manifest in coordinate systems which are consistent with the definable structure. Upon choosing a 'bad' coordinate system, the coupling functions may appear to be non-tame. This is no different than what happens in, for example, differential geometry: when non-smooth coordinates are chosen, the smoothness of the manifold is not seen. However, a major difference between tame geometry and differential geometry is that definable coordinates are typically much rarer than smooth coordinates, in the sense that it is often easy to think of a coordinate system that spoils tameness.

This means that, when discussing tameness in the context of effective field theories, extra care must be taken in ensuring that the chosen coordinates on the field space are definable. As an example, consider the strange-looking scalar potential on the punctured disk

$$V(r,\theta) = e^{-2\pi r}\cos\big(2\pi(\theta - \log r)\big),$$

with (r,θ) polar coordinates. In these coordinates, this potential is clearly not definably in any o-minimal structure (for instance due to the presence of infinitely many discrete zeros on the line segment $\theta = 0$). However, when applying a non-definable coordinate transformation to 'spiraling polar coordinates'

$$\begin{aligned}\tilde{r} &= r,\\ \tilde{\theta} &= \theta - \log r,\end{aligned}$$

the potential takes the form

$$V(\tilde{r},\tilde{\theta}) = e^{-2\pi\tilde{r}}\cos\Big(2\pi\tilde{\theta}\Big),$$

which is $\mathbb{R}_{\text{an,exp}}$-definable. Hence, the coordinates $(\tilde{r},\tilde{\theta})$ could be the definable coordinates making the scalar potential a definable function.

5.3.2 Towards Bottom-Up Evidence

As we have seen in our discussion of the swampland in Section 2.6, the swampland conjectures with the strongest evidence are those for which the evidence consists of both top-down and bottom-up arguments. The tameness in effective theories arising from F-theory and Type IIB compactifications provides strong top-down evidence, and this naturally motivates us to think about the Tameness Conjecture from a bottom-up perspective. At this point, convincing bottom-up evidence is not there yet. Nonetheless, in this subsection we speculate on the form that such arguments could take.

In general, bottom-up arguments for tameness in the form of o-minimal structures should address two questions:

(i) Why is the notion of a definability in a *structure* natural for effective field theories?

(ii) Why should this structure be *o-minimal* for effective field theories arising from quantum gravity?

Since the definition of a structure reflects logical operations, the first question is centered around the use of logic in physics. This is related to the question whether the universe can be fully described using mathematics. While this general theme may lead to an interesting philosophical discussion, let us focus more on the question at hand[6]. Recall that a structure on the field of real numbers is essentially a collection of subsets of the Euclidean spaces which is closed under basic logical operations. This collection can be as large as we like. We can even go as far as to consider a structure $\mathbb{R}_{\mathrm{all}}$ in which *all* sets are definable. In a tautological way, essentially any geometric object that we can associate to an effective field theory is then definable in $\mathbb{R}_{\mathrm{all}}$. This observation is not very insightful, but at least it provides a starting point for the meaning of definability in effective field theories.

To really address question (i), it is necessary to think about non-trivial structures. Suppose that we have an effective field theory which is definable in a structure which is smaller than $\mathbb{R}_{\mathrm{all}}$. The Tameness Conjecture only asserts definability of a number of specific objects associated to the effective field theory (the field space, coupling functions, and the parameter space). The specification of a definable structure on the field space $\mathcal{M}$ has strong consequences for the geometry of $\mathcal{M}$, but it also provides a distinctive criterion for subsets of $\mathcal{M}$: they can be definable or non-definable. A natural question that then emerges is whether definability of subsets of $\mathcal{M}$ has a physical meaning. In a similar spirit, it is natural to ask whether there is a physical meaning for the geometric manifestations of the logical operations defining a structure (unions, intersections, projections). Currently, such interpretations are not known.

Whereas the first question is about logic, the second one is about tameness, in the form of the o-minimal condition. An essential observation is that o-minimality is implemented at a one-dimensional level, and that the definition of a structure ensured that this finiteness condition affects all higher dimensions. Thus, assuming that an adequate answer to question (i) can be formulated, a natural setting to argue for the Tameness Conjecture from a bottom-up perspective consists of theories with a one-dimensional scalar field space. Since we want the field space to be non-compact in view of the Distance Conjecture, we are led to consider a theory with a scalar field taking values in the real line, coupled to gravity, described by an action of the form

$$S = \int_{M_D} d^D x \sqrt{-g} \left(\mathcal{R} - \frac{1}{2} \mathcal{G}(\phi)\, \partial_\mu \phi\, \partial^\mu \phi - V(\phi) \right) \tag{5.33}$$

Here the function $\mathcal{G}$ appearing in the kinetic term is interpreted as the field space metric on $\mathcal{M} = \mathbb{R}$. The presence of $\mathcal{G}$ can be removed by performing an appropriate coordinate transformation on the field space. By integrating the separable differential equation

$$\frac{d\phi}{d\phi'} = \sqrt{\mathcal{G}(\phi')} \tag{5.34}$$

and using the solution as a new variable, the kinetic term becomes canonically normalized, and the action becomes

$$S = \int_{M_D} d^D x \sqrt{-g} \left(\mathcal{R} - \frac{1}{2} \partial_\mu \phi\, \partial^\mu \phi - V(\phi) \right). \tag{5.35}$$

[6] As a side comment, we do point out that logic is closely related to the idea of *decidability*. Recently an undecidable problem in quantum field theory was highlighted in [81], and the construction of this problem involved the use of a non-definable function. It would thus be interesting to explore the connection between o-minimal structures and decidability in physics.

In this setting, the Tameness Conjecture says that the scalar potential V should be definable in an o-minimal structure in order for this effective field theory to be consistent with quantum gravity. Likely, the most promising angle from which to approach this problem is to ask

what goes wrong if V is not tame?

Addressing this question requires us to have a good understanding of definability and non-definability. The key signature of non-definability is infinite discreteness. Therefore, the first step to take is to consider a potential where infinite discreteness is manifest. An evident way for infinite discrete behavior to appear in a potential is the breaking of the monotonicity theorem (Theorem 3.3.1), which means that we should look at potentials with infinitely many isolated critical points. A natural candidate is the potential

$$V(\phi) = \sin(\phi). \tag{5.36}$$

An immediate consequence of this potential is that the effective theory acquires a $\mathbb{Z}$-global symmetry

$$\phi \mapsto \phi + 2\pi. \tag{5.37}$$

According to the No Global Symmetries Conjecture, this symmetry should be broken or gauged. Gauging it amounts to replacing the field space $\mathbb{R}$ by $\mathbb{R}/\mathbb{Z} = S^1$, and since the circle is compact, this potential is restricted analytic and hence definable $\mathbb{R}_{\text{an}}$ on S^1. This is consistent with the Tameness Conjecture.

This observation holds for any periodic potential. Therefore, in order to investigate how the effective field theory could become inconsistent from a scalar potential with infinite and discrete behavior, we have to focus on non-periodic potentials. For example, we could consider a potential of the form

$$V(\phi) = v(\phi)\sin(\phi), \tag{5.38}$$

where $v(\phi)$ is some simple non-periodic function (e.g. a polynomial or real exponential). The argument that such a 'wild' potential leads to inconsistencies would then constitute bottom-up evidence for the Tameness Conjecture.

Note that, since this is a swampland conjecture, the role of gravity played in this story is essential. Indeed, from purely field-theoretic considerations there appears a priori to be no principle that forbids a scalar potential as in equation (5.38). The inclusion of gravity into the theory should, if the Tameness Conjecture is true, somehow alter this picture. Therefore, the best setting to look for evidence is probably one where the role of gravity is paramount, namely the setting of black holes. As noted in Chapter 2, black hole solutions in effective field theories have indeed provided a rich bottom-up testing ground[7] for swampland principles, for instance for the No Global Symmetries Conjecture [82], the Weak Gravity Conjecture [83], and more recently also the Distance Conjecture [77].

[7] In many cases, black hole solutions in fact provide part of the original motivation for swampland conjectures.

Chapter 6

Tameness Conjecture and Supergravity

In testing the validity of swampland conjectures, it is often valuable to explore the conjectures in the context of effective field theories with a large amount of supersymmetry. While such theories may be unrealistic from a phenomenological perspective, the mathematically rigid structure imposed by supersymmetry allows for precise mathematical arguments in favor or against a swampland conjecture and often leads to important insights. In [1], evidence for the Tameness Conjecture from this supersymmetric perspective was suggested. In this chapter we aim to formulate the details of this evidence. More precisely, we will provide evidence in supergravity theories with more than eight supercharges, a setting which we will refer to as *higher supergravity*.

In ten spacetime dimensions, higher supergravities are encountered as the low-energy effective description of the five superstring theories. We have seen that Calabi-Yau compactifications, possibly supplemented with an orientifold projection, reduce the large amount of supersymmetry to $\mathcal{N} = 1$ in four dimensions. If we instead compactify on a simpler space such as a torus, a K3 surface, or a product thereof, we obtain lower-dimensional higher supergravity theories. For example, a compactification of Type IIB supergravity on the space $\mathrm{K3} \times \mathbb{T}^2$ yields $\mathcal{N} = 4$ supergravity in four dimensions [23].

Supergravity is a vast area of research and has many technical aspects. In view of regarding Hodge theory, string theory, and tame geometry as the three pillars of this thesis, supergravity could in fact rightfully be regarded as a fourth pillar. We have already briefly made contact with it in earlier chapters, but we did not discuss many details. In this chapter we delve into the technical structure of higher supergravities, but in view of constraints of space and time we focus on the applications that we are interested in.

We begin this chapter with a general overview of how supersymmetry constrains the geometry of the theory. In particular, we will begin with a review of the scalar field space geometry, and we will see that when more than eight supercharges are present, this geometry is almost uniquely determined. In this discussion we encounter the arithmetic quotients, which have played an important role in our study of the period map throughout this thesis. We emphasize some arithmetic aspects of higher supergravity and show how this could relate tameness to some other swampland principles, in particular the No Global Symmetries Conjecture and the Completeness Hypothesis.

6.1 Higher Supergravity - the Scalar Sector

6.1.1 Supersymmetry and Locally Symmetric Spaces

As discussed in Chapter 2, the amount of supersymmetry in a theory is determined by the number of supersymmetry generators $\mathcal{N}$, whose components count the number of supercharges. These supersymmetry generators are spinors on the spacetime manifold, and their presence severely constrains the geometric structure of the theory. For instance, the field space $\mathcal{M}$ in which the scalars of the theory take values, has a type of geometry which is determined by the number of supercharges. Though we will not review in detail how these constraints arise, the essence of the argument is that the supersymmetry algebra requires the existence of several *parallel* geometric objects on $\mathcal{M}$, restricting its holonomy [25]. For instance, in minimal supersymmetry, i.e. $\mathcal{N} = 1$, the field space $\mathcal{M}$ is required to be Kähler. For higher supergravity theories, which corresponds to $\mathcal{N} > 2$ in four dimensions, the situation is as follows [25].

Proposition 6.1.1. *Let $\mathcal{M}$ be the scalar field space of a higher supergravity theory in four, six, or ten dimensions, and let $\widetilde{\mathcal{M}}$ be its universal cover. Then $\widetilde{\mathcal{M}}$ is a symmetric space.*

To see what this implies for the geometry of the field space, let us recall the definition of a symmetric and locally symmetric space. If X is a Riemannian manifold, then a *geodesic symmetry* at a point x is a locally defined map s_x which inverts the geodesics through x. In terms of the exponential map $\exp_x : T_xX \to X$, the geodesic symmetry s_x sends a point $\exp_x(v)$ to $\exp_x(-v)$.

Definition 6.1.2. A complete Riemannian manifold X is *locally symmetric* if for any point $x \in X$, there exist a neighbourhood $x \in U \subseteq X$ on which the geodesic symmetry $s_x : U \to U$ is an isometry. If for every $x \in X$, the geodesic symmetry s_x exists globally and defines an isometry $s_x : X \to X$, then X is *symmetric*.

As a more practical definition, we remark that X is locally symmetric if and only if its Riemann tensor satisfies $\nabla_\lambda \mathcal{R}_{\mu\nu\rho}{}^\sigma = 0$ [25]. Symmetric spaces have a special structure which is closely related to a type of space we have encountered many times in this thesis. If X is a symmetric space and G is the connected component of the group of isometries of X, then it can be shown that G is in fact a Lie group and that it acts transitively on X. In order words, every pair of points in X can be connected via an isometry. Thus, if we let $K \subseteq G$ denote the stabilizer subgroup of an arbitrary point in X, which is a maximal compact subgroup, we deduce the following [66].

