Imperial College London

Machine Learning Calabi-Yau Manifolds

MSc Dissertation

Imperial College London

Department of Physics

Author:

Alex Wilkinson

Supervisor:

Prof. Dan Waldram

Submitted in partial fulfilment of the requirements for the degree of Master of Science of Imperial College London

September 25, 2020

Abstract

The applicability of machine learning techniques to the landscape of Calabi-Yau manifolds is examined. A review of the necessary mathematics is undertaken followed by a review of the Calabi-Yau manifold and its role in physics. The complete intersection Calabi-Yau manifold is explained and a machine learning approach is motivated. A working introduction to machine learning, with a focus on neural networks, is provided. Neural networks are then employed to learn the Hodge numbers of complex dimension 4 complete intersection Calabi-Yau manifolds. After training the networks on 10% of the data, they are able to predict $h^{1,1}$, $h^{2,1}$, $h^{3,1}$, and $h^{2,2}$ correctly with accuracies 0.82, 0.63, 0.16, and 0.04 respectively. The predicted distributions of Hodge numbers are also analysed and training with different fractions of the data is tested. The results suggest that machine learning may be applicable to model building where it can be utilised to identify promising regions in the landscape of string theory vacua.

Acknowledgements

I would like to thank my supervisor, Professor Dan Waldram, for his guidance and valuable discussions throughout the summer. I would also like to thank Professor Yang-Hui He for his helpful correspondence and being the inspiration for this project.

I am of course grateful to my parents and family for their support in all of my endeavours and extend a special thank you to Jade for always being there for me.

Contents

Introduction

String theory grew from a theory of the strong interaction in the 1960s [53] to the leading candidate for unification of the geometric ideas of general relativity with the quantum field theory of the Standard Model. The first formulation with a string as the fundamental object was bosonic string theory. This requires a 26-dimensional ($D = 26$) space for consistency and is plagued with a tachyon and no fermionic content. The superstring revolution in the 1980s led to five theories of a supersymmetric string: type I, type IIA, type IIB, and heterotic $E_8 \times E_8$ and $SO(32)$. They incorporate fermions and, as dictated by anomaly cancellation, are consistent in $D = 10$ spacetime. These superstring theories were all later found to be limits of the same 11-dimensional theory called M-theory [31]. Another notable development in string theory has been F-theory [51]. This offers a 12-dimensional description of type IIB string theory that gives a geometric reason for some of the peculiarities present in type IIB.

To make contact with the physical world we must address the extra 6 dimensions of the superstring theories. Doing this would permit string theory to give a phenomenologically realistic and mathematically consistent description of nature. This is typically done by supposing that the 10 dimensions of spacetime is a product of a familiar 4 dimensional manifold and some compact 6 dimensional manifold. In the spirit of Kaluza-Klein theory, the dimensions associated with the 6-manifold are made small in a process called compactification. The resulting compactified superstring theory can then be consistent with experimental evidence for $D = 4$. In order for it to also produce a supersymmetric gauge theory with a realistic particle spectrum, the 6-manifold must be a Calabi-Yau manifold. Through this realisation, the study of Calabi-Yau manifolds became an integral part of model building in string theory.

A Calabi-Yau manifold is defined as a Ricci-flat Kähler manifold. For now, this can

just be thought of as a complex manifold obeying some specific conditions. We will often refer to a Calabi-Yau manifold of n complex dimensions as a Calabi-Yau nfold or simply as CY_n. The threefold allows superstring theories to make contact with experimental truths. The fourfold is also of interest as a certain subset of them can be used in F-theory compactifications in a similar way to how the threefold is used. The study and cataloguing of the three and fourfold is of great importance as the topological properties of the compactified manifold dictates certain properties of the resulting $D = 4$ theory. We will refer to the collection of topologically distinct Calabi-Yau manifolds as the Calabi-Yau landscape. It forms part of the more general landscape of string theory vacua. This is the collection of all $D = 4$ Lorentz symmetric effective field theories that can be found by compactifications of string theories. In this way, string theory offers to the theoretical physicist the momentous task of understanding and selecting from the landscape of string theory vacua in return for the quantisation of gravity. The search for a vacuum that reproduces the particle physics and cosmology of our universe is known as the vacuum selection problem and is yet to be solved. A complete classification of the Calabi-Yau landscape would aide in the search for realistic vacua by allowing favourable Calabi-Yau manifolds to be identified.

It is interesting to view the challenge of understanding the Calabi-Yau landscape as a big data problem. Viewed through this lens, a machine learning approach is compelling. Taking a machine learning approach means using computer algorithms that will teach themselves to learn complex patterns in data without needing any specific programming. The Calabi-Yau landscape is primed for the implementation of these techniques as decades of study means there exists large data sets of inequivalent manifolds. The use of machine learning in the study of string vacua has recently seen substantial development with many techniques and algorithms being applied. To name a few, artificial neural networks (NNs) have been applied to the Hodge numbers of the Calabi-Yau threefolds [9], genetic algorithms used in the search for phenomenologically viable string vacua [1], and linear regression has been used to generate a previously proven conjecture regarding the gauge group rank in F-theory compactifications [16]. Much of this work is exploratory as the role of machine learning in the vacuum selection problem and string theory in general is yet to be properly established. It is certainly an avenue worth exploring as learning

landscape data may reveal unseen patterns and help to generate new data.

This work focuses on one aspect of the wider aim of developing machine learning techniques in string theory: using NNs to learn complete intersection Calabi-Yau manifolds (CICYs). The first part of this work is to understand what a CICY and a NN is. Understanding the CICY will take up a large portion of the dissertation as we must first build foundations in complex geometry and then define the Calabi-Yau manifold. The second part of this work will be a study of how NNs can be applied to learn topological data of CICY fourfolds (CICY4s). We will attempt to construct NNs that can predict the Hodge numbers (topological invariants) of unseen CICY4s, as opposed to seen CICY4s which are used to teach the NNs patterns in the data, and document the results. This has not yet been done in the literature[1] but is a natural extension to the work in [9, 28]. The study does not claim to be definitive but is undertaken to explore the applicability of machine learning to CICY4s and to give an original example of how it can contribute to string theory in general.

In § 1, the reader is made familiar with the basic concepts from complex geometry needed to understand the Calabi-Yau manifold and the topological invariants we will encounter in later sections. This section is structured around taking the reader from real to complex manifolds, and then to Kähler manifolds. Detours are made into holonomy, Hodge theory, holomorphic vector bundles and Chern classes as these are all necessary tools for later sections. The complex projective space ($\mathbb{CP}^N$) is frequently used as an example for various concepts as it is the space CICYs are built in.

With the review of complex geometry complete, § 2 is on the Calabi-Yau manifold. Four equivalent conditions are used to define the Calabi-Yau manifold and their equivalence is explained. To motivate the study of Calabi-Yau manifolds, the initial argument for why the compactified dimensions of heterotic string theory must be Calabi-Yau are outlined. Lastly, the Hodge numbers of a Calabi-Yau manifold are examined thoroughly with particular attention to the fourfold.

In § 3, we introduce the complete intersection construction of Calabi-Yau manifolds,

[1] A preprint titled 'Machine Learning Calabi=Yau Four-folds' [29] became available a few weeks before submission of this dissertation. At this time, the majority of the work for this dissertation was complete. Encouragingly, the results of the preprint are in agreement with the results found in this work.

again with a focus on fourfolds. An understanding is built by successive generalisation of an initially simple construction until we reach the most general CICY. After this we are ready to look at the Calabi-Yau landscape. The data available from the complete intersection construction is described and the size of the landscape is gauged by briefly outlining an additional method for constructing Calabi-Yau manifolds that uses reflexive polytopes. With some knowledge of the Calabi-Yau landscape, we motivate the application of machine learning techniques and state the goals of such an application.

Section 4 describes how NNs can be applied to learn data. The aspects covered are selected to give the reader an understanding of the techniques used in the study that is the focus of the next section. The concept of a neuron is introduced and we describe how multiple neurons can be connected to form a basic NN. The process of training a NN to a particular data set is explained using a typical network as an example. We then describe the wider process of constructing an optimal NN and explain some common techniques to do this. It should be noted that although NNs are a very interesting part of machine learning, they are a small corner in the wider field. In addition to this, the section only introduces the most basic NN architectures, completely ignoring convolutional NNs, regularisation techniques, data augmentation, and many more of the advanced topics that are crucial in some of the cutting edge NNs being developed today. For this reason, § 4 should be treated as a working introduction to machine learning and an invitation to read further.

In § 5 the dissertation transitions into a study of the ability of NNs to predict Hodge numbers of unseen CICY4s. First, the results of relevant publications are summarised and the aims of the study are outlined. The CICY4 data set being used is then examined and the process of creating and training the NNs is documented. We then present a comprehensive analysis of their performance in predicting Hodge numbers.

1 Complex Geometry

To understand Calabi-Yau manifolds a firm grasp of complex geometry is required. This section provides a brief overview of the mathematics needed to understand the Calabi-Yau manifold and its complete intersection construction. The reader is assumed to have a good grasp of real manifolds, Riemannian geometry, differential forms, and de Rham cohomology. A recap of these topics can be found in [44]. In an effort to keep this section brief, familiarity with fibre bundles has also been assumed. It is important that the reader has at least a working knowledge of vector bundles and the tangent bundle. An excellent introduction to these topics can be found in Chapter 5 of [35]. The following is based off [44, 12, 52, 8, 41, 23].

1.1 Complex Manifolds

The definitions of complex and real manifolds appear quite similar but have one crucial difference: the transition functions of a complex manifold must be holomorphic.

Definition 1.1.1. A function $f : \mathbb{C}^m \to \mathbb{C}$ is holomorphic if $f = f_1 + if_2$ satisfies the Cauchy-Riemann relations

$$\frac{\partial f_1}{\partial x^\mu} = \frac{\partial f_2}{\partial y^\mu}, \qquad \frac{\partial f_2}{\partial x^\mu} = -\frac{\partial f_1}{\partial y^\mu}, \tag{1.1.1}$$

for each $z^\mu = x^\mu + iy^\mu$.

Definition 1.1.2. M is a complex manifold if:

(i) M is a topological space.

(ii) M is endowed with a family of pairs $\{(U_i, \varphi_i)\}$.

(iii) $\{U_i\}$ is a family of open sets which covers M. The map φ_i is a homeomorphism from U_i to an open subset U of $\mathbb{C}^m$.

(iv) Given U_i and U_j such that $U_i \cap U_j \neq \emptyset$, the map $\psi_{ji} = \varphi_j \circ \varphi_i^{-1}$ from $\varphi_i(U_i \cap U_j)$ to $\varphi_j(U_i \cap U_j)$ is holomorphic.

The transition functions f_{ij} of a complex manifold relate the coordinates on overlapping coordinate patches U_i and U_j by

$$z_i^\mu = f_{ij}^\mu(z_j), \tag{1.1.2}$$

where z^μ are complex coordinates. The condition that f_{ij} is holomorphic amounts to z_i^μ being functions of z_j^μ but not their complex conjugates $z_j^{\bar{\mu}}$. The holomorphic condition is more restrictive than the C^∞ condition of real manifolds. All m-dimensional complex manifolds are also $2m$-dimensional real manifolds but not vice versa.

Remark. For real coordinates we use Latin indices and for complex coordinates we use Greek indices. We will use a bar over the index to denote the complex conjugate coordinate so that $z^{\bar{\mu}} = \bar{z}^{\bar{\mu}}$. Any complex manifold with $2m$ real coordinates has m complex coordinates z^μ each with a complex conjugate $z^{\bar{\mu}}$.

Using the holomorphic function we can define a holomorphic map.

Definition 1.1.3. Let $f : M \to N$ with M and N being complex manifolds of complex dimension m and n respectively. Take a point p in a chart (U, φ) of M and let (V, ψ) be a chart of N such that $f(p) \in V$. Assign coordinates $\{z^\mu\} = \varphi(p)$ and $\{w^\mu\} = \psi(f(p))$ so that there is a map $\psi \circ f \circ \varphi^{-1} : \mathbb{C}^m \to \mathbb{C}^n$. If each function w^ν is a holomorphic function of z^μ, then f is called a holomorphic map. If the inverse map $f^{-1} : N \to M$ exists and is also holomorphic we call f biholomorphic and say that M is biholomorphic to N.

In going from a $2m$-dimensional real manifold to an m-dimensional complex manifold we have a notion of complexification. The complexified tangent space at a point p on a manifold is simply $T_pM^{\mathbb{C}} = T_pM \otimes \mathbb{C}$ with elements $z = x + iy$ for $x, y \in T_pM$. The cotangent space $T_p^*M^{\mathbb{C}}$ can also be complexified in this way and will remain dual to the complexified tangent space. Taking the basis of T_pM and T_p^*M for a manifold with m complex dimensions to be

$$\left\{ \frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^m}; \frac{\partial}{\partial y^1}, \ldots \frac{\partial}{\partial y^m} \right\}, \tag{1.1.3}$$

$$\{\mathrm{d}x^1, \ldots, \mathrm{d}x^m; \mathrm{d}y^1, \ldots, \mathrm{d}y^m\} \tag{1.1.4}$$

respectively, we find a natural basis for $T_pM^{\mathbb{C}}$ and $T_p^*M^{\mathbb{C}}$ formed by

$$\frac{\partial}{\partial z^\mu} = \frac{1}{2}\left(\frac{\partial}{\partial x^\mu} - i\frac{\partial}{\partial y^\mu}\right), \quad \frac{\partial}{\partial z^{\bar{\mu}}} = \frac{1}{2}\left(\frac{\partial}{\partial x^{\bar{\mu}}} + i\frac{\partial}{\partial y^{\bar{\mu}}}\right), \tag{1.1.5}$$

$$\mathrm{d}z^\mu = \mathrm{d}x^\mu + i\mathrm{d}y^\mu, \quad \mathrm{d}z^{\bar{\mu}} = \mathrm{d}x^{\bar{\mu}} - i\mathrm{d}y^{\bar{\mu}} \tag{1.1.6}$$

respectively, where μ and $\bar{\mu}$ run from 1 to m.

We would like some way to recognise if a real manifold can be considered a complex manifold. To enable this we introduce an alternative definition of a complex manifold via complex structure. Denoting the tangent bundle with TM,

Definition 1.1.4. A smooth tensor field $J \in \Gamma(TM \otimes T^*M)$ is the almost complex structure of a manifold M if it is an endomorphism $J : TM \to TM$ that obeys $J^2 = -\mathbb{I}$. A manifold with an almost complex structure is called an almost comlpex manifold.

At a given point on the manifold the almost complex structure acts on the real basis as

$$J_p\left(\frac{\partial}{\partial x^a}\right) = \frac{\partial}{\partial y^a}, \quad J_p\left(\frac{\partial}{\partial y^a}\right) = -\frac{\partial}{\partial x^a}. \tag{1.1.7}$$

So that in this basis it can be written in matrix form as

$$J_p = \begin{pmatrix} 0 & -\mathbb{I}_m \\ \mathbb{I}_m & 0 \end{pmatrix} \tag{1.1.8}$$

for a $2m$-dimensional manifold. The almost complex structure is naturally extended to $T_pM^{\mathbb{C}}$ where we can use (1.1.5) to see that it acts on the basis of complex coordinates as

$$J_p\left(\frac{\partial}{\partial z^\mu}\right) = i\frac{\partial}{\partial z^\mu}, \quad J_p\left(\frac{\partial}{\partial z^{\bar{\mu}}}\right) = -i\frac{\partial}{\partial z^{\bar{\mu}}}. \tag{1.1.9}$$

So in this basis the almost complex structure can be expressed as

$$J_p = i\mathrm{d}z^\mu \otimes \frac{\partial}{\partial z^\mu} - i\mathrm{d}z^{\bar{\mu}} \otimes \frac{\partial}{\partial z^{\bar{\mu}}}. \tag{1.1.10}$$

In components this takes the diagonal form

$$J_{p\,\nu}^{\mu} = i\delta^\mu{}_\nu, \quad J_{p\,\bar{\nu}}^{\bar{\mu}} = -i\delta^{\bar{\mu}}{}_{\bar{\nu}} \tag{1.1.11}$$

with all other components zero. We will refer to this as the canonical form of J.