Theorem 6.1.3. *The symmetric space X is isometric to G/K.*

In the context of higher supergravity, we thus find that the universal covers of the scalar field spaces are symmetric spaces of the form G/K for a Lie group G and a maximal compact subgroup K. In fact, higher supersymmetry does not only specify that $\widetilde{\mathcal{M}}$ must be symmetric, but it also specifies precisely which groups G and K should be. For various dimensions D and number of supersymmetry generators $\mathcal{N}$, the uniquely determined groups G and K are shown [25]. Some of the groups in this table are non-standard, such as $\mathrm{SO}^*(12)$ or $\mathrm{E}_{7(7)}$. Since we will not study these cases explicitly, we refer to [25] for an explanation. In some cases the number $\mathcal{N}$ is split up into $(\mathcal{N}_\mathrm{L}, \mathcal{N}_\mathrm{R})$, referring to the chirality of the supersymmetry generators.

D	$\mathcal{N}$	G	K
4	3	$\mathrm{SU}(3,k)$	$\mathrm{SU}(3)\times\mathrm{U}(k)$
4	4	$\mathrm{SL}(2,\mathbb{R})\times\mathrm{SO}(6,k)$	$\mathrm{SO}(2)\times\mathrm{SO}(6)\times\mathrm{SO}(k)$
4	5	$\mathrm{SU}(5,1)$	$\mathrm{U}(5)$
4	6	$\mathrm{SO}^*(12)$	$\mathrm{U}(6)$
4	8	$\mathrm{E}_{7(7)}$	$\mathrm{SU}(8)$
6	(2,2)	$\mathbb{R}\times\mathrm{SO}(4,k)$	$\mathrm{SO}(4)\times\mathrm{SO}(k)$
6	(4,0)	$\mathrm{SO}(5,k)$	$\mathrm{SO}(5)\times\mathrm{SO}(k)$
6	(4,2)	$\mathrm{SO}(5,1)$	$\mathrm{SO}(5)$
6	(6,0)	$\mathrm{SU}^*(6)$	$\mathrm{Sp}(6)$
6	(4,4)	$\mathrm{SO}(5,5)$	$\mathrm{SO}(5)\times\mathrm{SO}(5)$
6	(6,2)	$\mathrm{F}_{4(4)}$	$\mathrm{Sp}(6)\times\mathrm{SU}(2)$
6	(8,0)	$\mathrm{E}_{6(6)}$	$\mathrm{Sp}(8)$
10	(1,0)	$\mathbb{R}$	-
10	(1,1)	$\mathbb{R}$	-
10	(2,0)	$\mathrm{SL}(2,\mathbb{R})$	$\mathrm{SO}(2)$

Table 6.1: Classical global symmetry groups G for higher supergravity theories and their compact subgroup K for various choices of dimension D and amount of supersymmetry $\mathcal{N}$.

As a familiar example, the bottom entry of the table corresponds to Type IIB supergravity in ten dimensions. In this theory there are two real scalars, namely the axion and the dilaton, which we combined into the axio-dilaton τ. In Chapter 2 we have indeed seen that the axio-dilaton takes values in the field space $\mathbb{H}=\mathrm{SL}(2,\mathbb{R})/\mathrm{SO}(2)$. In general, the group G has an interpretation in higher supergravity as the classical global symmetry group of the Lagrangian. By this we mean that the classical equations of motion of the theory are invariant under an action of G, but that this invariance is broken at the quantum level. For example, the Type IIB supergravity action shown in equation 2.2 is invariant under $\mathrm{SL}(2,\mathbb{R})$, but the presence of D7-branes breaks this symmetry to $\mathrm{SL}(2,\mathbb{Z})$. Generally, the interpretation of G as a symmetry group is clear for the scalar sector, where G is a group of isometries of the scalar field space $\mathcal{M}$. For the other sectors, the fields are organized in representations of G, as we will explain in more detail later.

6.1.2 Field Spaces, Arithmeticity, and the Swampland

The fact that the universal cover of the field space in higher supergravity should be a symmetric space can be thought of as specifying the local geometry of the field space. The most general space that preserves this local geometry is a quotient of the form

$$\mathcal{M}=\Gamma\backslash\widetilde{\mathcal{M}}, \tag{6.1}$$

where Γ is a discrete subgroup of the isometry group of $\widetilde{\mathcal{M}}$ [25]. We note that the resulting field space is not necessarily smooth, since Γ could act non-freely[1]. Since $\mathcal{M}$ is a symmetric space isomorphic to G/K, it follows that the field space in a higher supergravity is of the form $\mathcal{M}=\Gamma\backslash G/K$, which appears

[1] Despite their non-smoothness, spaces of this form are still often called locally symmetric spaces [66].

to be precisely the sort of space we have encountered in the previous chapters. There is an important subtlety here, however: supersymmetry does not constrain Γ, other than that it should be a discrete subgroup of G. In particular, Γ is not required to be an arithmetic subgroup, which means that $\mathcal{M}$ is *not* necessarily an arithmetic quotient. To discuss this further, we first require a better understanding of Γ.

The appearance of the discrete subgroup Γ has two interpretations. One of them is geometric, since the unspecified global geometry of $\mathcal{M}$ allows for a quotient by a discrete subgroup. The second interpretation is physical and originates from quantizing the theory. At the quantum level, only symmetry transformations which preserve Dirac quantization are allowed. The largest group preserving this quantization condition is $G_{\mathbb{Z}}$, and this breaks the global symmetry group G to a discrete subgroup $\mathcal{U} \subseteq G_{\mathbb{Z}}$, which is sometimes called the U-duality[2] subgroup [84]. In the setting of string theory, these U-dualities are expected to be exact symmetries of the quantized theory [85]. Taking a quotient $G/K \to \Gamma\backslash G/K$ by a subgroup $\Gamma \subseteq \mathcal{U}$ amounts to 'gauging' a subgroup of U-dualities. This notion of gauging should be understood as identifying scalar field values which lie in the same Γ-orbit.

According to the Tameness Conjecture, effective field theories arising from quantum gravity should have a field space which is definable in an o-minimal structure. In the present setting of higher supergravity theories, we have seen in Chapter 4 that the double quotient space $\Gamma\backslash G/K$ has a definable structure provided that Γ is an *arithmetic* subgroup of G, i.e. a finite index sugroup of $G_{\mathbb{Z}}$. Supersymmetry does not require Γ to be arithmetic, which suggests that there are potentially higher supergravities with a non-definable field space. However, there are relations between arithmeticity and other swampland conjectures.

Recall that the Completeness Hypothesis (Conjecture 2.6.4) states that an effective field theory in the landscape must have states of all possible charges consistent with Dirac quantization [82]. In [25] it was noted that this strongly suggests that the U-duality group $\mathcal{U}$ should be equal to the full group $G_{\mathbb{Z}}$, which is the largest group consistent with Dirac quantization. This allows a complete set of charges to be reached by U-duality transformations.

The subgroup $\Gamma \subseteq \mathcal{U}$ plays the role of the gauged subgroup of the U-duality group. By the arguments above, we find that Γ is a subgroup of $G_{\mathbb{Z}}$. The No Global Symmetries Conjecture (Conjecture 2.6.3) asserts that a consistent theory of quantum gravity may not possess any global symmetries, so that any apparent global symmetry must be broken or gauged [82]. A weaker variant of this conjecture claims that global symmetries may be present, but that the global symmetry group must be finite. In the setting of higher supergravities, the global symmetry group is the U-duality group $\mathcal{U}$, before the quotient by Γ is taken. Here it is important to note that the U-duality group is an exact symmetry of the quantum theory which is believed to extend to a UV-completion in string theory.

[2] The 'U' stands for *unified*, which refers to its origin as a unification of S- and T-duality in string theory.

As noted in [25], consistency with the weaker No Global Symmetries Conjecture implies that almost all of $\mathcal{U}$ must be gauged. Here the notion of 'almost all' is made precise by requiring that the gauged subgroup $\Gamma \subseteq \mathcal{U}$ must be such that the remaining global symmetry group, given by the quotient $\mathcal{U}/\Gamma$, is finite. Taking $\mathcal{U}$ to be $G_{\mathbb{Z}}$ this means that we require that Γ is a subgroup of finite index in $G_{\mathbb{Z}}$, which is precisely the statement that Γ is arithmetic. We thus find that the presence of higher supersymmetry and the assumption of some of the widely accepted swampland conjectures together imply the first part of the Tameness Conjecture, as summarized in the following diagram:

$$\begin{gathered}\text{higher supersymmetry } + \text{ completeness } + \text{ no global symmetries}\\ \Downarrow \\ \text{tameness of field space}\end{gathered}$$

At present, the precise meaning of this implication arrow is unclear. The arithmeticity of Γ implies that the tools of arithmetic group theory are available, which allowed us to construct a fundamental set for arithmetic quotients by means of Siegel sets. If Γ is not arithmetic then it is not clear how to construct such a fundamental set, thus obscuring the tameness of the field space $\mathcal{M} = \Gamma\backslash G/K$. Even without arithmeticity, the field space can still be definable in an o-minimal structure, but the existence of a fundamental set is not clear in this case.

To close this subsection, let us comment on the non-smoothness of $\mathcal{M}$. As mentioned above, if the action of Γ on G/K is not free, then the double quotient $\Gamma\backslash G/K$ need not be smooth. However, the observation that $\Gamma \subseteq G_{\mathbb{Z}}$ and that K is compact implies that the stabilizers of this action are finite, so that the singularities of $\mathcal{M}$ are of orbifold type, admitting a finite smooth covering. In [86], a more detailed discussion of the singularities of the field space is given, where it is proposed that one can work with a finite smooth cover of $\mathcal{M}$. Instead, we will work around the singularities in a naive way by focusing on the non-singular part $\mathcal{M}_{\mathrm{s}} \subseteq \mathcal{M} \setminus \mathcal{M}_{\mathrm{sing}}$.

6.1.3 Siegel Sets for Supergravity

We have now argued that field spaces in higher supergravity theories consistent with string theory are arithmetic quotients $\mathcal{M} = \Gamma\backslash G/K$. This enables us to use the theory of arithmetic groups to construct a fundamental set for the action of Γ, consisting of a finite number of Siegel sets. In this subsection we construct a Siegel set for the group $\mathrm{SO}(p,q)$, which appears in the field space of several higher supergravity theories, as shown in Table 6.1. An example of particular interest is $\mathcal{N} = 4$ supergravity in four dimensions, which we will discuss in more detail later. This subsection has the dual purpose of increasing our understanding of the higher supergravity field space and further exemplifying the constructiong of Siegel sets.

In the following we assume that $p \leq q$ without loss of generality. The first step to take is to write down a minimal parabolic subgroup P of $\mathrm{SO}(p,q)$. In other words, we are interested in a Borel subgroup of $\mathrm{SO}(p,q)$. Since Borel subgroups are unique up to conjugation, other parabolic subgroups may be obtained as subgroups containing a conjugate of P [71]. To this end, it is convenient to consider a different basis in which there is a minimal parabolic subgroup P with a simple structure.