Given an almost complex manifold (M, J) and any single point p on M, one can always find local coordinates z^μ such that J takes the canonical form at the point p. The important question is can we choose complex coordinates that put J in the canonical form for the whole open set near p? Such coordinates are called local holomorphic coordinates and if they exist we say that the almost complex structure is integrable. Drawing inspiration from general relativity's Riemann tensor, we attempt to answer this question by defining a new tensor field from J and its derivatives called the Nijenhuis tensor

$$N^c{}_{ab} = J^d{}_a(\partial_d J^c{}_b - \partial_b J^c{}_d) - J^d{}_b(\partial_d J^c{}_a - \partial_a J^c{}_d). \tag{1.1.12}$$

Note that the proof that this is a tensor can be reformulated by writing

$$\tilde{N}^c{}_{ab} = J^d{}_a(\nabla_d J^c{}_b - \nabla_b J^c{}_d) - J^d{}_b(\nabla_d J^c{}_a - \nabla_a J^c{}_d), \tag{1.1.13}$$

where ∇ are covariant derivatives, and showing that $\tilde{N}$, which is manifestly a tensor, is equal to N. This can be done fairly easily using $J^2 = -\mathbb{I}$ and the definition of the covariant derivative in general relativity.

The importance of the Nijenhuis tensor comes from the Newlander-Nirnberg theorem (for a proof see [54]).

Theorem 1.1.1 (**Newlander and Nirnberg**). *An almost complex structure J is integrable if and only if its Nijenhuis tensor vanishes.*

The importance of integrability comes from

Definition 1.1.5. For a real manifold M with an integrable almost complex structure J, (M, J) defines a complex manifold. An integrable almost complex structure is referred to as a complex structure.

To see this, consider a manifold M covered with open sets $U_{(\alpha)}$ that has an integrable complex structure J. Since J is integrable, we can find a local holomorphic coordinate system $z^\mu_{(\alpha)}$ that puts it in its canonical form globally. This means the coordinates $z^\mu_{(\alpha)}$ and $z^\mu_{(\beta)}$ on the overlap $U_{(\alpha)} \cap U_{(\beta)}$ must be related in a way that preserves J. To preserve the diagonal nature of the canonical form, the coordinates on the overlap must be related by holomorphic coordinate transformations $z^\mu \to \tilde{z}^\nu(z^\mu)$. Recalling that holomorphic

transition functions mean that the manifold is complex, we confirm that (M, J) is a complex manifold if J is integrable. Proof of the converse proceeds in a similar way.

Understanding that all complex manifolds are equipped with a complex structure allows for a deeper analysis of the tangent space. Consider the two vectors $Z = Z^\mu \partial/\partial z^\mu$ and $\bar{Z} = \bar{Z}^{\bar{\mu}} \partial/\partial z^{\bar{\mu}}$. By acting on them with J_p we get $J_p Z = iZ$ and $J_p \bar{Z} = -i\bar{Z}$ respectively i.e. Z and $\bar{Z}$ are eigenvectors of J_p with eigenvalue $+1$ and -1 respectively. This tells us that $T_p M^{\mathbb{C}}$ can be decomposed into two disjoint vector spaces as

$$T_p M^{\mathbb{C}} = T_p M^+ \oplus T_p M^-, \tag{1.1.14}$$

where

$$T_p M^\pm = \{Z \in T_p M^{\mathbb{C}} | J_p Z = \pm iZ\}. \tag{1.1.15}$$

Since this can be done at all points p on M, the decomposition extends to the full complexified tangent bundle $TM^{\mathbb{C}}$. We refer to TM^+ as the holomorphic tangent bundle and TM^- as the anti-holomorphic tangent bundle. The duals of these can be taken to give the holomorphic and anti-holomorphic cotangent bundles. It is clear that TM^+ and TM^- are related by complex conjugation and so isomorphic. Upon complexifying the tangent bundle this isomorphism means we can choose to use only TM^+ in place of the real tangent bundle TM. Because of this we will often write TM in the context of complex manifolds to mean the holomorphic tangent bundle.

Projection tensors can be defined as

$$P^a{}_b = \frac{1}{2}(\delta^a{}_b - iJ^a{}_b), \quad Q^a{}_b = \frac{1}{2}(\delta^a{}_b + iJ^a{}_b) \tag{1.1.16}$$

so that any $Z \in T_p M^{\mathbb{C}}$ can be decomposed as $Z = Z^+ + Z^-$, where $Z^\pm \in T_p M^\pm$. We refer to $Z \in T_p M^+$ as a holomorphic vector and $Z \in T_p M^-$ as an anti-holomorphic vector. Note that this tangent space decomposition occurs for almost complex as well as complex manifolds.

1.1.1 Example: Complex Projective Space $\mathbb{CP}^N$

The complex projective space is an important family of complex manifolds as they facilitate a particular construction of Calabi-Yau manifolds. For this reason, it is important that they are understood. We can think of $\mathbb{CP}^N$ as the space of complex lines through

the origin of $\mathbb{C}^{N+1}$. Consider $\mathbb{C}^{N+1}\backslash\{0\}$ which is characterised by the complex numbers z^μ, $\mu = 0, \ldots, N+1$ such that there is always at least one that is non-zero. These are often called the homogeneous coordinates. $\mathbb{CP}^N$ is then obtained by quotienting this space by the equivalence relation

$$(z, \ldots, z^{N+1}) \sim \lambda(z, \ldots, z^{N+1}), \quad \forall \lambda \in \mathbb{C}. \tag{1.1.17}$$

We define the i-th inhomogeneous coordinates in a region where $z^i \neq 0$ by

$$\xi^\mu_{(i)} = \frac{z^\mu}{z^i}, \tag{1.1.18}$$

which clearly obeys the equivalence relation. These define a coordinate system for the region where $z^i = 0$. They cover all of $\mathbb{CP}^N$ and since $\xi^i_{(i)} = 1$ they describe N independent coordinates so that $\mathbb{CP}^N$ has complex dimension N. Since the z^μ are never all zero, we can construct an atlas from these coordinate patches using the coordinates $\xi^\mu_{(i)}$ with the open subsets $U_i := \{z^i \neq 0\}$.

Theorem 1.1.2. *$\mathbb{CP}^N$ is a complex manifold.*

Proof. On the non-trivial intersection $U_i \cap U_j$ we have

$$\xi^\mu_{(j)} = \frac{z^\mu}{z^j} = \frac{z^\mu}{z^i}\frac{z^i}{z^j} = \frac{\xi^\mu_{(i)}}{\xi^i_{(j)}} \tag{1.1.19}$$

which is holomorphic. This means that the transition functions are holomorphic and so $\mathbb{CP}^N$ is a complex manifold. □

It should also be noted that $\mathbb{CP}^N$ is compact [12].

1.1.2 Differential Forms

The complex structure of a complex manifold M allows for the construction of projection tensors (1.1.16) which can be used to refine exterior calculus on real manifolds. The projection tensors have a natural action on complex-valued 1-forms $\omega = \omega_a \mathrm{d}x^a$ that induces a split

$$\omega^{(1,0)} = P\omega = P_a{}^b \omega_b \mathrm{d}x^a, \qquad \omega^{(0,1)} = Q\omega = Q_a{}^b \omega_b \mathrm{d}x^a \tag{1.1.20}$$

of the complex-valued 1-form into a $(1,0)$-form and a $(0,1)$-form. By the usual properties of projectors we have

$$\omega = \omega^{(1,0)} + \omega^{(0,1)}. \tag{1.1.21}$$

This extends to complex-valued k-forms to give the decomposition

$$\Omega^k(M)^{\mathbb{C}} = \bigoplus_{r+s=k} \Omega^{(r,s)}(M), \tag{1.1.22}$$

where $\Omega^k(M)^{\mathbb{C}}$ denotes the space of complex-valued k-forms and $\Omega^{(r,s)}(M)$ denotes the space of (r,s)-forms that are obtained by the application of r of the P projectors and s of the Q projectors to a complex-valued $(r+s)$-form. In terms of the k-th exterior power of the complexified cotangent bundle, which the space of complex-valued k-forms are sections of, the decomposition follows from (1.1.14)

$$\bigwedge^k T^*M^{\mathbb{C}} = \bigoplus_{r+s=k} \bigwedge^{r,s} T^*M, \tag{1.1.23}$$

where $\bigwedge^{r,s} T^*M = \bigwedge^r T^*M^+ \oplus \bigwedge^s T^*M^-$. Sections of $\bigwedge^{r,s} T^*M$ compose the space of (r,s)-forms. The result of the decomposition (1.1.22) is that any complex k-form can be written as a sum of (r,s)-forms

$$\omega = \sum_{r+s=k} \omega^{(r,s)}. \tag{1.1.24}$$

We find that $\mathrm{d}z^\mu$ and $\mathrm{d}z^{\bar{\mu}}$ of the basis (1.1.6) is a $(1,0)$-form and a $(0,1)$-form respectively. This basis can then be used to write an (r,s)-form as a complex-valued differential form with r holomorphic pieces and s anti-holomorphic pieces

$$\omega = \frac{1}{r!s!}\omega_{\mu_1\ldots\mu_r\bar{\nu}_1\ldots\bar{\nu}_s}\mathrm{d}z^{\mu_1}\wedge\ldots\wedge\mathrm{d}z^{\mu_r}\wedge\mathrm{d}z^{\bar{\nu}_1}\wedge\ldots\wedge\mathrm{d}z^{\bar{\nu}_s}. \tag{1.1.25}$$

The way that the projection operators are applied to k-forms results in the components of an (r,s)-form in the above basis being totally antisymmetric in μ and $\bar{\mu}$ separately. Notice that if $\omega \in \Omega^{(r,s)}$ then $\bar{\omega} \in \Omega^{(s,r)}$ where the bar denotes complex conjugation.

The decomposition of differential forms this way occurs on almost complex and complex manifolds. On a complex manifold we get some additional structure. The exterior derivative on a complex manifold admits a simple decomposition into holomorphic and anti-holomorphic directions as $d = \partial + \bar{\partial}$, where

$$\partial : \Omega^{(r,s)}(M) \to \Omega^{(r+1,s)}(M), \qquad \bar{\partial} : \Omega^{(r,s)}(M) \to \Omega^{(r,s+1)}(M) \tag{1.1.26}$$

are called Dolbeault operators. This decomposition of the exterior derivative is due to the complex structure being constant and so only occurs on a complex manifold. The nilpotency of the exterior derivative, $d^2 = 0$, carries over to the Dolbeault operators, $\partial^2 = \bar{\partial}^2 = 0$.

1.1.3 Dolbeault Cohomology

To probe the fundamental properties of a complex manifold we need to understand the cohomology. The Dolbeault operators associated with complex manifolds allows for a refinement of de Rham cohomology. We define Dolbeault cohomology for a complex manifold M as follows:

Definition 1.1.6. The set of $\bar{\partial}$-closed (r, s)-forms ($\omega \in \Omega^{(r,s)}(M)$ such that $\bar{\partial}\omega = 0$) on a complex manifold M is called the (r, s)-cocycle and is denoted $Z^{(r,s)}_{\bar{\partial}}(M)$.

Definition 1.1.7. The set of $\bar{\partial}$-exact (r, s)-forms ($\omega \in \Omega^{(r,s)}(M)$ such that $\omega = \bar{\partial}\eta$ for some $\eta \in \Omega^{(r,s-1)}(M)$) on a complex manifold M is called the (r, s)-coboundary and is denoted $B^{(r,s)}_{\bar{\partial}}(M)$.

Definition 1.1.8. The (r, s)th $\bar{\partial}$-cohomology group is the quotient space

$$H^{(r,s)}_{\bar{\partial}}(M) = Z^{(r,s)}_{\bar{\partial}}(M)/B^{(r,s)}_{\bar{\partial}}(M). \tag{1.1.27}$$

An element of a Dolbeault cohomology group $[\eta] \in H^{(r,s)}{}_{\bar{\partial}}(M)$ is an equivalence class of $\bar{\partial}$-closed (r, s)-forms which differ from η by a $\bar{\partial}$-exact form.

In analogy with the Betti numbers of de Rham cohomology, we have for Dolbeault cohomology

Definition 1.1.9. The Hodge numbers are defined by $h^{r,s} = \dim_{\mathbb{C}} H^{(r,s)}_{\bar{\partial}}(M)$.

Hodge numbers of a complex dimension m manifold are often arranged using the Hodge

diamond

$$
\begin{array}{ccccccc}
 & & & h^{m,m} & & & \\
 & & h^{m,m-1} & \vdots & h^{m-1,m} & & \\
 & \iddots & & \vdots & & \ddots & \\
h^{m,0} & \cdots & \cdots & & \cdots & \cdots & h^{0,m}. \\
 & \ddots & & \vdots & & \iddots & \\
 & & h^{1,0} & \vdots & h^{0,1} & & \\
 & & & h^{0,0} & & &
\end{array}
$$

There are $(m+1)^2$ Hodge numbers in the diamond but much fewer independent Hodge numbers.

1.2 Kähler Manifolds

To understand the Kähler manifold we must first introduce and analyse the Hermitian manifold.

Definition 1.2.1. Let (M, J) be a complex manifold with a Riemannian metric g. We call g a Hermitian metric if

$$g(v,u) = g(Jv, Ju) \quad \forall v, u \in \Gamma(TM). \tag{1.2.1}$$

A complex manifold with a Hermitian metric is called a Hermitian manifold.

An equivalent definition of a Hermitian metric is that its pure holomorphic and pure anti-holomorphic components vanish i.e. $g_{\mu\nu} = g_{\bar{\mu}\bar{\nu}} = 0$.

Theorem 1.2.1. *A complex manifold (M, J) always admits a Hermitian metric.*

Proof. If g is a Riemannian metric on M we can always define for $u, v \in \Gamma(TM)$,

$$h(u,v) := \frac{1}{2}\big(g(u,v) + g(Ju, Jv)\big). \tag{1.2.2}$$

If g is positive definite then so is h. Since it can easily be seen that h satisfies the condition (1.2.1), h is a Hermitian metric. □

On a Hermitian manifold we can always define the Hermitian 2-form.

Definition 1.2.2. Let (M, J) be a complex manifold with a Hermitian metric g. The Hermitian 2-form is defined as

$$\omega(u, v) = g(Ju, v) \quad \forall v, u \in \Gamma(TM). \tag{1.2.3}$$

The Hermicity of the metric can be used to confirm that ω is a 2-form

$$\omega(u, v) = g(Ju, v) \overset{(1.2.1)}{=} g(J^2u, Jv) = -g(u, Jv) = -g(Jv, u) - -\omega(v, u). \tag{1.2.4}$$

In terms of a coordinate basis the Hermitian 2-form is written $\omega_{ab} = J_a{}^c g_{cb}$. So using the complex basis we can write the 2-form as the $(1, 1)$-form

$$\omega = i g_{\mu\bar{\nu}} \mathrm{d}z^\mu \wedge \mathrm{d}z^{\bar{\nu}}. \tag{1.2.5}$$

To study the curvature of the Hermitian manifold we need to first look at connections.

Definition 1.2.3. Let (M, J, g) be a Hermitian manifold. The manifold permits a connection that is compatible with both the Hermitian metric and complex structure, meaning

$$\nabla g = \nabla J = 0, \tag{1.2.6}$$

called the Hermitian connection.

To obtain a unique Hermitian connection we must make one further restriction. We impose that the anti-holomorphic covariant derivative of a holomorphic vector field and the holomorphic covariant derivative of an anti-holomorphic vector field must obey

$$\nabla_{\bar{\mu}} v^\nu = \partial_{\bar{\mu}} v^\nu, \quad \nabla_\mu v^{\bar{\nu}} = \partial_\mu v^{\bar{\nu}}. \tag{1.2.7}$$

Such a connection is called the Chern connection.

Theorem 1.2.2. *On a Hermitian manifold there exists a unique Chern connection.*

Proof. To prove the existence and uniqueness of the Chern connection we will attempt to construct it. The condition that the Chern connection is compatible with J implies

$$\Gamma^{\bar{\rho}}{}_{\mu\nu} = 0, \quad \Gamma^{\bar{\rho}}{}_{\bar{\mu}\nu} = 0, \quad \Gamma^{\rho}{}_{\mu\bar{\nu}} = 0, \quad \Gamma^{\rho}{}_{\bar{\mu}\bar{\nu}} = 0. \tag{1.2.8}$$

The condition unique to the Chern connection (1.2.7) implies

$$\Gamma^{\rho}{}_{\bar{\mu}\nu} = 0, \quad \Gamma^{\rho}{}_{\mu\bar{\nu}} = 0. \tag{1.2.9}$$

The only non zero components of the Chern connection are therefore $\Gamma^{\rho}{}_{\mu\nu}$ and $\Gamma^{\bar{\rho}}{}_{\bar{\mu}\bar{\nu}}$. In reference to there being no components with mixed holomorphic and anti-holomorphic indices, we say that the Chern connection is pure in its indices. All that is left is to impose the metric compatibility condition. Using that the connection is pure in its indices and that a Hermitian metric is mixed in its indices, this condition reads

$$\nabla_{\mu} g_{\nu\bar{\rho}} = \partial_{\mu} g_{\nu\bar{\rho}} - \Gamma^{\sigma}{}_{\mu\nu} g_{\sigma\bar{\rho}} = 0, \quad \nabla_{\bar{\mu}} g_{\nu\bar{\rho}} = \partial_{\bar{\mu}} g_{\nu\bar{\rho}} - \Gamma^{\bar{\sigma}}{}_{\bar{\mu}\bar{\rho}} g_{\nu\bar{\sigma}} = 0. \tag{1.2.10}$$

Solving for the connection, we get a unique Chern connection

$$\Gamma^{\rho}{}_{\mu\nu} = g^{\rho\bar{\sigma}} \partial_{\mu} g_{\bar{\sigma}\nu}, \quad \Gamma^{\bar{\rho}}{}_{\bar{\mu}\bar{\nu}} = g^{\bar{\rho}\sigma} \partial_{\bar{\mu}} g_{\sigma\bar{\nu}}. \tag{1.2.11}$$

□

We are now equipped with a unique connection to examine the curvature of Hermitian manifolds with. The form of the Chern connection greatly simplifies the Riemann tensor. Starting with the usual expression for the Riemann tensor in terms of the connection and its derivative we find that the only non-zero components are

$$R^{\mu}{}_{\nu\rho\bar{\sigma}} = -\partial_{\bar{\sigma}} \Gamma^{\mu}{}_{\rho\nu}, \quad R^{\bar{\mu}}{}_{\bar{\nu}\bar{\rho}\sigma} = -\partial_{\sigma} \Gamma^{\bar{\mu}}{}_{\bar{\rho}\bar{\nu}}. \tag{1.2.12}$$

When equipped with only a metric the next step is to contract indices of the Riemann tensor to form the Ricci tensor. The complex structure of the Hermitian manifold permits an additional way to contract the indices of the Riemann tensor,

$$\mathcal{R} = \frac{1}{4} R^{a}{}_{bcd} J^{b}{}_{a} \mathrm{d}x^{c} \wedge \mathrm{d}x^{d}, \tag{1.2.13}$$

to form what is called the Ricci 2-form. The simple form of the Riemann tensor for the Chern connection allows the Ricci 2-form to be written in a much neater way. In the decomposition (1.1.24) of the 2-form, the $(2,0)$-form and $(0,2)$-form parts vanish due to the Riemann tensor vanishing. We are then left with the Ricci 2-form written as a $(1,1)$-form with non-vanishing components

$$\mathcal{R}_{\mu\bar{\nu}} = \frac{i}{2} (R^{\rho}{}_{\rho\mu\bar{\nu}} - R^{\bar{\rho}}{}_{\bar{\rho}\mu\bar{\nu}}), \tag{1.2.14}$$

where we have made use of the canonical form of J. Evaluating the first Riemann tensor explicitly gives

$$R^{\rho}{}_{\rho\mu\bar{\nu}} = -\partial_{\bar{\nu}} \Gamma_{\mu\rho}{}^{\rho} = -\partial_{\bar{\nu}} \left(g^{\rho\bar{\sigma}} \partial_{\mu} g_{\bar{\sigma}\rho} \right) = -\partial_{\bar{\nu}} \partial_{\mu} \log \sqrt{|g|}, \tag{1.2.15}$$

where $g := \det g$. Evaluation of the second Riemann tensor yields the same but with a factor of -1. We then see that the Ricci 2-form can be written as

$$\mathcal{R} = -i\partial\bar{\partial}\log\sqrt{|g|}. \tag{1.2.16}$$

We can use the identity $\partial\bar{\partial} = (1/2)\,\mathrm{d}\big(\partial - \bar{\partial}\big)$ for complex manifolds to show that the Ricci 2-form is closed i.e. $\mathrm{d}\mathcal{R} = 0$.

We are now ready to introduce the Kähler manifold.

Definition 1.2.4. A Kähler manifold is a Hermitian manifold (M, J, g) with a Hermitian 2-form that is closed $d\omega = 0$. We call M the Kähler manifold, g the Kähler metric, and ω the Kähler form.

The Kähler form being closed has some interesting consequences. An extension of Poincaré's lemma to complex coordinates gives that any closed (r, s)-form α can be written locally as $\alpha = \partial\bar{\partial}\eta$ for some $(r-1, s-1)$-form η. Since the Kähler form is closed this means it can be written locally as $\omega_{\mu\bar{\nu}} = i\partial_\mu\partial_{\bar{\nu}}\mathcal{K}$ for some scalar function $\mathcal{K}(z, \bar{z})$ called the Kähler potential. By looking at (1.2.5) we then see that the Kähler potential can be used to write the Kähler metric locally as

$$g_{\mu\bar{\nu}} = \partial_\mu\partial_{\bar{\nu}}\mathcal{K}. \tag{1.2.17}$$

On overlaps of coordinate patches $U \cap U'$ the Kähler potentials $\mathcal{K}$ and $\mathcal{K}'$ must encode a common metric i.e. $\partial\bar{\partial}\mathcal{K} = \partial\bar{\partial}\mathcal{K}'$. This is ensured by the Kähler potentials being related by a Kähler transformation

$$\mathcal{K}'(z, \bar{z}) = \mathcal{K}(z, \bar{z}) + f(z) + \bar{f}(\bar{z}), \tag{1.2.18}$$

where $f(z)$ is a holomorphic function. If a Kähler potential can be globally defined up to a Kähler transformation such that the metric is given by (1.2.17), then the manifold is Kähler. This is often how in practice whether a complex manifold is Kähler or not is determined.

The condition that ω is closed leads to an important result for the connection on a Kähler manifold. It reads explicitly

$$\begin{aligned} \mathrm{d}\omega &= \partial\omega + \bar{\partial}\omega \\ &= i\partial_\rho g_{\mu\bar{\nu}}\mathrm{d}z^\rho \wedge \mathrm{d}z^\mu \wedge \mathrm{d}z^{\bar{\nu}} + i\partial_{\bar{\rho}} g_{\mu\bar{\nu}}\mathrm{d}z^{\bar{\rho}} \wedge \mathrm{d}z^\mu \wedge \mathrm{d}z^{\bar{\nu}} = 0. \end{aligned} \tag{1.2.19}$$

The two parts must vanish separately so that we have the following conditions on the Kähler metric

$$\partial_\rho g_{\mu\bar{\nu}} = \partial_\mu g_{\rho\bar{\nu}}, \qquad \partial_{\bar{\rho}} g_{\mu\bar{\nu}} = \partial_{\bar{\nu}} g_{\mu\bar{\rho}}. \tag{1.2.20}$$

This property of the Kähler metric means that the Chern connection (1.2.11) is symmetric in its lower indices and so has vanishing torsion. To summarise, the Chern connection is unique, metric-compatible, and is also torsion-free for a Kähler manifold. However, all Hermitian manifolds are equipped with the Levi-Civita connection which is unique, metric-compatible, and torsion-free. What this means is that the Chern and Levi-Civita connection must be the same for a Kähler manifold.

On a Kähler manifold we find that the Ricci 2-form and the usual Ricci tensor are closely related. By contracting (1.2.12) to get the Ricci tensor we see that that the pure components of the Ricci tensor vanish. Computing the non-vanishing mixed components gives

$$R_{\mu\bar{\nu}} = R^{\rho}{}_{\mu\rho\bar{\nu}} = -\partial_{\bar{\nu}}\Gamma^{\rho}{}_{\rho\mu} = -\partial_{\bar{\nu}}\Gamma^{\rho}{}_{\mu\rho}. \tag{1.2.21}$$

By comparison with the components of the Ricci 2-form we find the relation

$$R_{\mu\bar{\nu}} = -i\mathcal{R}_{\mu\bar{\nu}}. \tag{1.2.22}$$

From this relation it is clear that on a Kähler manifold the Ricci tensor and Ricci 2-form encode the same information. An important result of this is that a Kähler manifold with a Ricci-flat (vanishing Ricci tensor) metric is equivalent to one that has $\mathcal{R} = 0$.

1.2.1 Example: Complex Projective Space $\mathbb{CP}^N$

We will see that the compact complex manifold $\mathbb{CP}^N$ introduced in § 1.1.1 is an example of a Kähler manifold.

Theorem 1.2.3. *The complex projective space $\mathbb{CP}^N$ is Kähler.*

Proof. On a patch U_i we can construct the scalar function

$$\mathcal{K}_i = \log\left(1 + \sum_{a=1}^{N} |\xi^a_{(i)}|^2\right). \tag{1.2.23}$$

On the non-trivial intersection of patches $U_i \cap U_j$ the relation (1.1.19) means that the scalar functions defined on these patches differ by

$$\mathcal{K}_i(z,\bar{z}) = \mathcal{K}_j(z,\bar{z}) - \log \xi^i_{(j)} - \log \bar{\xi}^i_{(j)} \tag{1.2.24}$$

i.e. a Kähler transformation. This ensures that on $\mathbb{CP}^N$

$$g_{\mu\bar{\nu}} = \partial_\mu \partial_{\bar{\nu}} \mathcal{K}_i \tag{1.2.25}$$

is globally defined. This is known as the Fubini-Study metric. If we can show that it is positive definite on TM then $\mathcal{K}$ will be a Kähler potential and so g a Kähler metric.

Working on a specific coordinate patch we can evaluate (1.2.25) explicitly giving

$$g_{\mu\bar{\nu}} = \frac{\delta_{\mu\bar{\nu}}(1+|\xi|^2) - \bar{\xi}_\mu \xi_\nu}{(1+|\xi|^2)^2}, \tag{1.2.26}$$

where $|\xi|^2 = \sum_{a=1}^N |\xi^a|^2$. Then for a real vector field $v \in \Gamma(TM)$ we get

$$g_{ab} v^a v^b = \frac{|v|^2(1+|\xi|^2) - |(\bar{\xi} v)|^2}{(1+|\xi|^2)^2}. \tag{1.2.27}$$

By application of the Cauchy-Schwartz inequality to the above, we confirm that the Fubini-Study metric is positive definite on TM and so $\mathbb{CP}^N$ is Kähler. □

1.2.2 Holonomy

The holonomy of a manifold is expressed as a subgroup of $GL(n,\mathbb{R})$ and measures how vectors are transformed by parallel transport around a closed loop on the manifold. Let (M,g) be a Riemannian manifold with affine connection Γ and consider a vector $X \in T_pM$. Parallel transport X around a closed curve C to get a new vector $X_C \in T_pM$. The loop C and connection Γ has induced a linear transformation $P_C : T_pM \to T_pM$. The set of all these transformations is the holonomy group at p and is denoted $\mathrm{Hol}_p(M)$. For connected manifolds it can be shown that the holonomy groups at different points are all isomorphic to each other so we consider the holonomy group to be independent of the point p.

If Γ is a metric connection and M an n-dimensional orientable Riemannian manifold then the length of a vector is preserved on parallel transport. This implies that the holonomy group is $\subseteq SO(n)$.

We have seen that for a Kähler manifold the Chern connection is pure in its holomorphic and anti-holomorphic indices. This means that a (anti-)holomorphic vector will

always remain (anti-)holomorphic after parallel transport as well as preserving the length of vector. This tells is that the holonomy group for a Kähler manifold of complex dimension m must be contained in $U(m) \subset SO(2m)$.

It will be useful to examine the holonomy of a Ricci-flat Kähler manifold. Let M be a Kähler manifold with complex dimension m. The vector $V \in T_pM^+$ is parallel transported around an infinitesimal rectangle with area δa^{bc} and edges that run parallel to the basis vectors ∂_b and ∂_c to get a new vector $V' \in T_pM^+$. It is a standard result that

$$V'^a = V^a + \delta a^{bc} R_{bc}{}^a{}_d V^d. \tag{1.2.28}$$

From this, we see that the matrix $\delta^a{}_d + \delta a^{bc} R_{bc}{}^a{}_d$ is an element of the holonomy group near the identity. Since the metric is Kähler these matrices are elements of the Lie algebra $\mathfrak{u}(m)$ which can be decomposed into a traceless part and a trace as

$$\mathfrak{u}(m) = \mathfrak{su}(m) \oplus \mathfrak{u}(1). \tag{1.2.29}$$

The $\mathfrak{u}(1)$ element is given by the trace

$$\delta^{bc} R_{bc}{}^a{}_a = -4\delta a^{\mu\bar{\nu}} R_{\mu\bar{\nu}}, \tag{1.2.30}$$

where we have used that the only non-vanishing components of the Riemann tensor are of the form $R_{\mu\bar{\nu}\rho\bar{\sigma}}$.

It is then clear that if the metric is Ricci-flat the $U(1)$ part of the holonomy group vanishes and the manifold has a holonomy group contained in $SU(m)$ rather than in $U(m)$. Note that the converse is also true, meaning a Kähler manifold has $\subseteq SU(m)$ holonomy if and only if the Kähler metric is Ricci-flat.

1.2.3 Hodge Theory

To understand Hodge theory on Kähler manifolds we must introduce some additional structure. Much of this will be stated without proof for brevity, [44] is a good starting point for a developing a deeper understanding. Many of the following definitions apply more generally to complex manifolds but we restrict ourselves to a Kähler manifold M with complex dimension m.

The Hodge star operator of real manifold extends naturally to the complexified tangent space

Definition 1.2.5. The Hodge star $*$ is a map

$$\bar{*} : \Omega^{(r,s)}(M) \to \Omega^{(m-r,m-s)}(M), \tag{1.2.31}$$

where $\bar{*}\beta = *\bar{\beta}$.

Definition 1.2.6. The inner product between $\alpha, \beta \in \Omega^{(r,s)}(M)$ is defined as

$$(\alpha, \beta) = \int_M \alpha \wedge \bar{*}\beta. \tag{1.2.32}$$

Definition 1.2.7. The adjoint Dolbeault operators $\partial^\dagger$ and $\bar{\partial}^\dagger$ are maps

$$\partial^\dagger : \Omega^{(r,s)}(M) \to \Omega^{(r-1,s)}(M), \qquad \bar{\partial}^\dagger : \Omega^{(r,s)}(M) \to \Omega(r, s-1)(M) \tag{1.2.33}$$

defined by

$$(\alpha, \partial\beta) = (\partial^\dagger\alpha, \beta), \qquad (\alpha, \bar{\partial}\beta) = (\bar{\partial}^\dagger\alpha, \beta). \tag{1.2.34}$$

A Kähler manifold has the additional Laplacians

$$\Delta_\partial = (\partial + \partial^\dagger)^2 = \partial\partial^\dagger + \partial^\dagger\partial, \qquad \Delta_{\bar{\partial}} = (\bar{\partial} + \bar{\partial}^\dagger)^2 = \bar{\partial}\bar{\partial}^\dagger + \bar{\partial}^\dagger\bar{\partial} \tag{1.2.35}$$

as well as the usual $\Delta = \mathrm{d}\mathrm{d}^\dagger + \mathrm{d}^\dagger\mathrm{d}$. Recall that since the Laplacian is a positive operator any k-form that is closed under it must also be closed under d and $\mathrm{d}^\dagger$. The same is true for (r, s)-forms closed under Δ_∂ or $\Delta_{\bar{\partial}}$ and their corresponding Dolbeault operators. We denote

$$\mathrm{Harm}^{(r,s)}_{\bar{\partial}}(M) = \{\omega \in \Omega^{(r,s)}(M) | \Delta_{\bar{\partial}}\omega = 0\}. \tag{1.2.36}$$

An element of $\mathrm{Harm}^{(r,s)}_{\bar{\partial}}(M)$ is called $\bar{\partial}$-harmonic form.

We can now introduce Hodge decomposition on a complex manifold.