Consider the symmetric $(p+q)\times(p+q)$ matrix

$$Q = \begin{pmatrix} 0 & 0 & J_p \\ 0 & I_{q-p} & 0 \\ J_p & 0 & 0 \end{pmatrix}, \tag{6.2}$$

where the matrices I_{q-p} and J_p are given by

$$I_{q-p} = \underbrace{\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}}_{q-p}, \quad J_p = \underbrace{\begin{pmatrix} 0 & & \cdots & 1 \\ \vdots & & \iddots & \vdots \\ 0 & 1 & & 0 \\ 1 & 0 & \cdots & 0 \end{pmatrix}}_{p}. \tag{6.3}$$

The matrix Q induces a bilinear pairing of indefinite signature (p,q), and thus the resulting group of determinant 1 matrices preserving this pairing is $\mathrm{SO}(p,q)$. It is conjugate to the standard parametrization[3] of $\mathrm{SO}(p,q)$ via the transformation $A\mapsto \beta^{-1}A\beta$, where β is the change of basis matrix

$$\beta = \begin{pmatrix} \frac{1}{\sqrt{2}}I_p & 0 & \frac{1}{\sqrt{2}}J_p \\ 0 & I_{q-p} & 0 \\ -\frac{1}{\sqrt{2}}J_p & 0 & \frac{1}{\sqrt{2}}I_p \end{pmatrix}, \quad \beta^{-1} = \begin{pmatrix} \frac{1}{\sqrt{2}}I_p & 0 & -\frac{1}{\sqrt{2}}J_p \\ 0 & I_{q-p} & 0 \\ \frac{1}{\sqrt{2}}J_p & 0 & \frac{1}{\sqrt{2}}I_p \end{pmatrix}. \tag{6.4}$$

A minimal parabolic subgroup of $\mathrm{SO}(p,q)$ is then given by [87]

$$P = \left\{ \begin{pmatrix} A & C & D \\ 0 & B & E \\ 0 & 0 & A^\dagger \end{pmatrix} \middle| \begin{array}{c} A\in \mathrm{GL}(p,\mathbb{R}) \text{ upper triangular},\ B\in\mathrm{SO}(q-p), \\ E^{\mathrm{T}}B + J_pA^{-1}C = 0, \\ E^{\mathrm{T}}E + J_pA^{-1}D + (J_pA^{-1}D)^{\mathrm{T}} = 0 \end{array} \right\}, \tag{6.5}$$

where the adjoint $A^\dagger$ is defined as

$$A^\dagger = J_p(A^{-1})^{\mathrm{T}}J_p.$$

The group P now has a Langlands decomposition $P = N_P\cdot A_P\cdot M_P$, where N_P is equal to the subgroup $\mathrm{U}_{p+q}\cap P$ of upper triangular matrices in P with 1s on the diagonal, and

$$A_P = \left\{ \begin{pmatrix} A & 0 & 0 \\ 0 & I_{q-p} & 0 \\ 0 & 0 & A^\dagger \end{pmatrix} \middle| A = (t_1,\dots,t_n),\, t_1,\dots,t_n > 0 \right\}, \tag{6.6}$$

$$M_P = \left\{ \begin{pmatrix} A & 0 & 0 \\ 0 & B & 0 \\ 0 & 0 & A^\dagger \end{pmatrix} \middle| A = (\pm1,\dots,\pm1) \right\}. \tag{6.7}$$

In order to obtain Siegel sets for $\mathrm{SO}(p,q)$, we have to identify a set of simple characters. Consider a general element $a\in A_P$ given by $a = \mathrm{diag}(t_1,\dots,t_p,1,\dots,1,t_p^{-1},\dots,t_1^{-1})$. A possible choice of simple characters is then

$$\chi_1(a) = t_1/t_2, \quad \chi_2(a) = t_2/t_3, \quad \dots \quad \chi_{p-1}(a) = t_{p-1}/t_p, \quad \chi_p(a) = t_1t_p. \tag{6.8}$$

[3]By the standard representation we mean the group of determinant 1 matrices which preserve the signature (p,q) pairing $\eta = \mathrm{diag}(-1,\dots,-1,1,\dots,1)$.

Let us illustrate this for a low-dimensional example with $(p, q) = (2, 4)$. Elements in A_P take the form

$$a = \begin{pmatrix} t_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & t_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & t_2^{-1} & 0 \\ 0 & 0 & 0 & 0 & 0 & t_1^{-1} \end{pmatrix}, \tag{6.9}$$

and according to the prescription above we should find two simple characters, with a possible choice given by $\chi_1(a) = t_1/t_2$, $\chi_2(a) = t_1 t_2$. At the level of the algebra $\mathfrak{a}_P$, these may be realized using the root vectors

$$\nu_1 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \nu_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{6.10}$$

i.e. we have $a\,\nu_1\,a^{-1} = t_1 t_2^{-1}\,\nu_1$ and $a\,\nu_2\,a^{-1} = t_1\,t_2\,\nu_2$ with $\nu_1, \nu_2 \in \mathfrak{n}_P$.

In general, the construction of a Siegel set is now completed by selecting a real number λ, and bounded open sets $U \subseteq N_P$ and $V \subseteq M_P \cdot K$, where K is the maximal compact subgroup $\mathrm{SO}(p) \times \mathrm{SO}(q)$. Note that the group M_P is compact, so for V one may choose the whole factor $M_P \cdot K$ of the Langlands decomposition. The resulting Siegel set is given by

$$\mathfrak{S} = U \cdot \left\{ \mathrm{diag}(t_1, \ldots, t_p, 1, \ldots, 1, 1/t_p, \ldots, 1/t_1) \,\middle|\, t_1/t_2 > \lambda, \ldots, t_{p-1}/t_p > \lambda, t_1 t_p > \lambda \right\} \cdot V. \tag{6.11}$$

6.1.4 The Field Space Metric

We now turn to a discussion of tameness of the coupling functions of the scalar fields, i.e. the metric $\mathcal{G}$ on the field space. In view of the fact that G acts as a global symmetry group, the crucial property of this metric is that it is G-invariant. As such, the metric on $\mathcal{M}$ originates from a G-invariant metric $\tilde{\mathcal{G}}$ on the group G, so we begin our discussion by discussing the definability of this metric. We provide two perspectives: a practical perspective in terms of coordinates, and a more intrinsic perspective in terms of the geometry of G.

For the coordinate perspective on the metric, we consider a set of local coordinates ϕ^i parametrizing a chart $U \subseteq G$. Since the group G is a matrix Lie group, it is definable in $\mathbb{R}_{\mathrm{alg}}$, and hence we choose these coordinates to be definable. We think of these definable coordinates as parametrizing a matrix $\mathcal{V}(\phi)$ in G. In this setting, the invariant metric $\tilde{\mathcal{G}}$ which is relevant for higher supergravity is given by

$$\tilde{\mathcal{G}} = \tilde{\mathcal{G}}_{ij}\, d\phi^i\, d\phi^j = -\mathrm{tr}[(\mathcal{V}^{-1}\partial_i \mathcal{V})(\mathcal{V}^{-1}\partial_j \mathcal{V})]\, d\phi^i\, d\phi^j, \tag{6.12}$$

where the indices $i, j = 1, \ldots, m$ refer to the scalar fields (for instance $\partial_i \mathcal{V} = \partial \mathcal{V}(\phi)/\partial \phi^i$). The interpretation of this formula, is that it is the trace of the square of the $\mathfrak{g}$-valued Maurer-Cartan form

$$\mathcal{V}^{-1}\partial_i \mathcal{V}\, d\phi^i, \tag{6.13}$$

which is left-invariant by construction. This metric is thus invariant under global G-transformations of the form $\mathcal{V}(\phi) \mapsto \Lambda\mathcal{V}(\phi)$ with $\Lambda \in G$. Note that, as a metric for G itself, $\tilde{\mathcal{G}}_{ij}$ is not guaranteed to be positive definite if G is non-compact. The equation above shows that the invariant metric on G is $\mathbb{R}_{\text{alg}}$-definable, since matrix inversion, differentiation, and taking the trace are operations that preserve definability. Here we use Proposition 3.6.5 to prove definability of the metric $\tilde{\mathcal{G}}$ in terms of the component functions $\tilde{\mathcal{G}}_{ij}$.

Let us now provide a more geometric point of view, from which the definability of this invariant metric is not surprising. Recall that any Lie group comes with a set of left-invariant vector fields, which are defined as follows. Every element $g \in G$ defines a diffeomorphism $L_g : G \to G$ by left translation by g. The left-invariant vector fields on G are characterized by

$$v(gh) = (TL_g)_h v(h), \tag{6.14}$$

and are in one-to-one correspondence with Lie algebra elements $X \in \mathfrak{g}$ via

$$v_X(g) = (TL_g)_e X. \tag{6.15}$$

> **Proposition 6.1.4.** *Let G be a matrix Lie group. Then the left-invariant vector fields on G are definable in $\mathbb{R}_{\text{alg}}$.*

Proof. Fix an element $X \in \mathfrak{g}$. Since G is a matrix Lie group, the group multiplication map $G \times G \to G$ is $\mathbb{R}_{\text{alg}}$-definable, and by differentiating with respect to the second factor along vector X, we find that the left invariant vector field v_X are definable, since differentation preserves differentiability. □

If we pick a basis $X_1, \ldots, X_m$ of the Lie algebra $\mathfrak{g}$, we thus obtain a definable global frame $v_{X_1}, \ldots, v_{X_m}$ of the tangent bundle TG. It follows that there is an $\mathbb{R}_{\text{alg}}$-definable global trivialization

$$TG \to G \times \mathfrak{g}. \tag{6.16}$$

Consequently, any tensor bundle on G can be globally trivialized by a definable map. From the point of view of this trivialization, a left-invariant tensor field $X : G \to T^{p,q}G$ is just constant, i.e. a map

$$G \to G \times \mathfrak{g}^p \otimes (\mathfrak{g}^*)^q \tag{6.17}$$

$$g \mapsto (g, X_0) \tag{6.18}$$

for some fixed element $X_0 \in \mathfrak{g}^p \otimes (\mathfrak{g}^*)^q$. We thus conclude the following.

> **Proposition 6.1.5.** *Let G be a matrix Lie group. Then any left-invariant tensor field on G is definable in $\mathbb{R}_{\text{alg}}$.*

In particular, it follows that the invariant metric $\tilde{\mathcal{G}}$ on G is definable.

The next step is to consider the metric on the symmetric space G/K, and again we start from a coordinate point of view. In the supergravity literature, the conventional way of working with the metric on G/K is to start with the metric on G, and view the action of the compact group K as an artificial 'gauge symmetry' [88]. This gauge symmetry is not associated to any propagating degrees of freedom, but only a formal procedure to remove the redundant dependence on K.

The Lie algebra $\mathfrak{g}$ of G decomposes as $\mathfrak{g} = \mathfrak{p} \oplus \mathfrak{k}$, where $\mathfrak{k}$ is the Lie algebra of the maximal compact subgroup K and $\mathfrak{p}$ is a complementary Lie algebra, orthogonal to $\mathfrak{k}$ with respect to the Killing form on $\mathfrak{g}$. The Maurer-Cartan form $\mathcal{V}^{-1}\partial_i\mathcal{V}\, d\phi^i$ on G is $\mathfrak{g}$-valued, and to obtain a metric for G/K we project out the $\mathfrak{k}$-component via a linear projection map $(\cdots)_{\mathfrak{p}} : \mathfrak{g} \to \mathfrak{p}$. In terms of a matrix parametrization $\mathcal{V}(\phi)$, the metric on G/K is then given by

$$\tilde{\mathcal{G}} = \tilde{\mathcal{G}}_{ij}\, d\phi^i\, d\phi^j = -\mathrm{tr}[(\mathcal{V}^{-1}\partial_i\mathcal{V})_{\mathfrak{p}}(\mathcal{V}^{-1}\partial_j\mathcal{V})_{\mathfrak{p}}]\, d\phi^i\, d\phi^j, \tag{6.19}$$

This suggests that the metric on the homogeneous space G/K is definable as well, since linear projections are always definable by the third axiom of o-minimal structures.

We note that it is possible to formulate the metric on G/K in an equivalent way that avoids the need to project down to the Lie algebra $\mathfrak{p}$ [88]. The idea behind this formulation is to work with manifestly K-invariant objects. If η is a K-invariant matrix, i.e. $\eta = k\,\eta\,k^{\mathrm{T}}$ for any $k \in K$, then the matrix $M(\phi) = \mathcal{V}(\phi)\,\eta\,\mathcal{V}(\phi)^{\mathrm{T}}$ is invariant under local K-transformations $\mathcal{V} \mapsto \mathcal{V}k(\phi)$. A calculation shows that, expressed using the matrix $M(\phi)$, the metric in (6.12) becomes

$$\tilde{\mathcal{G}} = \frac{1}{4}\,\mathrm{tr}[(\partial_i M)(\partial_j M^{-1})]\, d\phi^i\, d\phi^j. \tag{6.20}$$

Let us illustrate this for a familiar example.

Example 6.1.6. Consider the symmetric space $G/K = \mathrm{SL}(2,\mathbb{R})/\mathrm{SO}(2)$, which is relevant for the scalar field space of Type IIB supergravity. We can parametrize an arbitrary element in $\mathrm{SL}(2,\mathbb{R})$ by using the Langlands decomposition discussed in Chapter 4,

$$\mathcal{V}(t,s,\theta) = \begin{pmatrix} t & 0 \\ 0 & t^{-1} \end{pmatrix} \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}, \tag{6.21}$$

with $t \neq 0$. The maximal compact subgroup $K = \mathrm{SO}(2)$ preserves the standard Euclidean metric, so the matrix $M = \mathcal{V}\,\mathcal{V}^{\mathrm{T}}$ takes the form

$$M(t,s) = \begin{pmatrix} (1+s^2)t^2 & s \\ s & t^{-2} \end{pmatrix}. \tag{6.22}$$

Introducing the complex coordinate $\tau = st^2 + i\,t^2 \in \mathbb{H}$, this matrix becomes

$$M(\tau) = \frac{1}{\mathrm{Im}\,\tau} \begin{pmatrix} |\tau|^2 & \mathrm{Re}\,\tau \\ \mathrm{Re}\,\tau & 1 \end{pmatrix}, \tag{6.23}$$

which is precisely the parametrization for the $\mathrm{SL}(2,\mathbb{R})/\mathrm{SO}(2)$ quotient often encountered in the supergravity literature [89]. Using equation (6.20), the resulting metric is the familiar hyperbolic metric for the upper half plane,

$$\tilde{\mathcal{G}} = \frac{1}{(\mathrm{Im}\,\tau)^2} d\tau\, d\bar{\tau}.$$

This is precisely the metric encountered in the scalar kinetic term of the action in equation (2.2).