Theorem 1.2.4 (Hodge's theorem). *$\Omega^{(r,s)}(M)$ has the following unique orthogonal decompostion*

$$\Omega^{(r,s)}(M) = \bar{\partial}\Omega^{(r,s-1)}(M) \oplus \bar{\partial}^\dagger\Omega^{(r-1,s)}(M) \oplus Harm^{(r,s)}_{\bar{\partial}}(M). \tag{1.2.37}$$

Hodge's theorem means that any (r, s)-form ω can be uniquely decomposed as

$$\omega = \bar{\partial}\alpha + \bar{\partial}^\dagger\beta + \gamma, \tag{1.2.38}$$

where $\alpha \in \Omega^{(r,s-1)}(M), \beta \in \Omega^{(r,s+1)}(M)$, and $\gamma \in \text{Harm}_{\bar{\partial}}^{(r,s)}(M)$. As a consequence of Hodge's theorem, cohomology groups have a unique harmonic representative. For de Rham and Dolbeault cohomology groups respectively this means

$$H_{dR}^k(M) = \{[\omega] | \omega \in \Omega^k(M), \Delta\omega = 0\}, \tag{1.2.39}$$

$$H_{\bar{\partial}}^{(r,s)}(M) = \{[\omega] | \omega \in \Omega^{(r,s)}(M), \Delta_{\bar{\partial}}\omega = 0\}. \tag{1.2.40}$$

A Kähler manifold has the property that if a form is harmonic with respect to one of d, ∂, or $\bar{\partial}$ it must be harmonic with respect the other two also. Specifically

$$\Delta = 2\Delta_{\partial} = 2\Delta_{\bar{\partial}}. \tag{1.2.41}$$

This has an important consequence. Consider the de Rham and Dolbeault cohomology groups in terms of their harmonic representative (1.2.39) and (1.2.40). Complexification of (1.2.39) gives $H_{dR}^k(M)^{\mathbb{C}} = \{[\omega] | \omega \in \Omega^k(m)^{\mathbb{C}}, \Delta\omega = 0\}$. Since M is Kähler we can invoke (1.2.41) along with the decomposition of the complexified space of k-forms into (r,s)-forms (1.1.22) to find

$$H_{dR}^k(M)^{\mathbb{C}} = \bigoplus_{r+s=k} H_{\bar{\partial}}^{(r,s)}(M). \tag{1.2.42}$$

From this, we see that the Betti and Hodge numbers of a Kähler manifold are related by

$$b^k = \sum_{r+s=k} h^{r,s}. \tag{1.2.43}$$

The Euler number is then related to the Hodge numbers by

$$\chi = \sum_{r,s} (-1)^{r+s} h^{r,s}. \tag{1.2.44}$$

1.2.4 Hodge Numbers

The Kählericity of a manifold M with $\dim_{\mathbb{C}} M = m$ puts restrictions on the number of independent Hodge numbers. Due to complex conjugation and the Hodge star respectively, we find

$$h^{r,s} = h^{s,r}, \tag{1.2.45}$$

$$h^{r,s} = h^{m-r,m-s}. \tag{1.2.46}$$

There are also the restrictions that odd Betti numbers are even and that even Betti numbers are positive. The proofs of these are omitted as we will not find use for them in later discussions.

It will be useful to note the form of the Hodge diamond for a generic Kähler manifold with complex dimension 4:

$$
\begin{array}{ccccccccc}
 & & & & h^{0,0} & & & & \\
 & & & h^{1,0} & & h^{1,0} & & & \\
 & & h^{2,0} & & h^{1,1} & & h^{2,0} & & \\
 & h^{3,0} & & h^{2,1} & & h^{2,1} & & h^{3,0} & \\
h^{4,0} & & h^{3,1} & & h^{2,2} & & h^{3,1} & & h^{4,0}. \\
 & h^{3,0} & & h^{2,1} & & h^{2,1} & & h^{3,0} & \\
 & & h^{2,0} & & h^{1,1} & & h^{2,0} & & \\
 & & & h^{1,0} & & h^{1,0} & & & \\
 & & & & h^{0,0} & & & &
\end{array}
$$

1.3 Holomorphic Vector Bundles

Having spent some time working towards an understanding of Kähler manifolds, we now make a slight change of direction to cover some topics important for later sections. Upon complexifying the tangent space of a complex manifold we saw via the eigenspaces of the complex structure that the complexified tangent bundle decomposes into a holomorphic and anti-holomorphic tangent bundle. We understood TM^+ to be the holomorphic tangent bundle only in name. Here we will formalise the notion of holomorphic vector bundles.

Definition 1.3.1. Let $E \xrightarrow{\pi} M$ with E and M complex manifolds and π a holomorphic map. E is a holomorphic vector bundle if the local trivialisation maps $\phi_i : \pi^{-1}(U) \to U_i \times \mathbb{C}^k$ are biholomorphic. This is equivalent to the complex vector bundle E admitting holomorphic transition functions.

The simplest holomorphic vector bundle is the trivial bundle $M \times \mathbb{C}^k$. The holomorphic tangent and cotangent bundles also meet the criteria of a holomorphic vector bundle. This extends to $\bigwedge^{r,0} T^*M$ as defined in § 1.1.2 to make it a holomorphic vector bundle too.

Holomorphic vector bundles also have the notion of a line bundle. A holomorphic vector bundle of rank one is called a holomorphic line bundle. The typical example is the canonical line bundle $K_M := \bigwedge^{m,0} T^*M$ for a manifold M with complex dimension m. We should recall a useful property of line bundles that the tensor product of any two line bundles is also a line bundle.

1.3.1 Example: Complex Projective Space $\mathbb{CP}^N$

We will now examine the holomorphic line bundles of $\mathbb{CP}^N$. There is the tautological line bundle

$$\mathscr{O}(-1) := \{(l, v) \in \mathbb{CP}^N \times \mathbb{C}^{N+1} | v \in l\} \tag{1.3.1}$$

which comes from attaching to each point in $\mathbb{CP}^N$ the line through the origin it represents in $\mathbb{C}^{N+1}$ as a fibre. The dual of $\mathscr{O}(-1)$ is denoted $\mathscr{O}(1)$ and is called the hyperplane line bundle. We can take tensor products of k hyperplane line bundles to give new holomorphic line bundles over $\mathbb{CP}^N$ that we denote $\mathscr{O}(k)$. For $k \geq 0$ the vector space of $\Gamma(\mathscr{O}(k))$ is identified with the set of homogeneous polynomials of degree k in $\mathbb{CP}^N$. This means that the homogeneous coordinates of $\mathbb{CP}^N$ (z^μ, $\mu = 0, \ldots, N$) are sections of $\mathscr{O}(1)$.

1.4 Chern Classes

Given a fibre F, a structure group G, and a base space M, we can construct many fibre bundles over M by choosing different transition functions. We would like a way of classifying these fibre bundles depending on how much they differ from the trivial bundle $M \times F$. Characteristic classes are concerned with this classification. These are subsets of cohomology classes of M that measure the obstruction of the bundle to being a trivial bundle. Chern classes are a particular type of characteristic class relevant to Calabi-Yau manifolds. The discussion of Chern classes will be done in the language of differential geometry, rather than the axiomatic approach of algebraic geometry, and with their application to defining the Calabi-Yau manifold in mind. A more general discussion of Chern classes and characteristic classes can be found in Chapter 11 of [44] and in Chapter 6 of [19]. These are good resources for proofs of the identities introduced in this section.

Definition 1.4.1. Let $\mathcal{V}$ be a complex vector bundle over a manifold M and let $\mathcal{F}$ be the curvature two-form of a connection on $\mathcal{V}$. The total Chern class $c(\mathcal{V})$ of $\mathcal{V}$ is

$$c(\mathcal{V}) := \det\left(1 + \frac{i\mathcal{F}}{2\pi}\right). \tag{1.4.1}$$

The total Chern class can be written as a direct sum of forms of even degree by an expansion in powers of curvature

$$c(\mathcal{V}) = c_0(\mathcal{V}) + c_1(\mathcal{V}) + \ldots + c_r(\mathcal{V}), \tag{1.4.2}$$

where $c_k(\mathcal{V})$ are called Chern forms and r is the complex rank of $\mathcal{V}$. The Chern forms are $2k$-forms that are representatives of the cohomology class they define. The cohomology classes are called the Chern classes.

The Chern classes do not depend on the choice of connection used to define the curvature two-form. Different curvature two-forms give different Chern forms that are in the same Chern class.

It is often useful to have explicit formulae for the Chern classes. This can be obtained by expanding the determinant to give

$$\begin{aligned}
c_0(\mathcal{V}) &= [1], \\
c_1(\mathcal{V}) &= \left[\frac{i}{2\pi}\operatorname{Tr}\mathcal{F}\right], \\
c_2(\mathcal{V}) &= \left[\frac{1}{2}\left(\frac{i}{2\pi}\right)^2 (\operatorname{Tr}\mathcal{F}\wedge\operatorname{Tr}\mathcal{F} - \operatorname{Tr}\mathcal{F}\wedge\mathcal{F})\right], \\
&\vdots \\
c_r(\mathcal{V}) &= \left[\left(\frac{i}{2\pi}\right)^k \det\mathcal{F}\right].
\end{aligned} \tag{1.4.3}$$

When the vector bundle E over the manifold M is the holomorphic tangent bundle TM, we call the Chern classes $c_k(TM)$ the Chern classes of the manifold and denote them $c_k(M)$ or just c_k. We will primarily be concerned with Chern classes of Kähler manifolds, in particular with the first chern class c_1. The Ricci two-form of a Kähler manifold can be used to write its first Chern class as

$$c_1(M) = \left[\frac{\mathcal{R}}{2\pi}\right]. \tag{1.4.4}$$

We will now show that the first Chern class of a Kähler manifold is invariant under a smooth variation of the metric $g_{ab} \to g'_{ab} = g_{ab} + \delta g_{ab}$. Under such a variation the Ricci two-form (1.2.16) changes by

$$\mathcal{R}' = \mathcal{R} - \frac{i}{2}\partial\bar{\partial}(g^{ab}\delta g_{ab}). \tag{1.4.5}$$

Using the identity $\partial\bar{\partial} = \frac{1}{2}\,\mathrm{d}\big(\partial - \bar{\partial}\big)$ we find that $\delta\mathcal{R}$ is an exact form so that c_1 is left invariant.

Associated with the Chern class is the Chern character. If we write the total Chern class in the form $c(\mathcal{V}) = \prod_{i=1}^{r}(1 + x_i)$ then the Chern character is $\mathrm{ch}(\mathcal{V}) = \sum_i e^{x_i}$. The utility of the Chern character is in the following identities:

$$\mathrm{ch}(\mathcal{V} \otimes \mathcal{W}) = \mathrm{ch}(\mathcal{V}) \wedge \mathrm{ch}(\mathcal{W}), \qquad \mathrm{ch}(\mathcal{V} \oplus \mathcal{W}) = \mathrm{ch}(\mathcal{V}) + \mathrm{ch}(\mathcal{W}). \tag{1.4.6}$$

The first few terms of the expansion of the Chern character are

$$\mathrm{ch}(\mathcal{V}) = r + c_1(\mathcal{V}) + \frac{1}{2}(c_1(\mathcal{V})^2 - 2c_2(\mathcal{V})) + \frac{1}{6}(c_1(\mathcal{V})^3 - 3c_1(\mathcal{V})c_2(\mathcal{V}) + 3c_3(\mathcal{V})) + \dots . \tag{1.4.7}$$

A property of the Chern classes that will be useful in later discussion requires an understanding of short exact sequences.

Definition 1.4.2. A sequence of maps and vector spaces

$$A_1 \xrightarrow{a_1} A_2 \xrightarrow{a_2} \dots \xrightarrow{a_n} A_{n+1} \tag{1.4.8}$$

is an exact sequence if for all $i = 1, \dots n-1$

$$\ker(a_{i+1}) = \mathrm{Im}(a_i). \tag{1.4.9}$$

A short exact sequence is an exact sequence of the form

$$0 \xrightarrow{a} B \xrightarrow{b} C \xrightarrow{c} D \xrightarrow{d} 0, \tag{1.4.10}$$

where 0 denotes a trivial vector space.

The short exact sequence (1.4.10) can be interpreted as the statement that $A \subseteq B$ and $C = B/A$.

If vector bundles $\mathcal{V}$, $\mathcal{V}'$ and $\mathcal{V}''$ obey the short exact sequence

$$0 \longrightarrow \mathcal{V}' \longrightarrow \mathcal{V} \longrightarrow \mathcal{V}'' \longrightarrow 0 \tag{1.4.11}$$

then their Chern classes obey the useful identity

$$c(\mathcal{V}) = c(\mathcal{V}') \wedge c(\mathcal{V}''). \tag{1.4.12}$$

Note that if $\mathcal{V}$ is the direct sum bundle $\mathcal{V}' \oplus \mathcal{V}''$ then the short exact sequence (1.4.11) holds.

Another useful identity is that the first Chern class of the tensor product of two line bundles L and L' is given by [7]

$$c_1(L \otimes L') = c_1(L) + c_1(L'). \tag{1.4.13}$$

1.4.1 Example: Complex Projective Space $\mathbb{CP}^N$

The homogeneous coordinates z^μ, $\mu = 0, \ldots, N$ can be used to span the holomorphic tangent bundle $T\mathbb{C}^{N+1}$ by the tangent vectors $s^\mu(z)\frac{\partial}{\partial z^\mu}$, where $s^\mu \in \mathscr{O}(1)$. We can also use the homogeneous coordinates to span $T\mathbb{CP}^N$ by the tangent vectors $s^\mu(z)\frac{\partial}{\partial z^\mu}$, where $s^\mu \in \mathscr{O}(1)$, but we must account for the equivalence relation (1.1.17) that makes overall rescalings in $\mathbb{CP}^N$ trivial. This equivalence relation enforces that the map from $\mathscr{O}(1)^{\oplus(N+1)}$ (the direct sum of $N+1$ hyperplane line bundles) to $T\mathbb{CP}^N$ has a kernel that is the trivial line bundle $\mathbb{C}$. This is summarised with the short exact sequence

$$0 \longrightarrow \mathbb{C} \longrightarrow \mathscr{O}(1)^{\oplus(N+1)} \longrightarrow T\mathbb{CP}^N \longrightarrow 0 \tag{1.4.14}$$

called the Euler sequence. For the trivial line bundle we have $c(\mathbb{C}) = 1$. If we then apply (1.4.12) to the Euler sequence we see

$$c(\mathbb{CP}^N) := c(T\mathbb{CP}^N) = c(\mathscr{O}(1)^{\oplus(N+1)}) = c(\mathscr{O}(1))^{N+1}. \tag{1.4.15}$$

Since $\mathscr{O}(1)$ is a line bundle it is rank one and the expansion of its total Chern class terminates at the first Chern class. If we denote $x = c_1(\mathscr{O}(1))$ so that $c(\mathscr{O}(1)) = 1 + x$ we then get the expression for the total Chern class of $\mathbb{CP}^N$

$$c(\mathbb{CP}^N) = (1 + x)^{N+1}. \tag{1.4.16}$$

2 The Calabi-Yau Manifold

The Calabi-Yau manifold has its roots in the 1950s conjecture by Calabi [11]

Conjecture 2.0.1 (**Calabi**). *Let (M, g, ω) be a compact Kähler manifold and $\mathcal{C}$ be an arbitrary $(1,1)$-form such that $[\mathcal{C}] = [c_1(M)]$. Then there exists a unique Kähler metric $\tilde{g}$ and Kähler form $\tilde{\omega}$ such that $[\omega] = [\tilde{\omega}] \in H^2_{dR}(M)$ and that $\mathcal{C} = \mathcal{R}(\tilde{\omega})$.*

Such a conjecture is valuable since it relates geometry (curvature) to topology (Chern classes). Calabi was able to prove uniqueness but not existence of $(\tilde{g}, \tilde{\omega})$. Such an existence proof came from Yau in the 1970s [56]

Theorem 2.0.1 (**Yau**). *The Calabi conjecture holds.*

In celebration of Calabi's conjecture and Yau's proof, the Calabi-Yau manifold is defined as the following special case of the Calabi conjecture:

Definition 2.0.1. A Calabi-Yau manifold is a compact Kähler manifold with a vanishing first Chern class $c_1 = 0$.

This definition can also be generalised to non-compact Kähler manifolds to give non-compact Calabi-Yau manifolds. These non-compact Calabi-Yau manifolds enter physics through brane-world descriptions of our universe [46]. In this work we only consider compact Calabi-Yau manifolds.

Yau's theorem ensures that given a Calabi-Yau manifold there always exists a Kähler metric with vanishing Ricci curvature (equivalent to a zero Ricci 2-form). Recalling that the first Chern class of a Kähler manifold M is invariant under metric variations, it is clear from (1.4.4) that if M admits a Ricci-flat metric its first Chern class vanishes. This means that we have an equivalent definition of a Calabi-Yau manifold as a compact Kähler manifold with zero Ricci 2-form. If we recall the analysis of holonomy in § 1.2.2, this is

quickly supplemented by a third equivalent definition: a Calabi-Yau manifold is a compact Kähler manifold of complex dimension m with a holonomy group contained in $SU(m)$.