Since this gauging procedure makes the mathematical structure of the metric slightly obscure, we now again turn to a more geometric perspective. The metric on G/K is a section of the tensor bundle $T^{0,2}G/K$, so the first step is to understand the structure of this bundle. The tangent bundle on G/K can be interpreted as an associated vector bundle to the principal K-bundle $G \to G/K$ and the adjoint representation[4] of K on $\mathfrak{g}/\mathfrak{k}$ [90, p. 55]. In other words, we have

$$T(G/K) \cong G \times_K \mathfrak{g}/\mathfrak{k} = (G \times \mathfrak{g}/\mathfrak{k})/K. \tag{6.24}$$

Here K acts on G via the group multiplication and on the quotient $\mathfrak{g}/\mathfrak{k}$ via the adjoint representation. This idea extends to any tensor bundle over G/K, and we have

$$T^{p,q}(G/K) \cong G \times_K \big((\mathfrak{g}/\mathfrak{k})^p \otimes (\mathfrak{g}/\mathfrak{k})^{*q}\big), \tag{6.25}$$

where the representation of K on $(\mathfrak{g}/\mathfrak{k})^p \otimes (\mathfrak{g}/\mathfrak{k})^{*q}$ is formed by taking tensor powers of the adjoint representation on $\mathfrak{g}/\mathfrak{k}$. Since G acts on G/K by diffeomorphisms, there is again a notion of left-invariant tensor fields.

Proposition 6.1.7. *Let G be a matrix Lie group, and let K be a maximal compact subgroup. Then any left-invariant tensor field on G/K is definable in $\mathbb{R}_{\mathrm{alg}}$.*

Proof. Let us fix the type (p, q). In the proof of this proposition it will be convenient to use the perspective on definable sections of bundles given in Observation 3.6.4. From this perspective, we have to show that the image of a left-invariant (p, q)-tensor field is a definable subset of $T^{p,q}(G/K)$. Fix an element $X_0 \in (\mathfrak{g}/\mathfrak{k})^p \otimes (\mathfrak{g}/\mathfrak{k})^{*q}$ which is invariant under the action of K, and consider the definable subset

$$G \times \{X_0\} \subseteq G \times \big((\mathfrak{g}/\mathfrak{k})^p \otimes (\mathfrak{g}/\mathfrak{k})^{*q}\big). \tag{6.26}$$

The tensor bundle $T^{p,q}(G/K)$ is obtained by taking the quotient by K, and since K is a compact group acting algebraically, the projection map

$$\pi : G \times \big((\mathfrak{g}/\mathfrak{k})^p \otimes (\mathfrak{g}/\mathfrak{k})^{*q}\big) \to \Big(G \times \big((\mathfrak{g}/\mathfrak{k})^p \otimes (\mathfrak{g}/\mathfrak{k})^{*q}\big)\Big)/K \tag{6.27}$$

is $\mathbb{R}_{\mathrm{alg}}$-definable (analogous to the argument of Example 3.4.9). The image of the slice $G \times \{X_0\}$ under π is a submanifold of $T^{p,q}(G/K)$ diffeomorphic to G/K; it coincides with the image of the section

$$\begin{aligned} \sigma_{X_0} : G/K &\to \Big(G \times \big((\mathfrak{g}/\mathfrak{k})^p \otimes (\mathfrak{g}/\mathfrak{k})^{*q}\big)\Big)/K \\ gK &\mapsto (g, X_0)K. \end{aligned}$$

By construction, this section defines a left-invariant tensor field. By a geometric version of the Frobenius reciprocity theorem [90, p. 54], there is a one-to-one correspondence between invariant sections of $T^{p,q}(G/K)$ and K-invariant elements in $(\mathfrak{g}/\mathfrak{k})^p \otimes (\mathfrak{g}/\mathfrak{k})^{*q}$. In other words, all left-invariant (p, q)-tensor fields on G/K are of the form constructed above. Since the image of this section coincides with the definable set $\pi(G \times \{X_0\})$, we conclude that σ_{X_0} is definable. □

[4] More precisely, the representation is induced by the restriction of the adjoint representation of G on $\mathfrak{g}$ to K.

We now turn to the invariant metric on the field space $\mathcal{M} = \Gamma\backslash G/K$. In view of the discussion in Section 6.1.2, we assume that Γ is an arithmetic subgroup so that $\mathcal{M}$ is an arithmetic quotient. As mentioned earlier in this chapter, this space may have orbifold singularities arising from fixed points of the action of Γ. In what follows we will work around these singularities by simply removing them from the field space. For our purposes, this is an acceptable procedure since the set of singularities $\mathcal{M}_{\text{sing}}$ is definable. Equivalenly, the non-singular space $\mathcal{M}_{\text{n.s.}} = \mathcal{M}_{\text{sing}}$ is definable. To see this, let $F \subseteq G/K$ be a fundamental set for the action of Γ, and let Γ_F be the finite set

$$\Gamma_F = \{\gamma \in \Gamma \,|\, \gamma F \cap F \neq \emptyset\}. \tag{6.28}$$

The set of fixed points

$$F_{\text{sing}} = \{z \in F \,|\, \gamma z = z \text{ for some non-trivial } \gamma \in \Gamma_F\} \tag{6.29}$$

is closed and definable, since it is constructed in terms of a first-order formula. Note that in order to conclude that this is a first-order formula, it is crucial that the set Γ_F is finite (and hence definable). The set of singularities of $\mathcal{M}$ is now given by the image $\pi(F_{\text{sing}})$, and since the projection $\pi : F \to \Gamma\backslash G/K$ is definable, we conclude that $\mathcal{M}_{\text{sing}}$ is definable. Moreover, since Γ acts freely on the open submanifold $F_{\text{s}} = F\backslash F_{\text{sing}}$, the space $\mathcal{M}_{\text{s}} = \pi(F_{\text{s}})$ is smooth [91].

Locally, the metric $\mathcal{G}$ on $\mathcal{M}_{\text{s}}$ is identical to the metric $\tilde{\mathcal{G}}$ on G/K. Therefore, from a local point of view, definability of the metric $\mathcal{G}$ follows almost directly. However, as we have seen several times in this thesis, definability can only be checked locally on a finite definable open cover. This cover should consist of a number of open subsets of G/K which may be identified to opens in $\mathcal{M}_{\text{sing}}$ by means of the projection map π.

We construct such a cover as follows. Let $\{U_i\}$ be a definable open cover of $\mathcal{M}_{\text{s}}$, where i ranges over a finite index set. By the triangulation theorem (Theorem 3.3.10), the opens U_i may be assumed to be simply connected. For each i, let V_i be the preimage of U_i under the projection map $\pi : F_{\text{s}} \to \mathcal{M}_{\text{s}}$. Since we restrict to a fundamental set, the preimage of any point in $\mathcal{M}_{\text{s}}$ will consist of finitely many points. In turn, this implies that each V_i has finitely many connected components. Denote these connected components by V_{ij}, with j ranging over a finite index set. Since the connected components of a definable set are definable, each V_{ij} is definable.

The set $\{V_{ij}\}$ provides the required open cover, since the restriction π_{ij} of the projection map π to V_{ij} is a diffeomorphism onto its image. To see this, it suffices to note that π_{ij} is injective. We argue that this is the case by contradiction. If π_{ij} would not be injective, then there would be a point $z \in V_{ij}$ such that there is an element $\gamma \in \Gamma$ with $\gamma z \in V_{ij}$ and $\gamma z \neq z$. Since V_{ij} is path-connected, there exists a path in V_{ij} from z to γz. By covering space theory, such a path projects down to a loop in U_i with a non-trivial homotopy class [92]. However, this is a contradiction, since U_i was assumed to be simply connected.

Letting $U_{ij} \subseteq \mathcal{M}_{\text{s}}$ denote the image of V_{ij} under the projection, we conclude that $\pi_{ij} : V_{ij} \to U_{ij}$ is a definable diffeomorphism. The metric on U_{ij} is now given by

$$\mathcal{G} = (\pi_{ij}^{-1})^* \tilde{\mathcal{G}}, \tag{6.30}$$

which is definable since it is the pullback of a definable metric along a definable map. Since the cover $\{U_{ij}\}$ of $\mathcal{M}_{\rm s}$ is definable, this completes the proof that the metric $\mathcal{G}$ on the field space $\mathcal{M}_{\rm s}$ is definable in $\mathbb{R}_{\rm alg}$. Altogether we have now shown that the scalar sector of a higher supergravity is definable in an o-minimal structure, showing that the Tameness Conjecture holds in this setting.

6.1.5 Infinite Distances in Higher Supergravity

To close the section on the scalar sector, we briefly discuss how Siegel sets provide insights on the structure of infinite distance limits in the field space, which are of interest for the Distance Conjecture. Starting with a simple example, consider the ten-dimensional Type IIB supergravity, whose scalar field space is the arithmetic quotient $\mathrm{SL}(2,\mathbb{Z})\backslash\mathrm{SL}(2,\mathbb{R})/\mathrm{SO}(2) = \mathrm{SL}(2,\mathbb{Z})\backslash\mathbb{H}$. In terms of real coordinates $x+iy \in \mathbb{H}$, the invariant metric on the upper-half plane constructed above takes the form

$$\tilde{\mathcal{G}} = \frac{1}{y^2}(dx^2+dy^2).$$

Restricting to the standard fundamental domain $\mathcal{F}$, we see that the field space of this metric has one infinite distance limit, corresponding to $y\to\infty$. From any point $x_0+iy_0 \in \mathcal{F}$, the geodesic distance to a point $x+iy$ with depends on the coordinate y as

$$d(x_0+iy_0, x+iy) \sim \left|\int_{y_0}^{y} \sqrt{1/t^2}\, dt\right| = \big|\log(y-y_0)\big|, \qquad (6.31)$$

where the absolute value accounts for the correct sign depending on whether $\tilde{y} > y$ or $\tilde{y} < y$. This logarithmic divergence has an important interpretation from the perspective of S-duality. In the present setting, S-duality is the $\mathrm{SL}(2,\mathbb{Z})$-transformation that acts as $\tau \mapsto -1/\tau$ on the upper-half plane. If we had taken our fundamental domain to be the S-dual image of $\mathcal{F}$, then the limit $y\to\infty$ is mapped to $y\to 0$, and the logarithm ensures that this limit is also at infinite distance[5]. Below we will argue that this logarithmic behavior is not specific to this example, but in fact a universal phenomenon in any higher supergravity theory whose field space is an arithmetic quotient.

Suppose that the subgroup Γ is arithmetic. Then any Siegel set $\mathfrak{S} \subseteq G/K$ has the general form

$$\mathfrak{S} = U \cdot A_{P,\lambda} \cdot V,$$

where we recall that U and V are chosen as bounded sets, and $A_{P,\lambda} \subseteq A_P$ is the cone defined by

$$A_{P,\lambda} = \{a \in A_P \,|\, \chi(a) > \lambda \ \text{ for all simple characters } \ \chi \in \Delta(A_P, N_P)\}.$$

The group A_P is a real torus, isomorphic to a power of the multiplicative group $(\mathbb{R}_{>0})^r$ for some $r \geq 0$. In any fundamental set F constructed from a finite union of Siegel sets, the boundedness of U and V implies that all the unbounded directions in F come from the cones $A_{P,\lambda}$. Identifying A_P with $\mathbb{R}^r_{>0}$ and using coordinates $(t_1,\dots,t_r) \in (\mathbb{R}_{>0})^r$, the invariant metric is given by

$$\tilde{\mathcal{G}} = \sum_{j=1}^{r} \frac{1}{t_j^2}\, dt_j^2.$$

[5] This statement is of course not surprising in view of the invariance of the metric.

The subset $A_{P,\lambda} \subseteq A_P$ is a cone whose direction is determined by the choice of simple characters. As a consequence of the structure of the metric $\tilde{\mathcal{G}}$, any path towards the boundary of A_P constitutes an infinite distance limit. More precisely, the geodesic distance from a generic point in A_P to another point $(t_1, \ldots, t_r)$ with $t_j \to 0$ or $t_j \to \infty$ is asymptotic to $|\log(t_j)|$. Since a fundamental set for the arithmetic quotient is constructed from a finite union of Siegel sets, we find:

> **Observation 6.1.8.** In a higher supergravity theory for which Γ is arithmetic, any infinite distance limit in the scalar field space comes from a cone $A_{P,\lambda}$ in a Siegel set. In any such limit, the distance diverges logarithmically in the multiplicative group coordinates on A_P.

As argued earlier in this section, the arithmeticity condition is fulfilled if consistency with the No Global Symmetries Conjecture is imposed.

6.2 Higher Supergravity - the Gauge Field Sector

We now turn to the gauge fields appearing in higher supergravity theories. Earlier, we focused on 1-form gauge fields A^a in a four-dimensional spacetime. In this chapter we will be more general and look at p-form gauge fields in a $D = 2(p+1)$-dimensional spacetime. The general Lagrangian describing the kinetic terms of such fields is given by

$$\mathscr{L} = -\frac{1}{2} f_{ab} F^a \wedge *F^b + \frac{1}{2} \tilde{f}_{ab} F^a \wedge F^b. \tag{6.32}$$

A remarkable feature of higher supergravity is that the geometry of the field space is so constrained that the gauge coupling functions are uniquely determined. To make sense of this statement, we require a more global perspective on the gauge coupling function, extending the local point of view taken in Chapter 5. In the local case, the f_{ab} and $\tilde{f}_{ab}$ were regarded as a set of real-valued functions, but in the global case we must specify in which space these functions take values. To do so, we follow the discussion in [25]. Note that the generalities presented here hold for any theory with abelian gauge fields described by the Lagrangian $\mathscr{L}$. Later in this section we specialize to higher supergravity again.

6.2.1 General Geometric Structure of Gauge Couplings

It is first convenient to switch to a basis of field strengths F^a and dual field strengths $G_a = *(\partial \mathscr{L} / \partial F^a)$ which is diagonal with respect to the Hodge star operator. In a $D = 2(p+1)$-dimensional spacetime with Minkowski signature, the Hodge star satisfies

$$*^2 = \begin{cases} +1, & p \text{ even}, \\ -1, & p \text{ odd}, \end{cases} \tag{6.33}$$

so it has eigenvalues ± 1 for p even and $\pm i$ for p odd. We therefore discuss these cases separately, starting with the even case.

Gauge couplings of p-form gauge fields - p even

Similar to Section 2.4, we decompose F^a and G_a into self-dual and anti-self-dual parts $F^{a\pm}$ and $G_a^\pm$, such that

$$F^a = F^{a+} + F^{a-}, \quad *F^{a\pm} = \pm F^{a\pm}, \tag{6.34}$$

and likewise for G_a. Calculating the variational derivative $\partial\mathscr{L}/\partial F^a$ and extracting the (anti)-self-dual parts, one finds

$$G_a^\pm = (-f_{ab} \pm \tilde{f}_{ab})F^{b\pm}. \tag{6.35}$$

We assemble the gauge coupling functions into a single object $\mathcal{N}_{ab}$ defined by

$$\mathcal{N}_{ab} = f_{ab} - \tilde{f}_{ab}. \tag{6.36}$$

Viewing the gauge couplings as matrices, we note that f_{ab} is always symmetric, and that $\tilde{f}_{ab}$ is anti-symmetric for p odd. We can therefore extract f_{ab} and $\tilde{f}_{ab}$ from $\mathcal{N}_{ab}$ by taking the symmetric and anti-symmetric parts:

$$f_{ab} = \frac{1}{2}(\mathcal{N}_{ab} + \mathcal{N}_{ba}), \quad \tilde{f}_{ab} = -\frac{1}{2}(\mathcal{N}_{ab} - \mathcal{N}_{ba}). \tag{6.37}$$

Denoting by $\mathcal{N}^\dagger_{ab} = f_{ab} + \tilde{f}_{ab}$ the matrix obtained from changing the sign of the anti-symmetric part of $\mathcal{N}_{ab}$, we have

$$G_a^+ = -\mathcal{N}_{ab}F^{a+}, \quad G_a^- = -\mathcal{N}^\dagger_{ab}F^{a-}. \tag{6.38}$$

Recall from the previous chapter that the gauge coupling functions depend on a choice of duality frame. If we collect the field strengths F^a and G_a into a vector

$$\mathcal{F}^\pm = \begin{pmatrix} F^{a\pm} \\ G_b^\pm \end{pmatrix} \tag{6.39}$$

consisting of $2n$ fields, then Equation (6.38) can be expressed in a 'covariant' way as

$$(1 - \Omega)\mathcal{E}\mathcal{F}^+ = 0, \tag{6.40}$$

with a similar equation holding for $\mathcal{F}^-$. Here Ω is the orthogonal matrix

$$\Omega = \begin{pmatrix} 0 & I_n \\ I_n & 0 \end{pmatrix}. \tag{6.41}$$

The matrix $\mathcal{E}$ is algebraically related to $\mathcal{N}$ via

$$\mathcal{E} = \begin{pmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{C} & \mathcal{D} \end{pmatrix}, \quad \mathcal{N} = (\mathcal{B} - \mathcal{D})(\mathcal{A} - \mathcal{C})^{-1}. \tag{6.42}$$

Though we are presently considering an arbitrary theory, for our later purposes we note that the relevance of $\mathcal{E}$ is that it plays a role in the formulation of the gauge couplings in higher supergravity. It can be shown that the matrix $\mathcal{E}$ must be an element of the group $\mathrm{O}(n, n)$, which is the group of matrices that preserves Ω. In fact, for p even, the group $\mathrm{O}(n, n)$ is the duality group for abelian p-form gauge fields in $D = 2(p + 1)$ [25].

From the above we see that every $\mathcal{E} \in \mathrm{O}(n,n)$ determines gauge coupling functions f_{ab} and $\tilde{f}_{ab}$ through the matrix $\mathcal{N}_{ab}$. However, different choices of $\mathcal{E}$ may lead to the same gauge couplings. The couplings are encoded in Equation (6.47), and we see that a transformation $\mathcal{E} \mapsto \tilde{\mathcal{E}} = \mathcal{S}\mathcal{E}$ preserves the structure of this equation if the matrix $\mathcal{S}$ commutes with Ω. The group of such $\mathcal{S}$ is the centralizer of Ω in $\mathrm{O}(n,n)$, which is the maximal compact subgroup $\mathrm{O}(n) \times \mathrm{O}(n)$. From this we infer the following observation [25].

Observation 6.2.1. Suppose for the moment that the field space $\mathcal{M}$ is simply connected. Then for even p-form abelian gauge fields in $D = 2(p+1)$ dimensions, there is a one-to-one correspondence

$$\{\text{gauge coupling functions } f_{ab} \text{ and } \tilde{f}_{ab}\} \xleftrightarrow{1:1} \{\text{maps } \mathcal{M} \to \mathrm{O}(n,n)/(\mathrm{O}(n)\times\mathrm{O}(n))\}.$$

The reason for assuming that $\mathcal{M}$ is simply connected is monodromy: as we have mentioned in the previous chapter, traversing a non-trivial loop in the field space may lead to a rotation of the duality frame in which the gauge couplings are formulated. This issue will be addressed in the next subsection.

Gauge couplings of p-form gauge fields - p odd

We now turn to the case that p is odd, following the same steps as in the even case. Since $*^2 = -1$ in this case, the field strengths are decomposed into imaginary self-dual and imaginary anti-self-dual parts, namely

$$F^a = F^{a+} + F^{a-}, \quad *F^{a\pm} = \pm i F^{a\pm}, \tag{6.43}$$

and similarly for G_a. This time, a calculation shows that the $G_a^\pm$ are related to the $F^{a\pm}$ via

$$G_a^\pm = (f_{ab} \pm i\tilde{f}_{ab})F^{b\pm}. \tag{6.44}$$

In analogy with the case that p is even, we combine the gauge couplings into a matrix

$$\mathcal{N}_{ab} = -\tilde{f}_{ab} + i f_{ab}. \tag{6.45}$$

In the present case where p is odd, both of the gauge couplings f_{ab} and $\tilde{f}_{ab}$ are symmetric, and they are obtained from $\mathcal{N}_{ab}$ as

$$f_{ab} = \mathrm{Im}\,\mathcal{N}_{ab}, \quad \tilde{f}_{ab} = -\mathrm{Re}\,\mathcal{N}_{ab}. \tag{6.46}$$

Again, we can encode the gauge couplings in the equation

$$(1 - i\Omega)\mathcal{E}\mathcal{F}^+ = 0, \tag{6.47}$$

where Ω is now defined by

$$\Omega = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}, \tag{6.48}$$

and $\mathcal{E}$ defines $\mathcal{N}$ via the algebraic relation

$$\mathcal{E} = \begin{pmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{C} & \mathcal{D} \end{pmatrix}, \quad \mathcal{N} = i(\mathcal{B} - i\mathcal{D})(\mathcal{A} - i\mathcal{C})^{-1}. \tag{6.49}$$

An analogous equation holds for $\mathcal{F}^-$. In this case, $\mathcal{E}$ is an element of the duality group $\mathrm{Sp}(2n,\mathbb{R})$. Note that we have already encountered this group in Chapter 5 as the duality group for $p = 1$. The centralizer subgroup of Ω is now the maximal compact subgroup $\mathrm{U}(n) \subseteq \mathrm{Sp}(2n,\mathbb{R})$, and we draw the following conclusion [25].

Observation 6.2.2. Assume that the field space $\mathcal{M}$ is simply connected. Then for odd p-form abelian gauge fields in $D = 2(p+1)$ dimensions, there is a one-to-one correspondence

$$\{\text{gauge coupling functions } f_{ab} \text{ and } \tilde{f}_{ab}\} \xleftrightarrow{\;1:1\;} \{\text{maps } \mathcal{M} \to \mathrm{Sp}(2n,\mathbb{R})/\mathrm{U}(n)\}.$$

6.2.2 Gauge Couplings as Morphisms of Arithmetic Quotients

We now return to the setting of higher supergravity. In this situation, the universal cover of the field space $\mathcal{M}$ is a symmetric space G/K. As a result of the previous subsection, the gauge couplings in higher supergravity are therefore associated to a map $\tilde{\mu}$ from G/K into the symmetric space

$$G^{\mathrm{d}}/K^{\mathrm{d}} = \begin{cases} \mathrm{O}(n,n)/(\mathrm{O}(n)\times\mathrm{O}(n)), & p \text{ even}, \\ \mathrm{Sp}(2n,\mathbb{R})/\mathrm{U}(n), & p \text{ odd}, \end{cases}$$

where G^{d} and K^{d} denote the duality group and its maximal compact subgroup, respectively. An elegant aspect of the higher supersymmetric setting is that this map originates from a representation $\rho : G \to G^{\mathrm{d}}$, and that in fact this representation is (up to isomorphism) uniquely determined [25]. The physical interpretation of this map is that the classical global symmetry group G has an action on the gauge fields, and that this action can only be given in terms of duality transformations in G^{d}, which is the largest possible symmetry group of the gauge kinetic terms. The 'lifted' gauge coupling map $\tilde{\mu}$ is related to ρ via the diagram

$$\begin{tikzcd} G \arrow[r, "\rho"] \arrow[d] & G^{\mathrm{d}} \arrow[d] \\ G/K \arrow[r, "\tilde{\mu}"] & G^{\mathrm{d}}/K^{\mathrm{d}}, \end{tikzcd} \tag{6.50}$$

which is well-defined because the image $\rho(K)$ must be contained in the maximal compact subgroup K^{d} by compactness. In view of the definability of such group quotients and algebraic maps between them, we thus already conclude:

Observation 6.2.3. The lifted gauge coupling map $\tilde{\mu}$ is $\mathbb{R}_{\mathrm{alg}}$-definable in higher supergravity.

An as example, consider the $\mathcal{N} = 4$ supergravity in four dimensions. According to Table 6.1, the group G is $\mathrm{SL}(2,\mathbb{R}) \times \mathrm{SO}(6,k)$, where $k \geq 6$ is the number of vector multiplets coupled to the theory, in addition to the six vectors appearing in the gravity multiplet as graviphotons [89]. The required representation

$$\mathrm{SL}(2,\mathbb{R}) \times \mathrm{SO}(6,k) \longrightarrow \mathrm{Sp}(2(6+k),\mathbb{R})$$

is then the bifundamental representation [93]. This is discussed more explicitly in the next section.