There are in fact many equivalent definitions of a Calabi-Yau manifold in the literature. We give an extended definition of the Calabi-Yau manifold using a select number of the definitions

Definition 2.0.2. A Calabi-Yau manifold is a compact m-dimensional Kähler manifold M that meets the following equivalent conditions:

(i) The first Chern class vanishes; $c_1(M) = 0$.

(ii) There exists a Ricci-flat metric.

(iii) Has holonomy contained in $SU(m)$.

(iv) There exists a globally defined and nowhere vanishing holomorphic m-form.

In what follows we will restrict our definition of a Calabi-Yau manifold to one with a holonomy group of exactly $SU(m)$. When we refer to the Calabi-Yau manifold we mean one with this extra restriction. To do so is common in the literature and is often done implicitly. The reason for this is that the landscape of Calabi-Yau manifolds with a holonomy group smaller than $SU(m)$ is populated by manifolds that are some lower dimensional Calabi-Yau producted with a $2n$-torus. Such manifolds are undesirable as they contain no new information and create redundancies in classification attempts. We will see another reason for restricting to $SU(m)$ holonomy shortly.

Before explaining the origin of condition (iv) we want to provide a motivation for the study of Calabi-Yau manifolds in string theory.

2.1 Application to String Theory

The Calabi-Yau manifold entered physics through the 1985 work of Candelas, Horowitz, Strominger and Witten [14]. We will outline this work and recreate in brief their reasoning for a simplified case. The original paper and [5, 23] offer more detail. The starting point is with the 10-dimensional theory of the heterotic superstring and the goal is to create a

sensible low-energy theory that incorporates Einstein's general relativity and the Standard Model.

We must first clarify what a sensible low-energy theory is. The most obvious requirement is that it is 4-dimensional. To achieve this we work with a 10-dimensional manifold $M_{10} = T_4 \times M_6$, where T_4 is a maximally symmetric manifold called the external manifold and M_6 some compact 6-dimensional manifold called the internal manifold. The 6-dimensional manifold can then be compactified so that in the low-energy limit physics is restricted to T_4. In what follows, the indices tangent to M_{10} are denoted by M and are decomposed into indices tangent to T_4 and M_6 as $M = (\mu, m)$ respectively.

We next need to deal with the supersymmetry (SUSY) of the heterotic string. Experiment has shown that the everyday world is not supersymmetric. However, this does not rule out keeping some SUSY from the high-energy theory that then breaks at an energy larger than what has been observed. This is actually desirable as such a scenario remedies one of the most troubling issues with the Standard Model, the gauge hierarchy problem. SUSY offers an explanation as to why the mass scale of the weak interaction symmetry breaking is so small compared to other fundamental mass scales such as the Planck mass. It does this by having the Higgs doublet be the the superpartner of a massless charged fermion, ensuring that it is exactly massless in the limit of unbroken SUSY. Another compelling reason for low-energy SUSY is that a state of unbroken SUSY in 4 dimensions always obeys the equations of motion. This is welcomed since finding compactified solutions to the equations of motion can be difficult. With these reasons in mind we seek a theory with unbroken $\mathcal{N} = 1$ supergravity ($\mathcal{N} \geq 1$ theories are not phenomenologically viable).

Heterotic string theory is a promising starting point for compactification since in addition to having an $E_8 \times E_8$ gauge group which naturally accommodates grand unified theories, it has $\mathcal{N} = 1$ SUSY in 10 dimensions. Obtaining $\mathcal{N} = 1$ SUSY in 4 dimensions then becomes the task of ensuring unbroken SUSY for the compactification. We denote the infinitesimal SUSY parameters $\eta(x^M)$, where x^M are the spacetime coordinates of the 10 dimensional manifold, and the conserved supercharges associated with these parameters are denoted Q. An unbroken SUSY is a conserved supercharge that annihilates the vacuum $|\Omega\rangle$. The statement $Q|\Omega\rangle = 0$ is equivalent to the statement $\langle\Omega|\{U, Q\}|\Omega\rangle = 0$ for

every field U in the theory. This is trivial for bosonic fields but for fermionic fields it results in a restriction on the compactification. For fermionic U we have that $\{U, Q\} = \delta_\eta U$, where δ_η denotes the SUSY variation associated with the supercharge. Then at tree level the condition of unbroken SUSY becomes that $\delta_\eta U = 0$ for every elementary field in the theory.

A theory of $\mathcal{N} = 1$ supergravity in 10 dimensions contains a three-form field strength H_{MNP}, some two-form Yang Mills field strengths F_{MN}, a dilaton Φ, and some fermionic fields coupled to gravity. Supposing that T_4 is maximally symmetric means that the mixed and pure T_4 indices of H and F vanish. We also make the simplifying assumptions that $H = 0$ for pure M_6 indices and that the dilaton is constant. Having made these assumptions, we are left two non-trivial conditions for unbroken SUSY coming from $\delta_\eta U = 0$ for the fermionic fields. These come from the variation of the gravitino ψ_M and the gluinos (adjoint fermions of the Yang-Mills supermultiplet) χ,

$$\delta_\eta \psi_M = \nabla_M \eta, \tag{2.1.1}$$

$$\delta\chi = -\frac{1}{4} F_{MN} \Gamma^{MN} \eta, \tag{2.1.2}$$

where Γ with multiple indices is the antisymmetrised product of individual gamma matrices. We will first look at the consequence of the gravitino SUSY variation vanishing.

The condition arising from the gravitino variation is the Killing spinor equation which is interpreted as demanding that η is a covariantly constant spinor. Supposing that the 10 dimensional manifold is $T_4 \times M_6$ means the covariantly constant spinor can be decomposed according to the product structure as

$$\eta = \eta_{(4)} \otimes \eta_{(6)}. \tag{2.1.3}$$

This allows for a separate examination of the Killing equation on the external and internal manifolds.

We consider the external manifold where the Killing equation reads $\nabla_\mu \eta_{(4)} = 0$ which implies the following integrability condition

$$[\nabla_\mu, \nabla_\nu] \eta_{(4)} = \frac{1}{4} R_{\mu\nu\rho\sigma} \Gamma^{\rho\sigma} \eta_{(4)} = 0. \tag{2.1.4}$$

For a maximally symmetric space the Riemann tensor is set by its symmetries to be $R_{\mu\nu\rho\sigma} = (R/12)(g_{\mu\rho} g_{\nu\sigma} - g_{\mu\sigma} g_{\nu\rho})$. Because of this, (2.1.4) for a maximally symmetric

space is the condition that the Ricci scalar R is zero. This means that the external manifold is Minkowski space $\mathbb{M}_4$. This is an assuring result since $\mathcal{N} = 1$ SUSY is also possible in the much less desirable anti-de Sitter space (maximally symmetric space with $R < 0$). For Minkowski space $\nabla_\mu = \partial_\mu$ so that $\eta_{(4)}$ is a constant spinor which for the compactified theory on $\mathbb{M}_4$ is the infinitesimal parameter of an unbroken global SUSY.

We now turn to the Killing equation on the internal manifold $\nabla_m \eta_{(6)} = 0$. Equipped with the covariantly constant spinor $\eta_{(6)}$ provided by the Killing equation, we will attempt to probe the structure of M_6. The SUSY parameter of the heterotic string is a Majorana-Weyl spinor [43] that is in the **16** irrep of $Spin(1,9)$. It can be decomposed under $SL(2,\mathbb{C}) \times SU(4)$ as

$$\mathbf{16} = (\mathbf{2},\mathbf{4}) \oplus (\bar{\mathbf{2}},\bar{\mathbf{4}}) \longrightarrow \eta = \eta_{(4)+} \otimes \eta_{(6)+} \oplus \eta_{(4)-} \otimes \eta_{(6)-}, \tag{2.1.5}$$

where the $\pm$ are used to denote the chirality of the representation. Dropping the notation for the manifold, we have η_+ and η_- on the internal manifold. Since these are covariantly constant we are free to choose the normalisation $\eta_\pm^\dagger \eta_\pm = 1$. Using the internal manifold gamma matrices γ we define

$$J^m{}_n := i\eta_+^\dagger \gamma^m{}_n \eta_+. \tag{2.1.6}$$

Application of a Fierz identity yields $J^m{}_n J^n{}_p = -\delta^m{}_p$ so that (M_6, J) is an almost complex manifold. Constructing the almost complex structure using the covariantly constant spinor ensures that it is also covariantly constant. Looking back to the discussion of almost complex structure we recall that the Nijenhuis tensor can be reformulated using the covariant derivatives of J (1.1.13). From this we see $\nabla_m J^n{}_p = 0$ implies that the Nijenhuis tensor vanishes. This makes J a complex structure and (M_6, J) a complex manifold.

As M_6 is a complex manifold it admits a Hermitian metric with a natural Hermitian two-form that can be written as $\omega = \frac{1}{2} J_{mn} \mathrm{d}x^m \wedge \mathrm{d}x^n$, where we have used the Hermitian metric to lower an index on J. For a metric connection ($\nabla g = 0$) we have that $\nabla\omega = 0$. We can always choose the Levi-Civita connection which is a metric connection with vanishing torsion. Using the torsion-free property of the Levi-Civita connection, we find that the connection terms in $\nabla\omega$ vanish due to symmetric indices being contracted with antisymmetric indices. Only the partial derivatives remain and we have that $\nabla\omega = \mathrm{d}\omega = 0$. A closed Hermitian form is a Kähler form, meaning that (M_6, ω) is a Kähler manifold.

The covariantly constant spinor also puts restrictions on the holonomy of the internal manifold we now know is Kähler. A general spin manifold (a manifold admitting spinors) with real dimension 6 has a holonomy group $Spin(d) \simeq SU(4)$. A covariantly constant spinor is invariant under parallel transport around a closed curve. Spinors are in the fundamental representation of $SU(4)$ so this invariance is the statement $U\eta = \eta$ for $SU(4)$ matrices U. The existence of a covariantly constant spinor then restricts the holonomy group to some subset of $SU(4)$ that obeys this invariance condition. If we suppose that the spinor η has positive chirality so that it is in the $\mathbf{4}$ of $SU(4)$ then there always exists an $SU(4)$ rotation that puts it in the form

$$\eta_- = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \eta_- \end{pmatrix}. \tag{2.1.7}$$

Using this, the subgroup $SU(3)$ that acts on the first three components is identified as the unbroken group that leaves a $\mathbf{4}$ of $SU(4)$ invariant. This tells us that the holonomy group of a spin manifold with a covariantly constant 6-dimensional spinor is $SU(3) \subseteq SU(4)$.

In summary, we have found that in order to have $\mathcal{N} = 1$ SUSY after compactification the internal manifold M_6 must be Kähler, have complex dimension 3, and have $SU(3)$ holonomy. This satisfies condition (iii) of our Calabi-Yau definition and so describes a Calabi-Yau threefold. This discovery of Candelas, Horowitz, Strominger and Witten was the initial motivation for the study of Calabi-Yau manifolds. We now see that by ignoring Calabi-Yau manifolds with $\subset SU(3)$ holonomy we are guaranteed $\mathcal{N} = 1$ SUSY after compactification.

An important realisation is that topological invariants of the threefold chosen dictates certain properties of the compactified model. For example, the number of generations of particles in the 4-dimensional theory is given by half the Euler number of the threefold. With the topological information of the Calabi-Yau manifold we can know the particle content and gauge group of the 4-dimensional theory. We can also obtain many exact 'topological', meaning they depend on cohomology classes, formulas that give important information such as Yukawa couplings [2]. This makes the study of the Calabi-Yau landscape an integral part model building in string theory. However, with only topological

information we cannot learn about overall normalisations. This means that the coupling constants and particle masses of the 4-dimensional theory are unknown. To deduce them we require the explicit Ricci-flat Kähler metric of the Calabi-Yau manifold. Unfortunately, no such analytic metric has been found. This presents another challenge to our understanding of the landscape of string vacua. In this work, we shall only consider the topological type the Calabi-Yau manifold and ignore the issue of finding an explicit metric.

We have not yet considered the gluino variation (2.1.2). The condition arising from it does not dictate properties of M_6 but for completeness we will briefly outline what it does imply. A more thorough explanation can be found in [23]. On the internal manifold, which we have shown is Calabi-Yau so can make use of complex coordinates z^m, the condition means that the internal Yang-Mills field strengths must obey

$$F_{mn} = F_{\bar{m}\bar{n}} = 0, \qquad g^{m\bar{n}} F_{m\bar{n}} = 0. \tag{2.1.8}$$

The field strengths are derived from some Yang-Mills gauge field that is a connection on some vector bundle E. Assuming that the gauge fields are real, the conditions on the $(1,0)$ and $(0,1)$ parts are equivalent via Hermitian conjugation. It can be shown that a solution to $F_{mn} = 0$ is a holomorphic vector bundle. The condition on the $(1,1)$ part is known as the Hermitian Yang-Mills equation. The choice of a connection on a holomorphic vector bundle that is a solution to the Hermitian Yang-Mills equation is something that needs to specified before the compactification. This forms a region in the landscape of string vacua distinct from the Calabi-Yau landscape.

It should also be noted that Calabi-Yau manifolds appear in string theory not only as threefolds but also as fourfolds via F-theory. Introduced by Vafa in 1996 [51], F-theory is a promising framework for understanding string theory that offers a non-perturbative way to compactify type IIB superstring theories with a non-constant dilaton. It is not feasible to give an overview of F-theory on Calabi-Yau fourfolds here but detailed reviews can be found in [55] and [5]. The important result is that F-theory compactifications with an internal manifold that is an elliptically fibred fourfold can lead to 4-dimensional theories with $\mathcal{N} = 1$ SUSY. An elliptic fibration is a bundle of elliptic curves and for a Calabi-Yau manifold to be elliptically fibred is common. So for the same reasons as for the

heterotic string compactification, it is important to understand the landscape of Calabi-Yau fourfolds. For fourfolds it must be determined whether the manifold is elliptically fibred in addition to computing their topological invariants.

2.2 Holomorphic $(m,0)$-form

Understanding condition (iv) of Definition 2.0.2 is important as a guaranteed nowhere vanishing holomorphic $(m,0)$-form is a useful tool for examining Calabi-Yau manifolds. We will start by assuming that the holomorphic $(m,0)$-form exists

Theorem 2.2.1. *A compact $2m$-dimensional Kähler manifold M with a nowhere vanishing holomorphic $(m,0)$-form Ω is Calabi-Yau.*

Proof. Any $(m,0)$-form $\Omega \in \Omega^{(m,0)}(M)$ must be proportional to the permutation symbol $\varepsilon_{\mu_1\ldots\mu_m}$. Since the $(m,0)$-form needs to be holomorphic it must be

$$\Omega_{\mu_1\ldots\mu_m} = f(z)\varepsilon_{\mu_1\ldots\mu_m}, \tag{2.2.1}$$

where $f(z)$ is a nowhere vanishing holomorphic function. Using the Hermitian metric on the Kähler manifold, we define a globally well-defined coordinate scalar from Ω

$$||\Omega||^2 = \frac{1}{m!}\Omega_{\mu_1\ldots\mu_m} g^{\mu_1\bar{\nu}_1}\ldots g^{\mu_m\bar{\nu}_m}\bar{\Omega}_{\bar{\nu}_1\ldots\bar{\nu}_m} \overset{(2.2.1)}{=} \frac{|f|^2}{\sqrt{g}}. \tag{2.2.2}$$

Rearranging this for $\sqrt{g}$ in terms of $||\Omega||^2$ we see that the Ricci 2-form (1.2.16) is given by

$$\mathcal{R} = i\partial\bar{\partial}\log||\Omega||^2. \tag{2.2.3}$$

Since $\log||\Omega||^2$ is a globally defined coordinate scalar, the identity $\partial\bar{\partial} = (1/2)\,\mathrm{d}\big(\partial - \bar{\partial}\big)$ implies that $\mathcal{R}$ is an exact form, meaning $c_1(M) \propto [\mathcal{R}] = 0$. □

We will now show that the $(m,0)$-form is unique (up to a constant). Suppose there are two globally defined holomorphic $(m,0)$-forms Ω and Ω'. They must both be proportional to the permutation symbol and so are related by

$$\Omega' = h(z)\Omega, \tag{2.2.4}$$

where $h(z)$ is a globally defined holomorphic function. The maximum modulus principle implies that a globally defined holomorphic function is a constant. Thus, $h(z)$ is constant so that Ω is indeed unique up to a constant.