The universal cover $\widetilde{\mathcal{M}}$ of the field space is G/K, and the true field space is obtained by taking a quotient by Γ. In view of the arguments in Section 6.1.2, we assume that Γ is arithmetic. The arithmetic quotient $\Gamma\backslash G/K$ is not simply connected, so in order to define the gauge couplings we have to deal with monodromy. This is done in analogy with the period map: the monodromy transformations of the duality frame determine a representation of $\pi_1(\mathcal{M})$ into the duality group G^{d}. We can thus obtain a well-defined gauge coupling map if we take a quotient by the monodromy group. As with the period map, we take a quotient by the arithmetic group $G^{\mathrm{d}}_{\mathbb{Z}}$, which contains the monodromy group. In fact, this quotient has a physical interpretation in terms of Dirac quantization of the fluxes of the gauge fields [25]. The map μ from the field space $\mathcal{M}$ obtained by taking the monodromy quotient of $G^{\mathrm{d}}/K^{\mathrm{d}}$ was called the *intrinsic gauge coupling map* in [86].

In the setting of higher supergravity, we now find that the intrinsic gauge coupling is a map μ between two arithmetic quotients, related to $\tilde{\mu}$ by the diagram

$$\begin{array}{ccc} G/K & \xrightarrow{\tilde{\mu}} & G^{\mathrm{d}}/K^{\mathrm{d}} \\ \downarrow & & \downarrow \\ \Gamma\backslash G/K & \xrightarrow{\mu} & G^{\mathrm{d}}_{\mathbb{Z}}\backslash G^{\mathrm{d}}/K^{\mathrm{d}}. \end{array} \tag{6.51}$$

Since the map μ is induced by a map of algebraic groups $\rho : G \to G^{\mathrm{d}}$ and since $\rho(\Gamma) \subseteq G^{\mathrm{d}}_{\mathbb{Z}}$ because $G^{\mathrm{d}}_{\mathbb{Z}}$ contains the monodromy group, we obtain the following result:

> **Observation 6.2.4.** The intrinsic gauge coupling map μ in a higher supergravity theory is a morphism of arithmetic quotients, as defined in [2]. Therefore, μ is definable in $\mathbb{R}_{\mathrm{alg}}$.

Here we refer to the morphisms of arithmetic quotients mentioned in Chapter 4, whose definability is one of the non-trivial results in [2]. Note that the gauge coupling map determines (the class of) a matrix $\mathcal{E}$, from which the gauge coupling functions f_{ab} and $\widetilde{f}_{ab}$ can be extracted in an algebraic manner which preserves $\mathbb{R}_{\mathrm{alg}}$-definability, as explained above. We thus conclude that, regardless of whether we choose to work with the field space $\mathcal{M}$ or its universal cover $\widetilde{\mathcal{M}}$, the second part of the Tameness Conjecture holds for higher supergravity theories for which Γ is arithmetic.

6.3 Four-Dimensional $\mathcal{N} = 4$ Supergravity

We now proceed with a concrete example of higher supergravity to see some of the abstract constructions discussed above in action, namely the $\mathcal{N} = 4$ supergravity in four dimensions. From a string theory perspective, this theory arises as the low-energy description of Type IIB compactified on K3 $\times\, \mathbb{T}^2$ or a $\mathbb{T}^6/\mathbb{Z}_2$ orientifold [23]. The main aim of this section is to briefly exemplify the discussion of the previous sections, and not to give a detailed description of this theory.

6.3.1 The Scalar Sector

As already mentioned earlier, the cover of the scalar field space of this theory is the symmetric space

$$\widetilde{\mathcal{M}} = G/K = \mathrm{SL}(2,\mathbb{R})/\mathrm{SO}(2) \times \mathrm{SO}(6,k)/(\mathrm{SO}(6) \times \mathrm{SO}(k)).$$

The first factor $\mathrm{SL}(2,\mathbb{R})/\mathrm{SO}(2)$ is the upper half plane $\mathbb{H}$, which we have encountered many times in this thesis. The second factor is of the form $\mathrm{SO}(p,q)/\big(\mathrm{SO}(p) \times \mathrm{SO}(q)\big)$, for which we have constructed a Siegel set earlier in this chapter. Let us denote an element of $\widetilde{\mathcal{M}}$ by a pair

$$\mathcal{V} = \left(\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}, g \right),$$

with $\alpha\delta - \beta\gamma = 1$. The matrix $g \in \mathrm{SO}(6,k)$ is such that it preserves the bilinear pairing $\eta = \mathrm{diag}(-1,\dots,-1,1,\dots,1)$ of signature $(6,k)$, i.e. $g^{\mathrm{T}}\eta g = \eta$. Following the literature and the procedure above, we formulate the field space metric in terms of $\mathcal{V}\mathcal{V}^{\mathrm{T}}$. The metric on the factor $\mathrm{SL}(2,\mathbb{R})$ was already considered in Example 6.1.6, and is commonly expressed in terms of the matrix

$$M = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}^{\mathrm{T}} = \frac{1}{\mathrm{Im}\,\tau} \begin{pmatrix} |\tau|^2 & \mathrm{Re}\,\tau \\ \mathrm{Re}\,\tau & 1 \end{pmatrix}.$$

Together with the matrix $N = gg^{\mathrm{T}}$, the scalar kinetic terms specified by equation (6.20) take the form

$$-\frac{1}{4(\mathrm{Im}\,\tau)^2}\partial_\mu\tau\,\partial^\mu\bar{\tau} + \frac{1}{16}\,\mathrm{tr}\Big(\partial_\mu N_{ab}\,\partial^\mu N^{ab}\Big), \tag{6.52}$$

as given in e.g. [89]. Here we have restored the indices on N, and the upper indices indicate that an inverse is taken.

6.3.2 The Gauge Field Sector

As explained above, the starting point of the construction of the gauge couplings is a representation $G \to G^{\mathrm{d}}$. In this case, following [93], the required symplectic representation of G is given by

$$\begin{gathered} \rho : \mathrm{SL}(2,\mathbb{R}) \times \mathrm{SO}(6,k) \to \mathrm{Sp}(2(6+n),\mathbb{R}) \\ \left(\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}, g\right) \mapsto \begin{pmatrix} \alpha g & \beta\eta g^{-\mathrm{T}} \\ \gamma\eta g & \delta g^{-\mathrm{T}} \end{pmatrix}. \end{gathered}$$

The matrix on the right defines the matrix $\mathcal{E}^{-1}$ from the previous section. Note that this representation is obtained by combining the two commuting symplectic representations

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \mapsto \begin{pmatrix} \alpha & \beta\eta \\ \gamma\eta & \delta \end{pmatrix} \quad \text{and} \quad g \mapsto \begin{pmatrix} g & 0 \\ 0 & g^{-\mathrm{T}}. \end{pmatrix}$$

The $\mathrm{SL}(2,\mathbb{R})$ structure present in this embedding provides a convenient choice of duality frame [89]. Writing $\mathcal{F} = (\mathcal{F}^{a+}, \mathcal{F}_b-)$, where a and b are $\mathrm{SO}(6,k)$-indices and $\pm$ is an $\mathrm{SL}(2,\mathbb{R})$-index, we may take the field strengths to be $F^a = \mathcal{F}^{a+}$, with their duals given by $G_b = \mathcal{F}_b^-$.

For the scalar kinetic terms, we noted that it is often convenient to remove the local K-dependence by working with $\mathcal{V}\mathcal{V}^{\mathrm{T}}$ instead of $\mathcal{V}$. When this prescription is employed, the gauge couplings are computed via a slightly different algebraic procedure, explained in e.g. [88]. The resulting kinetic terms for the gauge fields are given by

$$-\frac{1}{4}\,\mathrm{Im}\,\tau\,N_{ab}\,F^a \wedge *F^b + \frac{1}{4}\,\mathrm{Re}\,\tau\,\eta_{ab}\,F^a \wedge F^b. \tag{6.53}$$

In the notation used above, we thus identify

$$f_{ab} = \frac{1}{2}\,\mathrm{Im}\,\tau\,N_{ab}, \quad \tilde{f}_{ab} = \frac{1}{2}\,\mathrm{Re}\,\tau\,\eta_{ab},$$

which shows the explicit dependence of the gauge coupling functions on the scalar fields. The tameness of these couplings follows from the algebraic group structure, as argued on general grounds in the previous section.

6.4 The Parameter Space and Gauged Supergravity

The third aspect of the Tameness Conjecture concerns the parameter space of effective field theories. In this final section we briefly discuss this aspect in the setting of higher supergravity. Throughout this chapter we have learned that the supergravity theory is determined by the geometry of the scalar field space $\mathcal{M} = \Gamma\backslash G/K$. We have only discussed this explicitly for the scalar and gauge field sectors, but this idea extends to other couplings for the elementary fields in the theory [25].

The first instance of a parameter space in this setting is therefore the (discrete) space of choices G, K, and Γ. As shown in Table 6.1, the group G is almost uniquely determined by specifying the amount of supersymmetry $\mathcal{N}$ and the spacetime dimension D; the only non-uniqueness comes from the possible presence of an integer parameter k counting the number of vector multiplets coupled to the theory. From a bottom-up perspective, nothing appears to forbid any value of k, but in string theory constructions there often exist bounds on the matter content and on the rank of the gauge group. Bounds of this type have been discussed in e.g. [94, 95] and form on ongoing area of research. It is therefore not inconceivable that similar bounds exist on k, though this is at present not known.

When the group G is specified, K is uniquely fixed to be the maximal compact subgroup of K, which is required to ensure that the quotient G/K is a symmetric space as required by Proposition 6.1.1. The final choice in determining the field space then comes from the subgroup Γ. We have used the consistency with some of the swampland conjectures to argue that Γ must be an arithmetic subgroup. This argument assumed a weaker form of the No Global Symmetries Conjecture, which permits a global symmetry group of finite order. If we assume the original form of this conjecture in which global symmetries of finite order are also forbidden, we arrive at the conclusion that Γ must be equal to the U-duality group $\mathcal{U}$, which by invoking the Completeness Hypothesis was argued to be $\mathcal{U} = G_{\mathbb{Z}}$. We combine these ideas into the following conclusion.

> **Observation 6.4.1.** Suppose that string theory provides an upper bound on the rank of the group G. Then, assuming consistency with the No Global Symmetries Conjecture and the Completeness Hypothesis, there are finitely many choices of G, K and Γ for the field space $\mathcal{M} = \Gamma\backslash G/K$ of a higher supergravity theory.

Let us emphasize that this is only a conditional conclusion, assuming a hypothetical bound on the rank of the symmetry group G.

So far we have, albeit implicitly, focused on a setting called *ungauged* higher supergravity. The group G played the role of a classical global symmetry group, and whenever such a group is present one may consider gauging a subgroup G_0 of G. In other words, a choice of subgroup G_0 may be promoted to a local symmetry, and upon doing so the theory becomes a *gauged* higher supergravity theory [88]. The gauging procedure introduces several new couplings into the theory, including a scalar potential[6].

[6]Note that a non-trivial scalar potential is forbidden in ungauged higher supergravity, since the G-invariance of such a theory implies that a scalar potential must be given in terms of a G-invariant map $G \to \mathbb{R}$, i.e. a constant.

From a top-down perspective, the relevance of gauged supergravities comes from the fact that a gauging may be generated from the presence of various physical objects in the compactification, such as branes and fluxes. Thus, the four-dimensional higher supersymmetric theories arising in string theory constructions are typically given by gauged higher supergravity theories.

The possibility of gauging a subgroup of G presents a new spectrum of choices for the parameter space of higher supersymmetric theories, namely the choices of consistent gaugings. A useful procedure for encoding these choices is the *embedding tensor formalism*, reviewed in e.g. [88]. The idea behind this formalism is to encode the embedding of the gauged subgroup $G_0 \subseteq G$ into a tensorial object called the embedding tensor. Consistency of the gauging, for example the closure of the Lie algebra $\mathfrak{g}_0 \subseteq \mathfrak{g}$, places strong constraints on this tensor, in the form of a number of consistency equations. For instance, in the case of gauged $\mathcal{N} = 4$ supergravity in four and five dimensions, a complete list of constraints was found in [89].

The solutions set of these constraints thus constitutes a part of the parameter space of higher supergravity theories, and investigating the definability of this space is left for further work. At this point, let us already note that the top-down perspective on the gauging procedure provides a promising outlook on definability of this solution set, since it is closely related to the parameter space for the compactification, which includes for instance the locus of flux vacua.

Chapter 7

Discussion and Outlook

7.1 Summary and Discussion

In this thesis we explored the recent progress in physics and mathematics centered around the idea of tameness. This tameness was made precise by tame geometry, built on the concept of an o-minimal structure. We began by reviewing the two areas of mathematics and physics in which many of this progress has taken place, starting with Chapter 1 on Hodge theory. The essence of the chapter lied in the construction of a period map and a discussion of its properties. We then moved on to physics, reviewing the basic concepts of string theory in Chapter 2, emphasizing the idea that string theory compactifications lead to an enormous landscape of effective theories, lying inside an even more enormous swampland of theories inconsistent with quantum gravity. We then covered the basics of tame geometry and o-minimal structures in Chapter 3, where we saw that the o-minimality condition led to a host of tameness theorems for definable spaces and maps.

In the second part of the thesis we explored the connections between the chapters of the first part. Chapter 4 on tameness in Hodge theory culminated in a review of the proof of the tameness of the period map given in [2], and along the way we investigated the geometry of arithmetic quotients. The idea that tameness could be a general principle in physics, made precise by the Tameness Conjecture formulated in [1], formed the center of discussion in Chapter 5. One of the motivations of this conjecture originates from the recent finding that the self-dual locus of a variation of Hodge structures is definable [4], and we reviewed the proof of this result. In Chapter 6, the last chapter, we strengthened the evidence for the Tameness Conjecture by showing how tameness arises in higher supergravity theories. We focused on the couplings for the scalar and gauge field sectors, and argued on general grounds that these couplings are tame as a consequence of the algebraic structure imposed by the symmetry group G. Relying on results from [2] on the geometry of arithmetic quotients, we argued that tameness persists even after taking quotients by infinite discrete duality subgroups, and in doing so we noted various possible connections with other swampland conjectures. The investigation of tameness in higher supergravity is not complete, and some aspects are left for future research.

7.2 Outlook on Further Research

The theme of tame geometry opens up many new avenues for further work. In this concluding section we provide a number of ideas and directions to explore for future research. Many of these ideas are inspired by insightful discussions with the people thanked in the acknowledgements. To structure the outlook we separate it into directions for physics and for mathematics.

Outlook for Physics

The first direction for further work which naturally presents itself is the extension of the investigation of the higher supersymmetric setting in Chapter 6. An outlook on some aspects was already given at the end of that chapter, but let us reiterate the main points. For the tameness of the coupling functions, we have focused on two of the most essential sectors in the theory, namely the scalars and the abelian gauge fields. These are not the only fields in higher supergravity, and it would be interesting to explore the tameness of the couplings of other fields as well. As a consequence of the highly constrained algebraic structure of these theories, the Tameness Conjecture is likely to hold for these other fields. In fact, it might be possible to argue for tameness purely on the grounds of the invariance of the theory under the classical global symmetry group G. For instance, the definability of the metric on the field space was proven by using its G-invariance.

In Chapter 6 we have restricted ourselves to ungauged supergravity. Generalizations of this setting may be obtained by gauging a subgroup G_0 of G, i.e. promoting a subgroup of G to a local symmetry. From a stringy perspective, this gauging is effected by the presence of branes and fluxes [88], and the ubiquity of these objects in compactifications makes it important to understand the gauged higher supergravity as well. For example, G-invariance of the scalars forbids a scalar potential, but when a subgroup of G is gauged, a scalar potential is generated. It must be shown that this scalar potential is tame in order for the Tameness Conjecture to be valid.

An important aspect of the Tameness Conjecture for which we have not yet found conclusive evidence in this work in this work is the definability of the parameter space. In higher supergravity, there are several of these parameters. For instance, as seen in Table 6.1, several higher supergravities have an integer k counting the number of vector multiplets. The discreteness of this parameter implies that, according to the Tameness Conjecture, only finitely many values of k should be allowed for theories consistent with quantum gravity. This is closely related to bounds on the rank of the gauge group of an effective field theory, which has been discussed in e.g. [94, 95]. Another parameter comes from the possible gaugings of the theory. One way in which the gauging can be parametrized is the *embedding tensor* [88], which describes the embedding of the gauged subgroup G_0 in G. Consistency of the theory places several constraints on the embedding tensor, and to validate the Tameness Conjecture one must show that the solution set of these constraints is definable. Due to the relation between fluxes and gaugings, this solution set may be closely related to the locus of flux vacua.

In general, the proposal of the Tameness Conjecture presents many new ideas to be explored in further work. Among these ideas we identify three main directions:

(i) Find more evidence for the Tameness Conjecture;

(ii) Investigate phenomenological implications of the Tameness Conjecture;

(iii) Explore connections between the Tameness Conjecture and other swampland conjectures.

For (i), the top-down evidence presented in [1, 80] and this thesis could potentially be supplemented by evidence from other string theory settings, such as Type I, Type IIA, or Heterotic. As outlined in Section 5.3.2, an intriguing direction to explore in future work is the search for bottom-up arguments based on finiteness principles or black hole phenomenology. Such arguments could perhaps, for instance, be formulated along the lines of the evidence given for the Distance Conjecture in [77]. Motivation for tameness from black hole physics would elucidate the connection between the Tameness Conjecture and gravity, which is mysterious at present.

For the second point, it would be interesting to explore how tameness of an effective field theory affects the physics of such a theory. Some phenomenological consequences of tameness were noted in [80] in asymptotic regions of the field space. In an upcoming paper, it is argued that in a definable effective field theory, the amplitudes of scattering processes given in terms of Feynman graphs are definable [96]. This shows that tameness of effective field theories is preserved at the quantum level, at least at finite order in perturbation theory.

The connection between the Tameness Conjecture and other swampland conjectures appears to be a promising direction for future research. As seen in Chapter 2, the swampland conjectures form a web of interrelated ideas, and it would be interesting to investigate the position of the Tameness Conjecture inside this web. The first step towards this were taken in [80], in which the Tameness Conjecture was connected to the Distance Conjecture. Here the tameness of coupling functions, combined with the tameness theorems discussed in Chapter 2, was used to shed light on the path-dependence of infinite distance limits relevant to the Distance Conjecture. In this thesis we noted a connection between the tameness in higher supergravity theories and the No Global Symmetries Conjecture and the Completeness Hypothesis through the arithmetic aspects of the field space geometry. However, the necessity of the arithmeticity assumption for tameness remains dubious, and it would be worthwhile to explore this in more detail in future work. Other swampland conjectures whose relation to the Tameness Conjecture is currently being explored include the Weak Gravity Conjecture and the de Sitter Conjecture.

Another interesting direction to study, which is not directly related to the Tameness Conjecture, is the distribution of flux vacua. The definability of the locus of flux vacua proven in [4] may provide new insights into how these vacua are distributed inside the Hodge bundle, through various results on the distribution of lattice points in definable families [97].

Outlook for Mathematics

The recent success of tame geometry in driving new mathematical developments gives promising prospects for the future. Among these, the most promising direction appears to be algebraic geometry and number theory, where much progress was already made on several conjectures [3, 56, 57].

The direction for future research that lies closest to what has been discussed in this thesis is the union of tame geometry and differential geometry. In Section 3.6 we found that these types of geometry interact nicely, but this setting seems to be relatively unexplored in the literature. At this point, it therefore seems relatively unclear whether imposing o-minimality as an additional axiom could lead to any deep new results in differential geometry. A promising starting point would be to extend the various tameness theorems discussed in Chapter 3 to the setting of definable smooth manifolds. As mentioned in Chapter 3, these tameness theorems are known to generalize to definable spaces when imposing a mild topological regularity condition which allows the space to be definably embedded in Euclidean space. It would, however, be nice to be able to formulate these theorems without reference to an embedding.

Besides the tameness theorems discussed in Chapter 3, it would also be interesting to attempt to generalize other non-trivial results of tame geometry in Euclidean space to definable manifolds. A concrete example which relates to the outlook for physics is the generalization of the Łojasiewicz inequality for definable functions to definable manifolds.

A different topic which appears to be worthwhile to investigate is related to the arithmetic quotients $\Gamma\backslash G/K$ encountered many times in this thesis. We have seen in Chapter 4 that if the group Γ is arithmetic, the construction of Siegel sets is available, which allows one to prove that arithmetic quotients are definable. Arithmetic groups are rather large discrete groups, since they have finite index in $G_{\mathbb{Z}}$. It is currently unclear how this story changes when one takes a quotient by a smaller non-arithmetic group Γ. Intuitively, a quotient by a smaller discrete group should be tamer, and it is a natural expectation that such 'non-arithmetic quotients' are definable in an o-minimal structure as well. It would be interesting to investigate the possibility to extend the construction of Siegel sets to prove or disprove whether this is indeed true. Doing so would also shed light on the possible connection between the Tameness Conjecture and the No Global Symmetries Conjecture through arithmeticity.

Let us speculate on one final idea for further research, concerning the definability of the locus of self-dual classes. This locus is a generalization of the locus of Hodge classes, and according to the Hodge conjecture such classes have a geometric interpretation in terms of algebraic subvarieties. It would be interesting, though ambitious, to see whether it is possible to find a (possibly conjectural) geometric interpretation of self-dual classes, perhaps in terms of definable subvarieties. The algebraicity of Hodge loci gives evidence for the Hodge conjecture, and the definability of self-dual loci could serve as starting point for the search for such a geometric interpretation.

Bibliography

[1] T. W. Grimm, *Taming the Landscape of Effective Theories*, 2112.08383.

[2] B. Bakker, B. Klingler and J. Tsimerman, *Tame topology of arithmetic quotients and algebraicity of Hodge loci*, *J. Amer. Math. Soc.* **33** (2020) 917–939.

[3] B. Bakker and J. Tsimerman, *The Ax-Schanuel conjecture for variations of Hodge structures*, *Invent. Math.* **217** (2019) 77–94.

[4] B. Bakker, T. W. Grimm, C. Schnell and J. Tsimerman, *Finiteness for self-dual classes in integral variations of Hodge structure*, 2112.06995.

[5] B. S. Acharya and M. R. Douglas, *A Finite landscape?*, hep-th/0606212.

[6] M. R. Douglas, *The Statistics of string / M theory vacua*, *JHEP* **05** (2003) 046, [hep-th/0303194].

[7] C. Vafa, *The String landscape and the swampland*, hep-th/0509212.

[8] E. Palti, *The Swampland: Introduction and Review*, *Fortsch. Phys.* **67** (2019) 1900037, [1903.06239].

[9] M. van Beest, J. Calderón-Infante, D. Mirfendereski and I. Valenzuela, *Lectures on the Swampland Program in String Compactifications*, 2102.01111.

[10] C. Voisin, *Hodge theory and complex algebraic geometry. I*, vol. 76 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, English ed., 2007.

[11] D. Huybrechts, *Complex geometry*. Universitext. Springer-Verlag, Berlin, 2005.

[12] D. Arapura, *Algebraic geometry over the complex numbers*. Universitext. Springer, New York, 2012, 10.1007/978-1-4614-1809-2.

[13] J. Carlson, S. Müller-Stach and C. Peters, *Period mappings and period domains*, vol. 168 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2017.

[14] P. A. Griffiths, *Periods of Integrals on Algebraic Manifolds, I. (Construction and Properties of the Modular Varieties)*, *American Journal of Mathematics* **90** (1968) 568–626.

[15] W. Schmid, *Variation of Hodge structure: the singularities of the period mapping*, *Invent. Math.* **22** (1973) 211–319.

[16] P. Griffiths and J. Harris, *Principles of algebraic geometry.* Wiley Classics Library. John Wiley & Sons, Inc., New York, 1994, 10.1002/9781118032527.

[17] G. R. Cavalcanti, *Complex Geometry - Holomorphic vector bundles, elliptic operators and Hodge theory*, Lecture notes for the course Complex Geometry (Utrecht) (2022) .

[18] J. Frésan, *Hodge theory and o-minimality*, Notes from the Bourbaki Seminar (Paris) (2020) .

[19] H. Hironaka, *Resolution of singularities of an algebraic variety over a field of characteristic zero. I, II*, *Ann. of Math. (2)* **79** *(1964), 109–203; ibid. (2)* **79** (1964) 205–326.

[20] E. Cattani, A. Kaplan and W. Schmid, *Degeneration of Hodge structures*, *Ann. of Math. (2)* **123** (1986) 457–535.

[21] E. Cattani and A. Kaplan, *Degenerating variations of Hodge structure*, No. 179-180, pp. 9, 67–96. 1989.

[22] E. Cattani, P. Deligne and A. Kaplan, *On the locus of Hodge classes*, *J. Amer. Math. Soc.* **8** (1995) 483–506.