It will be useful to demonstrate that Ω is harmonic. To do this we must confirm $\mathrm{d}\Omega = \mathrm{d}^\dagger\Omega = 0$. We quickly see $\mathrm{d}\Omega = 0$ by using that Ω is an $(m,0)$-form, $\partial\Omega = 0$, and that it is holomorphic, $\bar{\partial}\Omega = 0$. To see $\mathrm{d}^\dagger\Omega = 0$ we need to work in local coordinates and evaluate explicitly

$$\begin{aligned}\mathrm{d}^\dagger\Omega &= -\frac{1}{(m-1)!}\nabla^\nu\Omega_{\nu\mu_1\ldots\mu_{m-1}}\mathrm{d}z^{\mu_1}\wedge\ldots\wedge\mathrm{d}z^{\mu_{m-1}} \\ &= -\frac{1}{(m-1)!}g^{\nu\bar{\rho}}\nabla_{\bar{\rho}}\Omega_{\nu\mu_1\ldots\mu_{m-1}}\mathrm{d}z^{\mu_1}\wedge\ldots\wedge\mathrm{d}z^{\mu_{m-1}} \\ &= -\frac{1}{(m-1)!}g^{\nu\bar{\rho}}\partial_{\bar{\rho}}\Omega_{\nu\mu_1\ldots\mu_{m-1}}\mathrm{d}z^{\mu_1}\wedge\ldots\wedge\mathrm{d}z^{\mu_{m-1}} = 0.\end{aligned} \tag{2.2.5}$$

To get to the last line recall that the Chern connection of a Kähler manifold is pure in its indices so that it vanishes in the anti-holomorphic covariant derivative of Ω.

2.3 Hodge Numbers

The Calabi-Yau manifold inherits the structure of the Kähler manifold, meaning that the relations between Hodge numbers introduced in § 1.2.4 apply to the Calabi-Yau manifold also. As we have just discussed, the Calabi-Yau condition creates some additional structure on the Kähler manifold. We will see that this additional structure restricts the number of independent Hodge numbers even further.

A Calabi-Yau manifold admits a unique holomorphic $(m,0)$-form Ω that we proved to be harmonic. This implies $h^{m,0} \geq 1$. We know that Ω is unique but we have not ruled out the existence of some non-holomorphic harmonic $(m,0)$-form Ω'. The permutation symbol can be used to write Ω' as

$$\Omega' = f(z,\bar{z})\varepsilon. \tag{2.3.1}$$

An $(m,0)$-form is trivially $\bar{\partial}$-closed so to be harmonic we must confirm only that it is ∂-closed. Requiring that Ω' is ∂-closed gives $\partial f(z,\bar{z}) = 0$ which means that f must actually be holomorphic. But we have already seen that the holomorphic $(m,0)$-form is unique, so we find that Ω' is not permitted and $h^{m,0} = 1$.

Similar to how the Hodge star causes the relation $h^{r,s} = h^{m-r,m-s}$, the holomorphic $(m,0)$-form of the Calabi-Yau manifold results in the holomorphic duality $h^{r,0} = h^{m-r,0}$. To show this we start with some $\omega \in \Omega^{(k,0)}$ for $0 \leq k \leq m$ and define the 'holomorphic Hodge dual' by

$$v_{\bar{\mu}_1 \dots \bar{\mu}_{m-k}} = \frac{1}{k!} \bar{\Omega}_{\bar{\mu}_1 \dots \bar{\mu}_{m-k} \bar{\mu}_{m-k+1} \dots \bar{\mu}_m} \omega^{\bar{\mu}_{m-k+1} \dots \bar{\mu}_m}. \tag{2.3.2}$$

Using that Ω is harmonic gives

$$\nabla^{\bar{\mu}_1} v_{\bar{\mu}_1 \dots \bar{\mu}_{m-k}} = \frac{1}{k!} \bar{\Omega}_{\bar{\mu}_1 \dots \bar{\mu}_{m-k} \bar{\mu}_{m-k+1} \dots \bar{\mu}_m} \nabla^{\bar{\mu}_1} \omega^{\bar{\mu}_{m-k+1} \dots \bar{\mu}_m}. \tag{2.3.3}$$

Next we invert (2.3.2) and make use of the coordinate scalar defined previously (2.2.2)

$$\omega^{\bar{\mu}_{m-k+1} \dots \bar{\mu}_m} = \frac{m!}{(m-k)!k!} \frac{\Omega_{\bar{\mu}_1 \dots \bar{\mu}_{m-k} \bar{\mu}_{m-k+1} \dots \bar{\mu}_m}}{||\Omega||^2} v_{\bar{\mu}_1 \dots \bar{\mu}_{m-k}} \tag{2.3.4}$$

$$\implies \nabla_{\bar{\mu}_m} \omega^{\bar{\mu}_{m-k+1} \dots \bar{\mu}_m} = \frac{m!}{(m-k)!k!} \frac{\Omega_{\bar{\mu}_1 \dots \bar{\mu}_{m-k} \bar{\mu}_{m-k+1} \dots \bar{\mu}_m}}{||\Omega||^2} \nabla_{\bar{\mu}_m} v_{\bar{\mu}_1 \dots \bar{\mu}_{m-k}}. \tag{2.3.5}$$

Writing Ω as in (2.2.1) it can be shown that $\frac{\bar{\Omega}}{||\Omega||^2}$ is holomorphic and non-singular. This then means that from (2.3.3) and (2.3.5) we can conclude that the $(0, m-k)$-form v is $\bar{\partial}$-harmonic if and only if the $(k,0)$-form ω is ∂-harmonic. For a Kähler manifold this implies that v is harmonic if and only if ω is harmonic. So given a harmonic $(k,0)$-form $\omega \in H_{\bar{\partial}}^{(k,0)}(M)$, there exists a unique harmonic $(m-k,0)$-form $v \in H_{\bar{\partial}}^{(m-k,0)}(M)$ giving the holomorphic duality $h^{k,0} = h^{m-k,0}$.

Further restriction to the Hodge numbers of a Calabi-Yau manifolds with exactly $SU(m)$ holonomy can be found by examining the harmonic representatives for the Dolbeault cohomology groups $H_{\bar{\partial}}^{(r,0)}(M)$. The action of the Laplacian on the components of a k-form is

$$\Delta \omega_{a_1 \dots a_k} = -\nabla^b \nabla_b \omega_{a_1 \dots a_k} - k R_{b[a_1} \omega^b{}_{a_2 \dots a_k]} - \frac{1}{2} k(k-1) R_{bc[a_1 a_2} \omega^{bc}{}_{a_3 \dots a_k]}, \tag{2.3.6}$$

where the curvature tensors come from the relation

$$[\nabla_{a_1}, \nabla_b] \omega^b{}_{a_2 \dots a_k} = -R_{b a_1} \omega^b{}_{a_2 \dots a_k} - \omega^b{}_{c[a_3 \dots a_k} R^c{}_{a_2] a_1 b}. \tag{2.3.7}$$

Recalling that for Kähler manifolds $\Delta = 2\Delta_{\bar{\partial}}$, examining $H_{\bar{\partial}}^{(r,0)}(M)$ requires that we consider the Laplacian Δ acting on an $(r,0)$-form. Using that for a Ricci-flat Kähler manifold $R_{\mu\bar{\nu}} = R_{\mu\nu\bar{\rho}\bar{\sigma}} = 0$, the condition that an $(r,0)$-form is harmonic simply reads

$\nabla^a \nabla_a \omega_{\mu_1 \dots \mu_r} = 0$. For a compact manifold we can contract this with $\omega^{\mu_1 \dots \mu_r}$ and integrate over the manifold

$$0 = \int_M \sqrt{g} \omega^{\mu_1 \dots \mu_r} \nabla^a \nabla_a \omega_{\mu_1 \dots \mu_r} = - \int_M \sqrt{g} \nabla^a \omega^{\mu_1 \dots \mu_r} \nabla_a \omega_{\mu_1 \dots \mu_r}. \tag{2.3.8}$$

Since the integrand is positive definite we must have

$$\nabla_\nu \omega_{\mu_1 \dots \mu_r} = 0, \qquad \nabla_{\bar{\nu}} \omega_{\mu_1 \dots \mu_r} = \partial_{\bar{\nu}} \omega_{\mu_1 \dots \mu_r} = 0. \tag{2.3.9}$$

These equations tell us that under parallel transport $\omega_{\mu_1 \dots \mu_r}$ transforms as a singlet of the holonomy group $SU(m)$. We know that $\omega_{\mu_1 \dots \mu_r}$ is in the $\bigwedge^r \mathbf{m}$ of $SU(m)$ so we require that the singlet appears in the decomposition. This only happens for the trivial cases of $r = 0$ and $r = m$. This contradiction means the harmonic representative cannot exist for the cohomology groups $H^{(r,0)}_{\bar{\partial}}(M)$ with $0 < r < m$. From this we see that $h^{r,0} = 0$ for $0 < r < m$.

Applying the holomorphic duality along with the Kähler manifold Hodge number relations (1.2.45) and (1.2.46) to the fixed Hodge numbers of a Calabi-Yau fourfold leaves only four independent Hodge numbers known as the Hodge data of the fourfold. Organised in a Hodge diamond, these are

$$\begin{array}{ccccccc}
 & & & 1 & & & \\
 & & 0 & & 0 & & \\
 & 0 & & h^{1,1} & & 0 & \\
0 & & h^{2,1} & & h^{2,1} & & 0 \\
1 & & h^{3,1} & & h^{2,2} & & h^{3,1} & & 1. \\
0 & & h^{2,1} & & h^{2,1} & & 0 \\
 & 0 & & h^{1,1} & & 0 & \\
 & & 0 & & 0 & & \\
 & & & 1 & & &
\end{array}$$

The Euler characteristic of a fourfold is related to the Hodge numbers by the expression

$$\chi = \sum_{k=0}^{8} (-1)^k b^k = 4 + 2h^{1,1} - 4h^{2,1} + 2h^{3,1} + h^{2,2}. \tag{2.3.10}$$

There also exists a relation between the Hodge numbers of a fourfold that reduces the number of independent Hodge numbers to three. The origin of this relation is beyond the

scope of this work so we will only present an outline. The holomorphic Euler characteristic of a holomorphic vector bundle E on a compact complex manifold M is

$$\chi(M,E) = \sum_{k=0}^{\dim_{\mathbb{C}} M} (-1)^k \dim H^k(M,E). \tag{2.3.11}$$

Defining $\chi_q := \chi(M, \bigwedge^q T^*M)$ and recognising $H^r(M, \bigwedge^s T^*M) = H^{(r,s)}(M)$ gives that for fourfolds

$$\chi_s = \sum_{r=0}^{4} (-1)^r h^{r.s}(M). \tag{2.3.12}$$

We then make use of the Hirzebruch-Riemann-Roch index theorem [30]

$$\chi_s = \int_M \mathrm{ch}\big(\bigwedge^s T^*M\big) \wedge \mathrm{Td}(TM), \tag{2.3.13}$$

where $\mathrm{Td}(TM)$ is the Todd class [44] of the tangent bundle of M. Both (2.3.12) and (2.3.13) can be used to compute the first three holomorphic Euler numbers of the fourfold resulting in [30, 37]

$$\chi_0 = 2 = \frac{1}{720}\int_M (3c_2^2 - c_4), \tag{2.3.14}$$

$$\chi_1 = -h^{1,1} + h^{2,1} - h^{3,1} = \frac{1}{180}\int_M (3c_2^2 - 31c_4), \tag{2.3.15}$$

$$\chi_2 = -2h^{2,1} + h^{2,2} = \frac{1}{360}\int_M (9c_2^2 + 237c_4), \tag{2.3.16}$$

where we have used that $c_1 = 0$ and made use of Hodge number relations. The calculation using the Hirzebruch-Riemann-Roch theorem reveals the relation $22\chi_0 - 4\chi_1 - \chi_2 = 0$ which implies the non-trivial relation between Hodge numbers

$$-4h^{1,1} + 2h^{2,1} - 4h^{3,1} + h^{2,2} = 44. \tag{2.3.17}$$

We now understand the Hodge numbers for CY_4. We can fully specify the Hodge data of a fourfold using only, say, $(h^{1,1}, h^{1,2}, h^{1,3})$ as the independent Hodge numbers with an Euler number given by the expression $\chi = 6(8 + h^{1,1} + h^{3,1} - h^{2,1})$.

3 The Calabi-Yau Landscape

The vital role Calabi-Yau manifolds play in dictating the physics of compactified theories motivates the exploration of the landscape. This is done by choosing a method of construction and exhausting all the Calabi-Yau manifolds it can yield. The result is a list of manifolds labelled by their topological invariants. We will focus on the earliest systematic method of construction developed in the original papers [15, 13, 32, 25] known as CICYs. We focus on CICYs since they are well established and so have complete data sets readily available. These data sets are of a size amicable to machine learning but small enough that they can be handled with a laptop. In addition to this, the construction is relatively simple allowing for a compact explanation. We will also give a brief outline of an additional construction to further our understanding of the Calabi-Yau landscape. Once the landscape is familiar, we will then be in a position to motivate a machine learning approach.

3.1 The Complete Intersection Calabi-Yau

A straight forward way to construct Calabi-Yau manifolds is to consider submanifolds of Kähler manifolds that obey the Calabi-Yau condition. This is because the induced metric of a submanifold of a Kähler manifold is also Kähler. As $\mathbb{C}^N$ is Kähler, it is an obvious starting point. However, it is known that none of the submanifolds of $\mathbb{C}^N$ are compact. We instead turn to submanifolds of complex projective space which are automatically Kähler and compact.

There is a theorem by Chow [17] that allows for a natural way to construct submanifolds of $\mathbb{CP}^N$ algorithmically. It states

Theorem 3.1.1 (**Chow**). *Any analytic subvariety of a projective space is algebraic.*

What this means for our purposes is that the analytic submanifolds of complex projective

spaces may be considered as the zero locus of a finite number of polynomials in the homogeneous coordinates z^μ. We will look at submanifolds that are complete intersections of l polynomials p^i. Here complete intersection means that on the submanifold the l-form $\mathrm{d}p^1 \wedge \ldots \wedge \mathrm{d}p^l$ does not vanish. This ensures that the hypersurface described by a complete intersection of polynomials is smooth. A complete intersection manifold is defined as the manifold embedded by the complete intersection of the hypersurfaces described by the polynomials.

We need to understand how the restriction that the complete intersection manifold is Calabi-Yau translates to the polynomials. The more traditional way of doing this is to attempt to find the globally defined and nowhere vanishing holomorphic $(m, 0)$-form whose existence defines a Calabi-Yau manifold. We will instead take the approach of explicitly computing the first Chern class of the submanifold which can be done quickly and neatly by making use of a few tricks. This is the approach taken in Hübsch's bestiary [33]. To understand the more traditional method the reader is referred to Candelas [15].

We first consider the zero locus of a single polynomial p that is homogeneous with degree d in the homogeneous coordinates of $\mathbb{CP}^N$. This defines a hypersurface

$$\mathcal{M} = \{z \in \mathbb{CP}^N | p(z) = 0\} \tag{3.1.1}$$

of dimension $N - 1$. Recall that sections of the hyperplane line bundle $\mathscr{O}(1)$ are the homogeneous coordinates of $\mathbb{CP}^N$. It follows that p is a section of $\mathscr{O}(d)$.

Let Γ_a be a connection on $\mathscr{O}(d)$ and $X^a\nabla_a$ represent a tangent vector at a point of $\mathbb{CP}^N$. There is then a map from $T\mathbb{CP}^N$ to $\mathscr{O}(d)$ via

$$X^a\nabla_a p = X^a(\partial_a p + \Gamma_a p) := p_X. \tag{3.1.2}$$

If we restrict ourselves to points on the hypersurface $\mathcal{M}$ only, then the mapping is holomorphic and independent of the choice of connection since $p = 0$ on $\mathcal{M}$. When X is tangential to $\mathcal{M}$ we have $X^a\partial_a p = 0$ along $\mathcal{M}$ since p vanishes everywhere on $\mathcal{M}$ and so has no gradient along $\mathcal{M}$. Thus, we have that (3.1.2) maps vectors of the $T\mathcal{M}$ subbundle of $T\mathbb{CP}^N$ to zero i.e. $\ker(\nabla p) = T\mathcal{M}$. We can use p as a local coordinate in the embedding space around $\mathcal{M}$ since it vanishes at $\mathcal{M}$ and is nonzero away from $\mathcal{M}$. Rescaling $X' = \lambda X$ allows the map (3.1.2) to cover the range of these local coordinates

so that locally it maps onto $\mathscr{O}(d)$. The map is also global since it is holomorphic and $\text{rank}\mathscr{O}(d) = \dim T\mathbb{CP}^N - \dim T\mathcal{M}$.

The relevance of this discussion is revealed when we summarise it with the short exact sequence

$$0 \longrightarrow T\mathcal{M} \xrightarrow{\text{id}} T\mathbb{CP}^N|_{\mathcal{M}} \xrightarrow{\nabla p} \mathscr{O}(d)|_{\mathcal{M}} \longrightarrow 0. \tag{3.1.3}$$

This allows the Chern class of $\mathcal{M}$ to be computed using $c(\mathcal{M}) = c(\mathbb{CP}^N)/c(\mathscr{O}(d))$. Recall from § 1.4.1 that $c(\mathscr{O}(1)) = 1+x$. Its Chern character is then simply $\text{ch}(\mathscr{O}(1)) = e^x$. Using the properties of the Chern character (1.4.6) we find $\text{ch}(\mathscr{O}(d)) = e^{dx}$. By comparison ($x$ is a closed 2-form and the Chern forms are closed $2k$-forms) with the explicit expansion of the Chern character (1.4.7) we see

$$c(\mathscr{O}(d)) = 1 + dx. \tag{3.1.4}$$

As we have already computed $c(\mathbb{CP}^N)$, we now have an expression for the Chern class of $\mathcal{M}$

$$c(\mathcal{M}) = \frac{(1+x)^{N+1}}{1+dx}. \tag{3.1.5}$$

Expansion of this expression as a power series in x gives the first Chern class

$$c_1(\mathcal{M}) = \big((N+1) - d\big)x. \tag{3.1.6}$$

The Calabi-Yau condition is then realised as a constraint on the degree of the polynomial that $d = N + 1$. We see immediately that the zero locus of a degree 6 polynomial in $\mathbb{CP}^5$ describes a Calabi-Yau fourfold.

The next step is to extend this treatment to the complete intersection of K polynomials of degree d_a, $a = 1, \ldots, K$ in $\mathbb{CP}^N$ that define a hypersurface $\mathcal{M}$ of dimension $N - K$. The relevant line bundle is now $\bigoplus_{a=1}^{K} \mathscr{O}(d_a)$. The argument used to reach the short exact sequence (3.1.3) remains valid when generalised to $\bigoplus_{a=1}^{K} \mathscr{O}(d_a)$. Using this and the identity for Chern classes of direct sum bundles (1.4.12), the Chern class of $\mathcal{M}$ generalises to

$$c(\mathcal{M}) = \frac{(1+x)^{N+1}}{\prod_{a=1}^{K}(1+d_a x)}. \tag{3.1.7}$$

From which the vanishing first Chern class condition is realised on the polynomials as

$$\sum_{a=1}^{K} d_i = N + 1. \tag{3.1.8}$$

At first glance it seems that we can find at least one Calabi-Yau manifold at every N. However, we must consider what it means for one of the polynomials to have a degree of homogeneity equal to one. A degree one polynomial after a coordinate change can be written in the coordinates of $\mathbb{CP}^N$ as $z^{N+1} = 0$. This reduces the manifold to $\mathbb{CP}^{N-1}$ meaning no new Calabi-Yau manifolds can be found using degree one polynomials. With the restriction $d_a > 1$ we find the new fourfolds

$$\begin{gathered}[6\|2\,5], \quad [6\|3\,4], \\ [7\|2\,2\,4], \quad [7\|2\,3\,3], \\ [8\|2\,2\,2\,3], \\ [9\|2\,2\,2\,2\,2].\end{gathered} \tag{3.1.9}$$

We have adopted the notation that the number to the left of the double vertical line is the dimension of the projective space and the numbers to the right are the degrees of the polynomials e.g. the zero locus of a single degree 6 polynomial in $\mathbb{CP}^5$ is written $[5||6]$. Note that the condition (3.1.8) means the number to the left of the double vertical line can be calculated from the numbers to the right so may be omitted without ambiguity.

For the most general treatment of CICYs we must consider an embedding space that is a product of projective spaces. This general embedding space is denoted

$$\mathcal{X} = \mathbb{CP}^{n_1}_1 \times \ldots \times \mathbb{CP}^{n_m}_m. \tag{3.1.10}$$

The hypersurfaces that define the complete intersection manifold $\mathcal{M}$ are themselves defined by polynomials p^a, $a = 1, \ldots, K$ with degree of homogeneity d^r_a with respect to the homogeneous coordinates of each $\mathbb{CP}^{n_r}_r$ in the product $\mathcal{X}$. The resulting complete intersection manifold has dimension $\sum_{r=1}^{m} n_r - K$. We can extend the notation introduced in (3.1.9) to describe complete intersection manifolds in products of projective space via configuration matrices

$$\left[\begin{array}{c|ccc} n_1 & d^1_1 & \cdots & d^1_K \\ \vdots & \vdots & \ddots & \vdots \\ n_m & d^m_1 & \cdots & d^m_K \end{array}\right]. \tag{3.1.11}$$

The column to the left of the double vertical line are the dimensions of the projective spaces in $\mathcal{X}$. Each column to the right of the double vertical line is the degree of a polynomial in the coordinates of the different projective spaces that make up $\mathcal{X}$. We will see that the condition imposed by the vanishing first Chern class means the the column to the left of the double vertical line can be omitted without ambiguity. For this reason, when working with the configuration matrix we consider the polynomial degrees only.

Each polynomial p^a is now a section of the line bundle $\bigotimes_{r=1}^{m} \mathscr{O}_r(d_a^r)$ where the subscript on the hyperplane line bundle denotes that it is over the space $\mathbb{CP}_r^{n_r}$. The short exact sequence (3.1.3) remains valid when generalised to $\bigoplus_{a=1}^{K} \left(\bigotimes_{r=1}^{m} \mathscr{O}_r(d_a^r) \right) := E$ so that the Chern class of $\mathcal{M}$ obeys $c(\mathcal{M}) = c(\mathcal{X})/c(E)$. Using (1.4.12) and (1.4.13) and noting that tensor products of line bundles are line bundles, we get

$$\begin{aligned} c(E) &= \prod_{a=1}^{K} c\Big(\bigotimes_{r=1}^{m} \mathscr{O}_r(d_a^r) \Big) \\ &= \prod_{a=1}^{K} \Big(1 + c_1\Big(\bigotimes_{r=1}^{m} \mathscr{O}_r(d_a^r) \Big) \Big) \\ &= \prod_{a=1}^{K} \Big(1 + \sum_{r=1}^{m} d_a^r x_r \Big). \end{aligned} \tag{3.1.12}$$

To calculate $c(\mathcal{X})$ we use that the tangent bundle of a product manifold is the direct sum of the tangent bundles of each manifold in the product. Thus, $c(\mathcal{X})$ is simply the product of the Chern classes of each $\mathbb{CP}_r^{n_r}$ in $\mathcal{X}$ via the application of (1.4.12). Therefore, the Chern class of $\mathcal{M}$ is

$$c(\mathcal{M}) = \frac{\prod_{r=1}^{m} (1 + x_r)^{n_r+1}}{\prod_{a=1}^{K} \Big(1 + \sum_{s=1}^{m} d_a^r x_s \Big)}. \tag{3.1.13}$$

Which if we expand to get the first Chern class and then apply the Calabi-Yau condition we have the restriction on the polynomials

$$\sum_{a=1}^{K} d_a^r = n_r + 1, \quad \forall r = 1, \ldots, m. \tag{3.1.14}$$

In the search for fourfolds there is also the additional constraint that the complete intersection manifold has complex dimension 4 which manifests as

$$\sum_{r=1}^{m} n_r = K + 4. \tag{3.1.15}$$

It should be noted that a configuration matrix obeying the Calabi-Yau condition does not specify a particular manifold but rather a family of all the possible complete intersections parameterised by the space of coefficients of the polynomials. A generic choice of coefficients would define a CICY that we say is a member of the configuration defined by the configuration matrix. An important proof [25] tells us that each configuration contains one deformation class of Calabi-Yau manifolds. This is where the strength of the configuration matrix notation lies; key properties of a CICY such as the Hodge numbers depend only on the configuration matrix and not on the specific choice of polynomial coefficients. For this reason, we make no distinction between a configuration and its members.

CICYs offer a method of generating fourfolds by finding the possible configuration matrices that satisfy (3.1.14) and (3.1.15). An important point is that it is possible for multiple configuration matrices to describe the same manifold. For this reason, we must be aware of equivalences when building landscape data. The most restrictive equivalence is regarding degree one polynomials and demands that

$$\sum_{r=1}^{m} d_a^r > 1, \quad a = 1, \ldots, K \tag{3.1.16}$$

in order to avoid generating redundant manifolds. An obvious equivalence is between configuration matrices that differ only by permutations of rows or columns as this corresponds to a reordering of ambient spaces and polynomials. There are numerous other equivalences that should be considered to avoid needless effort. Some of these are understood well and others are empirical observations. These are described thoroughly in [15].

It is fairly straightforward to compute an expression for the Euler number of a generic CICY directly from the configuration matrix. The same can be done for the Chern classes and for an additional topological invariant called the intersection number [12]. To completely classify a Calabi-Yau manifold the Hodge numbers are needed also, these can occasionally be calculated with relative ease due to their relation to the manifold's complex structure. However, there is no way to read the Hodge numbers directly from the configuration matrix. Their calculation employs spectral sequences and requires computing line bundle cohomologies which is computationally expensive when done on a large scale. These calculations are well documented in [33].

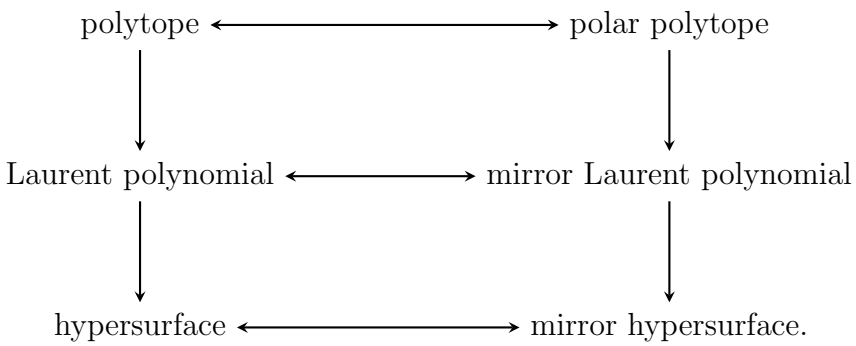
3.2 Why Machine Learning?

In the search for realistic compactifications it is important that we can easily traverse the Calabi-Yau landscape. What this means is having large data sets of Calabi-Yau manifolds with their topological properties computed to allow for identification of manifolds that lead to compactified theories with desirable properties. The benefits of the machine learning approach becomes apparent when we consider the size of the landscape.

In complex dimension 1 the 2-torus T^2 is the only compact Calabi-Yau manifold. Going to complex dimension 2 we have the only compact Calabi-Yau manifolds being the K3-surface (Chapter 8 of [4]) and the 4-torus T^4. At complex dimension 3 the Calabi-Yau manifold becomes important for building physical models in string theory, it should therefore not be surprising that the landscape is dramatically more complicated than for complex dimension 1 and 2. The number of inequivalent Calabi-Yau threefolds is in fact unknown. The same is true for complex dimension 4 and above. It is conjectured that the number of topologically distinct Calabi-Yau manifolds in each dimension is finite [57] but the size of the landscape is still very much an open question. We will briefly examine two important data sets of Calabi-Yau three and fourfolds. Other data sets exist and are covered in [27].

CICYs offer a glimpse into the vast Calabi-Yau landscape. The first complete data set was for the threefolds and consists of 7,890 configuration matrices [15]. The inequivalent configurations were classified and the Hodge numbers computed [24] revealing 266 distinct Hodge pairs (threefolds have two independent Hodge numbers). The data set of CICY4s is substantially larger, first computed in [21] the comprehensive list numbers 921,497 configuration matrices. The Hodge numbers were then computed in [22] revealing at least 36,779 topologically distinct manifolds.

The portion of the landscape the CICYs makeup is unknown but the existence of another method for generating Calabi-Yau manifolds implies that it is a small portion. A method generalised by Batyrev-Borisov in [3] allows mirror pairs of Calabi-Yau hypersurface families to be generated from reflexive polytopes. A good introduction to this is [18]. A crude understanding can be gleaned from the diagram



Starting with a reflexive polytope of dimension n and following the above steps will yield a complex dimension $n-1$ Calabi-Yau mirror pair. By mirror pair we mean Calabi-Yau manifolds that are dual to eachother via mirror symmetry, a reflection along the diagonal of the Hodge diamond. The construction of Calabi-Yau manifolds then becomes an exercise in finding reflexive polytopes.

The $n = 4$ case was first classified by Kreuzer and Sarke in [39] resulting in a list of 473,800,776 reflexive polytopes. From this list, 30,108 Calabi-Yau threefolds with distinct Hodge pairs were found. The number of distinct threefolds is in fact much greater than this because determining inequivalence requires additional topological data and depends on how singularities in the polytopes are dealt with [38]. Recently the $n = 5$ case was classified in [48] resulting in a staggering 185,269,499,015 reflexive polytopes giving 532,600,483 distinct sets of Hodge numbers after a computation taking 57,321 core hours.

The size of the Calabi-Yau landscape makes it fertile ground for the application of modern machine learning techniques. It is important that efforts are made to establish the potential of such an application to further our understanding of the Calabi-Yau landscape and string theory in general. The implementation of machine learning to the Calabi-Yau landscape can be done with two goals in mind. The first, is to evaluate its effectiveness as a predictor for incomplete data sets in string theory. If applied to the landscape of string vacua it could aide in targeted searches for vacua that resemble our universe. The CICY data sets being complete means that they can be used to establish how well machine learning techniques perform in this capacity.

We can also consider implementing machine learning to complete data sets to probe for unseen patterns. If, for example, we are able to train a NN on a complete CICY data set so that it can predict certain properties of the manifolds near perfectly in a fraction

of the time it takes conventionally, this may point to the machine using a simplification in the data that has been overlooked. For CICYs, this may lead to a conjecture relating the Hodge numbers to the configuration matrix. Machine learning techniques have been demonstrated to be capable of making analytic discoveries, an example of this is [45] where they have been used to make conjectures, some of which have since been proved, involving fundamental constants. The possibility that similar success will be found upon application to the Calabi-Yau landscape warrants further study.

4 Rudiments of Machine Learning

With a good understanding of the Calabi-Yau manifold and its complete intersection construction, we are almost in a position to apply machine learning to the CICY4 data. All that is left is to understand what is meant by machine learning. As machine learning is a vast field with many techniques and algorithms, we will focus on a branch of techniques known as artificial neural networks. NNs can be used in both regression and classification problems, understanding how a NN can be applied as a regressor is the purpose of this section. In addition to this, we shall only consider supervised learning which uses known input-output pairs for training. References for this section are [47, 6, 34].

It is important to note that much in the field of NNs, and machine learning in general, is not understood at a fundamental level. Precisely why a certain technique works is often unknown since the exact processes in the NN are obscure. Instead, we must use a heuristic understanding to offer explanations. The techniques presented in this section are well established in the field and have widely accepted explanations, but it should be kept in mind that much is left to understand. Research into fundamental explanations and a unified theory in machine learning is still very much underway [40, 42].

It is useful to first understand the basic problem a NN is designed for and how it approaches it. A NN is used in regression problems when there is a large but incomplete data set of input-output pairs which we shall label $\{\mathbf{x}_i, y_i\}$. What is meant by a large data set depends on the nature of the data, for example, if the input is high dimensional a larger number of them will be needed to make a NN viable. The available input-output pairs are used to train the NN so that it can recognise features of the data and learn how the outputs are related to the inputs. The goal of the training process is for the NN to adjust its parameters so that it can learn generalities of the data. If it has been successful in this the NN will be able to predict the output of unseen inputs by recognising the generalities it has learnt. A NN can be thought of as performing a nonlinear regression

of some complicated function that maps the inputs to the outputs. It is best suited for problems of larger dimensions where more conventional regression methods become untenable.

4.1 Neural Networks

NNs are a common machine learning technique modelled after their biological namesake. An input signal passes through connected layers activating a certain combination of neurons to produce a final output signal. For a regression problem a NN simply takes some n-dimensional input and gives a number $f : \mathbb{R}^n \to \mathbb{R}$.

Remark. In what followings there is no longer a notion of complex coordinates so we will repurpose the index notation introduced in § 1. We will use Latin indices to denote elements of a set and Greek indices to denote components of a vector or matrix.

To understand the whole network we must first look at a single neuron. A neuron is defined by an activation function σ, a weight vector w_i, and a bias b. It takes an input vector x_μ and produces the output $\sigma(w_\mu x_\mu + b) := \hat{y}$, where we make use of Einstein summation notation. This is best understood with the diagram in Figure 4.1.