[23] R. Blumenhagen, D. Lüst and S. Theisen, *Basic concepts of string theory.* Theoretical and Mathematical Physics. Springer, Heidelberg, Germany, 2013, 10.1007/978-3-642-29497-6.

[24] D. Z. Freedman and A. Van Proeyen, *Supergravity.* Cambridge Univ. Press, Cambridge, UK, 5, 2012.

[25] S. Cecotti, *Supersymmetric Field Theories: Geometric Structures and Dualities.* Cambridge University Press, 1, 2015.

[26] H. Murayama, *Supersymmetry phenomenology*, in *ICTP Summer School in Particle Physics*, pp. 296–335, 2, 2000. `hep-ph/0002232`.

[27] C. Vafa, *Evidence for F theory*, *Nucl. Phys. B* **469** (1996) 403–418, [`hep-th/9602022`].

[28] T. Weigand, *F-theory*, *PoS* **TASI2017** (2018) 016, [`1806.01854`].

[29] F. Denef, *Les Houches Lectures on Constructing String Vacua*, *Les Houches* **87** (2008) 483–610, [`0803.1194`].

[30] P. Candelas and X. de la Ossa, *Moduli Space of Calabi-Yau Manifolds*, *Nucl. Phys. B* **355** (1991) 455–481.

[31] G. Tian, *Smoothness of the universal deformation space of compact Calabi-Yau manifolds and its Petersson-Weil metric*, in *Mathematical aspects of string theory (San Diego, Calif., 1986)*, vol. 1 of *Adv. Ser. Math. Phys.*, pp. 629–646. World Sci. Publishing, Singapore, 1987.

[32] A. N. Todorov, *The Weil-Petersson geometry of the moduli space of* $\mathrm{SU}(n \geq 3)$ *(Calabi-Yau) manifolds. I*, *Comm. Math. Phys.* **126** (1989) 325–346.

[33] K. Hori and C. Vafa, *Mirror symmetry*, `hep-th/0002222`.

[34] Z. Lu and X. Sun, *Weil-Petersson geometry on moduli space of polarized Calabi-Yau manifolds*, *J. Inst. Math. Jussieu* **3** (2004) 185–229.

[35] E. Viehweg, *Quasi-projective moduli for polarized manifolds*, vol. 30 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1995, 10.1007/978-3-642-79745-3.

[36] R. Blumenhagen, B. Kors, D. Lust and S. Stieberger, *Four-dimensional String Compactifications with D-Branes, Orientifolds and Fluxes*, *Phys. Rept.* **445** (2007) 1–193, [hep-th/0610327].

[37] S. Gukov, C. Vafa and E. Witten, *CFT's from Calabi-Yau four folds*, *Nucl. Phys. B* **584** (2000) 69–108, [hep-th/9906070].

[38] T. W. Grimm, *Moduli space holography and the finiteness of flux vacua*, *JHEP* **10** (2021) 153, [2010.15838].

[39] E. Witten, *On flux quantization in M theory and the effective action*, *J. Geom. Phys.* **22** (1997) 1–13, [hep-th/9609122].

[40] I. Bena, J. Blåbäck, M. Graña and S. Lüst, *The tadpole problem*, *JHEP* **11** (2021) 223, [2010.10519].

[41] K. Dasgupta, G. Rajesh and S. Sethi, *M theory, orientifolds and G - flux*, *JHEP* **08** (1999) 023, [hep-th/9908088].

[42] M. Graña, T. W. Grimm, D. van de Heisteeg, A. Herraez and E. Plauschinn, *The Tadpole Conjecture in Asymptotic Limits*, 2204.05331.

[43] J. Cheeger, *Finiteness theorems for Riemannian manifolds*, *Amer. J. Math.* **92** (1970) 61–74.

[44] D. Skinner, *Quantum Field Theory II*, Lecture notes (Cambridge) (2017) .

[45] T. W. Grimm, *The N=1 effective action of F-theory compactifications*, *Nucl. Phys. B* **845** (2011) 48–92, [1008.4133].

[46] T. W. Grimm, *The Effective action of type II Calabi-Yau orientifolds*, *Fortsch. Phys.* **53** (2005) 1179–1271, [hep-th/0507153].

[47] J. McNamara and C. Vafa, *Cobordism Classes and the Swampland*, 1909.10355.

[48] M. Montero and C. Vafa, *Cobordism Conjecture, Anomalies, and the String Lamppost Principle*, *JHEP* **01** (2021) 063, [2008.11729].

[49] H. Ooguri and C. Vafa, *On the Geometry of the String Landscape and the Swampland*, *Nucl. Phys. B* **766** (2007) 21–33, [hep-th/0605264].

[50] T. W. Grimm, E. Palti and I. Valenzuela, *Infinite Distances in Field Space and Massless Towers of States*, *JHEP* **08** (2018) 143, [1802.08264].

[51] T. W. Grimm, C. Li and E. Palti, *Infinite Distance Networks in Field Space and Charge Orbits*, *JHEP* **03** (2019) 016, [1811.02571].

[52] T. W. Grimm, C. Li and I. Valenzuela, *Asymptotic Flux Compactifications and the Swampland*, *JHEP* **06** (2020) 009, [1910.09549].

[53] B. Bastian, T. W. Grimm and D. van de Heisteeg, *Weak gravity bounds in asymptotic string compactifications*, *JHEP* **06** (2021) 162, [2011.08854].

[54] A. Grothendieck, *Esquisse d'un programme*, in *Geometric Galois actions, 1*, vol. 242 of *London Math. Soc. Lecture Note Ser.*, pp. 5–48. Cambridge Univ. Press, Cambridge, 1997.

[55] J.-P. Serre, *Géométrie algébrique et géométrie analytique*, *Ann. Inst. Fourier (Grenoble)* **6** (1955/56) 1–42.

[56] F. Oort and J. Tsimerman, *The André-Oort conjecture*, in *Open problems in arithmetic algebraic geometry*, vol. 46 of *Adv. Lect. Math. (ALM)*, pp. 61–70. Int. Press, Somerville, MA, 2019.

[57] B. Bakker, Y. Brunebarbe and J. Tsimerman, *O-minimal GAGA and a conjecture of Griffiths*, 2018. 10.48550/ARXIV.1811.12230.

[58] L. van den Dries, *Tame topology and o-minimal structures*, vol. 248 of *London Mathematical Society Lecture Note Series.* Cambridge University Press, Cambridge, 1998, 10.1017/CBO9780511525919.

[59] B. Bakker, *Hodge theory and o-minimality*, Notes from the Felix Klein lecture series (Bonn) (2019) .

[60] A. Tarski, *A decision method for elementary algebra and geometry.* University of California Press, Berkeley-Los Angeles, Calif., 1951.

[61] M. Coste, *An introduction to o-minimal geometry*, Lecture notes (Pisa) (2002) .

[62] A. J. Wilkie, *On the theory of the real exponential field*, *Illinois Journal of Mathematics* **33** (1989) 384 – 408.

[63] A. M. Gabrielov, *Projections of semi-analytic sets*, *Functional Analysis and its applications* **2** (1968) 282–291.

[64] L. van den Dries and C. Miller, *On the real exponential field with restricted analytic functions*, *Israel Journal of Mathematics* **92** (Feb., 1995) 427.

[65] B. Bakker, Y. Brunebarbe, B. Klingler and J. Tsimerman, *Definability of mixed period maps*, 2020. 10.48550/ARXIV.2006.12403.

[66] L. Ji, *Lectures on locally symmetric spaces and arithmetic groups*, in *Lie groups and automorphic forms*, vol. 37 of *AMS/IP Stud. Adv. Math.*, pp. 87–146. Amer. Math. Soc., Providence, RI, 2006.

[67] Y. Peterzil and S. Starchenko, *Complex analytic geometry in a nonstandard setting*, in *Model theory with applications to algebra and analysis. Vol. 1*, vol. 349 of *London Math. Soc. Lecture Note Ser.*, pp. 117–165. Cambridge Univ. Press, Cambridge, 2008. DOI.

[68] R. Bianconi and R. Figueiredo, *O-minimal de Rham cohomology*, 2019. 10.48550/arxiv.1904.05485.

[69] A. Borel, *Introduction to arithmetic groups*, vol. 73 of *University Lecture Series.* American Mathematical Society, Providence, RI, 2019, 10.1090/ulect/073.

[70] A. Borel and L. Ji, *Compactifications of symmetric and locally symmetric spaces.* Mathematics: Theory & Applications. Birkhäuser Boston, Inc., Boston, MA, 2006.

[71] D. Prasad and S. M. Garge, *Lectures on Algebraic Groups*, Notes from Lectures at the Harish-Chandra Research Institute (Allahabad) (2003) .

[72] A. Borel and J.-P. Serre, *Corners and arithmetic groups*, *Comment. Math. Helv.* **48** (1973) 436–491.

[73] C. Soulé, *An introduction to arithmetic groups*, in *Frontiers in number theory, physics, and geometry. II*, pp. 247–276. Springer, Berlin, 2007. DOI.

[74] M. Kashiwara, *The asymptotic behavior of a variation of polarized Hodge structure*, *Publ. Res. Inst. Math. Sci.* **21** (1985) 853–875.

[75] C. Schnell, *A new finiteness theorem for variations of Hodge structure*, Talk at Zoom Algebraic Geometry seminar (2021) .

[76] S. Ashok and M. R. Douglas, *Counting flux vacua*, *JHEP* **01** (2004) 060, [hep-th/0307049].

[77] Y. Hamada, M. Montero, C. Vafa and I. Valenzuela, *Finiteness and the Swampland*, 2111.00015.

[78] B. Bakker and S. Mullane, *Definable structures on flat bundles*, 2022. 10.48550/ARXIV.2201.02144.

[79] M. Kneser, *Quadratische Formen.* Springer-Verlag, Berlin, 2002, 10.1007/978-3-642-56380-5.

[80] T. W. Grimm, S. Lanza and C. Li, *Tameness, Strings, and the Distance Conjecture*, 2206.00697.

[81] Y. Tachikawa, *Undecidable problems in quantum field theory*, 2203.16689.

[82] T. Banks and N. Seiberg, *Symmetries and Strings in Field Theory and Gravity*, *Phys. Rev. D* **83** (2011) 084019, [1011.5120].

[83] N. Arkani-Hamed, L. Motl, A. Nicolis and C. Vafa, *The String landscape, black holes and gravity as the weakest force*, *JHEP* **06** (2007) 060, [hep-th/0601001].

[84] M. Trigiante, *Gauged Supergravities*, *Phys. Rept.* **680** (2017) 1–175, [1609.09745].

[85] C. M. Hull and P. K. Townsend, *Unity of superstring dualities*, *Nucl. Phys. B* **438** (1995) 109–137, [hep-th/9410167].

[86] S. Cecotti, *Swampland geometry and the gauge couplings*, *JHEP* **09** (2021) 136, [2102.03205].

[87] D. W. Morris, *Introduction to arithmetic groups*, 2001. 10.48550/arxiv.math/0106063.

[88] H. Samtleben, *Lectures on Gauged Supergravity and Flux Compactifications*, *Class. Quant. Grav.* **25** (2008) 214002, [0808.4076].

[89] J. Schon and M. Weidner, *Gauged $\mathcal{N} = \mathbf{4}$ supergravities*, *JHEP* **05** (2006) 034, [hep-th/0602024].

[90] A. Čap and J. Slovák, *Parabolic geometries. I*, vol. 154 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2009, 10.1090/surv/154.

[91] J. M. Lee, *Introduction to smooth manifolds*, vol. 218 of *Graduate Texts in Mathematics*. Springer, New York, second ed., 2013.

[92] A. Hatcher, *Algebraic topology*. Cambridge University Press, Cambridge, 2002.

[93] R. D'Auria, S. Ferrara, F. Gargiulo, M. Trigiante and S. Vaula, *$\mathcal{N} = \mathbf{4}$ supergravity Lagrangian for Type IIB on T^6/Z_2 in presence of fluxes and D3-branes*, *JHEP* **06** (2003) 045, [hep-th/0303049].

[94] H.-C. Kim, H.-C. Tarazi and C. Vafa, *Four-dimensional $\mathcal{N} = \mathbf{4}$ SYM theory and the swampland*, *Phys. Rev. D* **102** (2020) 026003, [1912.06144].

[95] H.-C. Tarazi and C. Vafa, *On The Finiteness of 6d Supergravity Landscape*, 2106.10839.

[96] T. W. Grimm, L. Schlechter and M. R. Douglas, *To appear* (2022), .

[97] F. Barroero and M. Widmer, *Counting lattice points and o-minimal structures*, *Int. Math. Res. Not. IMRN* (2014) 4932–4957.