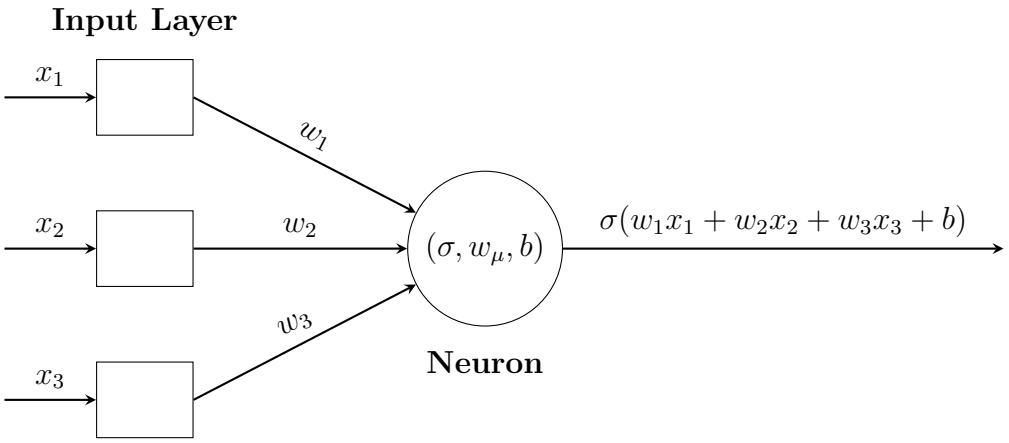


Figure 4.1: Diagram of a single neuron with 3 inputs. The input layer is used to feed the inputs into the neuron.

The activation function is used to determine how much a signal 'activates' the neuron. Many functions have seen use as activation functions, some common choices are

$$\begin{aligned}
&\text{Identity} && \sigma : \mathbb{R} \to \mathbb{R} && \sigma(x) = x, \\
&\text{Sigmoid} && \sigma : \mathbb{R} \to (0,1) && \sigma(x) = (1+e^{-x})^{-1}, \\
&\text{Tanh} && \sigma : \mathbb{R} \to (-1,1) && \sigma(x) = \tanh x, \\
&\text{ReLU} && \sigma : \mathbb{R} \to \mathbb{R}_0^+ && \sigma(x) = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases}.
\end{aligned} \tag{4.1.1}$$

The correct choice of activation function is not always obvious but some are better suited to certain types of data and whether the network is used for classification or regression. The weights dictate the weighting of each signal in the sum that gets passed to the activation function for output. These parameters are adjusted in the training process to allow the neuron to recognise features in the data. The bias is another parameter that is varied in the training process. It is used to offset the weighted sum and ensure that the signal that passes to the activation function remains in the 'active' region. This is best understood by considering the sigmoid function. Having a bias prevents inputs to the sigmoid that are too large from getting stuck outputting very close to either 0 or 1.

We are now ready to start building a network by connecting collections of neurons into layers. A layer is a group of neurons all with the same activation function and all connected to the same inputs. If there are k neurons in a layer then there are k outputs. When neurons are arranged like this, all the weight vectors w_μ of the neurons in the layer can be joined into a single weight matrix $W_{\mu\nu}$. The reason for doing this is that the output for the μ-th neuron can be written compactly as

$$l_\mu := \sigma(W_{\mu\nu} x_\nu + b_\mu). \tag{4.1.2}$$

Taking multiple layers and connecting them, by which we mean making the output of all the neurons in one layer the input for each neuron in the next layer, creates a fully-connected NN. Each layer has its own weight matrix, bias vector, and activation function. The output of the μ-th neuron in the i-th layer $l^{(i)}$ of a NN with s total layers is denoted $l_\mu^{(i)}$ so that $l_\mu^{(0)} = x_\mu$ and $l_\mu^{(s)} = \hat{y}_\mu$. We then have for the outputs at each neuron

$$l_\mu^{(i)} = \sigma^{(i)}(W_{\mu\nu}^{(i)} l_\nu^{(i-1)} + b_\mu^{(i)}). \tag{4.1.3}$$

When building a NN by connecting layers in this way we must always include an input and output layer. The input layer simply feeds the input data into the network. If we consider the constituents of this layer as neurons each would have a bias of zero, a single weight of one, and the identity activation function. These do not get adjusted in the training of the network. If the NN is to perform regression, the output layer must be a single neuron with the identity activation function. The output of the network is then a single, unrestricted value. This arrangement of neuron layers is a fully-connected feedforward NN as data flows from the input to the output layer. The diagram in Figure 4.2 is elucidating.

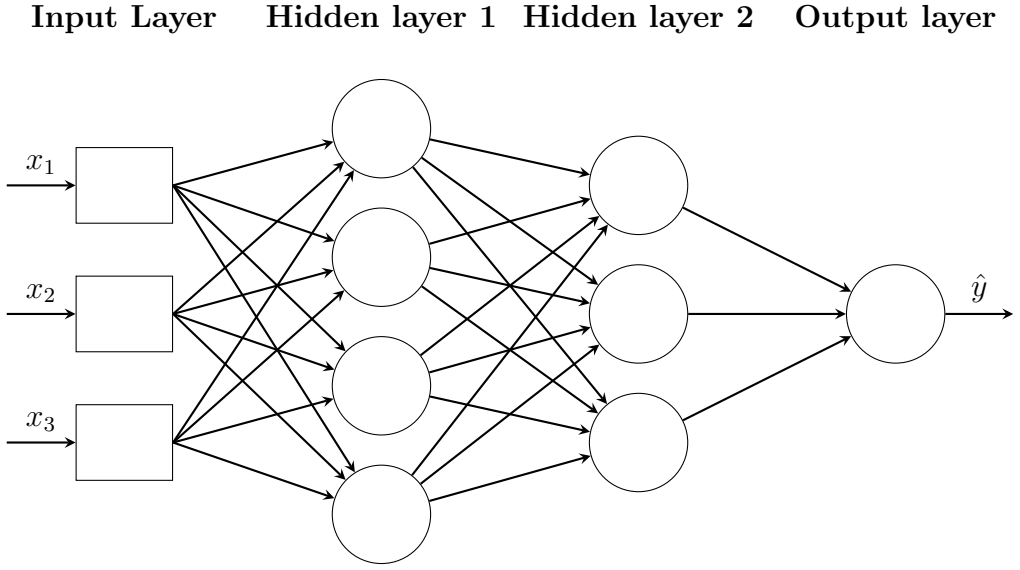


Figure 4.2: Diagram of a feedforward fully-connected NN with two hidden layers, a three-dimensional input, and a single output.

It is useful to establish some terminology relating to the architecture of a NN. Any layers in between the input and output layers are called hidden layers. NNs with many hidden layers are called deep while NNs with many neurons per layer are called wide. The weights and biases of the network are optimised during training while other parameters are set at design and do not change during training. These fixed parameters are called hyperparameters while the weights and biases are often referred to as just parameters. Although the hyperparameters are not changed during the training, they are adjusted after training and evaluation of a NN to find their optimal values. Note that the number of layers and the number of neurons in the layers are considered hyperparameters.

4.2 Backpropagation in Neural Networks

Once a NN is initialised it must go through a training process where known input-output pairs are used to adjust the weights and biases of the network. To find the optimal parameters we need to define a loss function that measures how well the NN is predicting the outputs. A typical choice of loss function is mean squared error but many others are viable. Training the NN is done by gradient descent where a minimum for the loss function in the parameter space is found by following the direction of steepest descent in discrete steps. For a loss function L and parameters $\theta_\mu^{(i)}$, a single iteration of gradient descent is the adjustment

$$\theta_\mu^{(i)} \to \theta_\mu^{(i)} - \eta \frac{\partial L}{\partial \theta_\mu^{(i)}}, \tag{4.2.1}$$

where η is an important hyperparameter called the learning rate. The choice of η dictates the size of the adjustment in parameter space which is vital in the search for the loss function's global minimum. Fortunately, there are many adaptive algorithms available for choosing an appropriate learning rate [47].

The algorithm used to evaluate the gradients used in (4.2.1) is called backpropagation. It is called this since input data is first fed forward through the network so that the outputs at each layer are computed and stored, and then a backward pass is performed where the gradients are calculated starting with the last layer and working backwards. We say that gradient descent is performed via a backpropagation algorithm.

To understand how the gradients are calculated, we consider the example of a NN with s layers trained on a single input-output pair of arbitrary dimensions $(\mathbf{x}, \mathbf{y})$ and we shall ignore the biases. We choose a mean squared error loss function

$$L := \frac{1}{n_s} \sum_{\mu=1}^{n_s} (y_\mu - l_\mu^{(s)})^2, \tag{4.2.2}$$

where n_i denotes the number of neurons in the i-th layer. Note that for a regression problem we would have $n_s = 1$. For compactness in what follows we write

$$z_\mu^{(i)} := \sum_{\nu=1}^{n_{i-1}} W_{\mu\nu}^{(i)} l_\nu^{(i-1)} \tag{4.2.3}$$

so that $l_\mu^{(i)} = \sigma^{(i)}(z_\mu^{(i)})$. We also turn off summation notation for this example. In the forward propagating stage, the input is fed into the network and $l_\mu^{(i)}$ and $z_\mu^{(i)}$ are computed

for all layers and the values stored. In the backward pass, the gradients of the loss function with respect to the weights is computed using this information.

The first thing to note is that L depends on a weight $W^{(i)}_{\mu\nu}$ only via the summed input $z^{(i)}_{\mu}$ to the neuron labelled with μ. This means that the chain rule can be applied as

$$\frac{\partial L}{\partial W^{(i)}_{\mu\nu}} = \frac{\partial L}{\partial z^{(i)}_{\mu}} \frac{\partial z^{(i)}_{\mu}}{\partial W^{(i)}_{\mu\nu}}. \tag{4.2.4}$$

It will be helpful to use the notation

$$\delta^{(i)}_{\mu} := \frac{\partial L}{\partial z^{(i)}_{\mu}}. \tag{4.2.5}$$

Looking at (4.2.3) we see that the second factor in (4.2.4) is given by

$$\frac{\partial z^{(i)}_{\mu}}{\partial W^{(i)}_{\mu\nu}} = l^{(i-1)}_{\nu}. \tag{4.2.6}$$

Substituting (4.2.5) and (4.2.6) into (4.2.4) gives

$$\frac{\partial L}{\partial W^{(i)}_{\mu\nu}} = \delta^{(i)}_{\mu} l^{(i-1)}_{\nu}. \tag{4.2.7}$$

All of the $l^{(i)}$ are known from the forward pass, so for the derivatives to be evaluated in the backward pass we need only to calculate the value of $\delta^{(i)}_{\mu}$ for the neurons in the output layer and in each hidden layer of the network. Calculating $\delta^{(i)}$ at the output layer can be done explicitly giving

$$\delta^{(s)}_{\mu} = -\frac{2}{n_s}\big(y_{\mu} - \sigma^{(s)}(z^{(s)}_{\mu})\big)\sigma'^{(s)}(z^{(s)}_{\mu}), \tag{4.2.8}$$

which can be computed easily from the information gathered in the forward pass.

To evaluate $\delta^{(i)}$ for the hidden layers we must work recursively. We employ the chain rule to calculate $\delta^{(s-1)}$

$$\begin{aligned} \delta^{(s-1)}_{\mu} = \frac{\partial L}{\partial z^{(s-1)}_{\mu}} &= \sum_{\nu=1}^{n_s} \frac{\partial L}{\partial z^{(s)}_{\nu}} \frac{\partial z^{(s)}_{\nu}}{\partial z^{(s-1)}_{\mu}} \\ &= \sum_{\nu=1}^{n_s} \delta^{(s)}_{\nu} \frac{\partial z^{(s)}_{\nu}}{\partial z^{(s-1)}_{\mu}} \\ &= \left(\sum_{\nu=1}^{n_s} W^{(s)}_{\nu\mu} \delta^{(s)}_{\nu} \right) \sigma'^{(s-1)}(z^{(s-1)}_{\mu}). \end{aligned} \tag{4.2.9}$$

We have used that variations in $z^{(s-1)}$ give rise to variations in L only through $z^{(s)}$. Note that the summation is over the first index since information is now propagating backwards

through the network. By induction we can produce an expression for any of the hidden layers $\delta^{(i)}$ known as the backpropagation formula

$$\delta_\mu^{(i)} = \left(\sum_{\nu=1}^{n_{i+1}} W_{\nu\mu}^{(i+1)} \delta_\nu^{(i+1)} \right) \sigma'^{(i)}(z_\mu^{(i)}). \tag{4.2.10}$$

Using backpropagation we can now compute the gradients (4.2.7) for every neuron. The weights can then be adjusted accordingly (4.2.1) and one iteration of gradient descent has been completed. The biases can be incorporated into the weight matrix by including an extra input fixed at +1, so no additional complexity is introduced by including biases.

Note that the above is an example of one step in a stochastic gradient descent where the parameters are adjusted after each input-output pair has been processed. It is most common that the input-output pairs are processed in batches and the parameters are adjusted after each batch. It should also be noted that during training the full input-output data set is passed through the network multiple times. The full training data being processed once comprises one training epoch.

4.3 Training and Evaluating Neural Networks

During supervised learning, the NN is provided with a number of seen input-output pairs. This is typically split into a training set, a validation set, and a test set. The majority of the data is usually contained in the training set with the test and validation set being similar sizes.

The training set is used to perform gradient descent and optimise the trainable parameters of the network as discussed in § 4.2. The validation set is used to evaluate the performance of the NN while the hyperparameters are being adjusted and the test set is used to evaluate the final performance after which the hyperparameters must be unchanged.

To understand why the seen data is split like this we need to look at underfitting and overfitting. Underfitting occurs when a network is not complex enough to learn the general features of the input set and performs poorly because of this. This is generally caused either by having a NN with too few neurons, or by training the NN for too few epochs. Overfitting happens when the NN becomes too specialised to the training data

and fails to learn general features which leads to poor perfomance on unseen data. This can occur if there are too many neurons as then the NN has the capacity to 'memorise' the training set causing it to struggle with unseen data. It can also occur if too many epochs are used in the training.

During the training of a NN the performance against the training set should increase convergently with each epoch. This is happening as the gradient descent guides the network through the paramter space to a minima of the loss function. To avoid overfitting, we use the validation set to evaluate how the NN performs on unseen data. This is important because it allows us to recognise when the NN is getting too spectialsed to the training set. The performance of the NN against the validation set is evaluated after each training epoch. If the error against the validation set stops improving, the training is halted. This is a technique known as early stopping. When early stopping is implemented we often allow for a set number epochs with no improvement before halting to allow for fluctuations. We call this number the patience. After the NN has finished training the performance against the validation set is also used to make adjustments to the hyperparameters. Finding the optimal hyperparameters is the main difficulty in building a effective NN. There are some guiding principles in this search but it is mainly a task of trial and error. There also exists many hyperparameter optimisation algorithms (HPOs) that can be read about in [34].

The final performance of the NN is evaluated with the as of yet unused test set. It is important that the test data is detached from the construction of the NN to ensure that it is representative of truly unseen data.

An additional technique for preventing overfitting is to make some of the neuron layers dropout layers [49]. For each training epoch a dropout layer will temporarily disable some of its neurons so that their output is ignored on the forward pass and the weights do not get adjusted on the backward pass. The neurons in a dropout layer to be disabled are chosen randomly and the fraction of the layer that is disabled is a hyperparameter. Randomly removing neurons forces the network to look for more features in the inputs rather than relying on a specific set of weights that may be particular to the training set. By forcing other representations of the input data to be learnt, the NN may perform better on unseen data. In terms of gradient descent, dropout layers mean that more of

the parameter space is explored so that the network is less likely to get stuck in local minima of the loss function.

5 Predicting Hodge Numbers

In this section a study of the viability of NNs as a tool in the Calabi-Yau landscape is undertaken. To do this, we attempt to use NNs to predict unseen Hodge data for CICY4s. The purpose of this is not to show that NNs can be of immediate application, but to develop a precedent for how the techniques of machine learning can intersect with the study of string theory. Work on this began in [26] where it was suggested that machine learning may be applicable to the string landscape and some basic NNs were built to demonstrate this point. Further work in [9, 10] showed that NNs can learn topological data of CICY3s, the Hodge number $h^{1,1}$ of unseen CICY3s were predicted correctly to an accuracy of 0.81 ± 0.01. The review [47] covers many other machine learning techniques that have seen application in various areas of string theory.

The CICY4 data set is chosen as it is a natural continuation on the work done for CICY3s. The fourfold data is of a similar form but is much larger and has a far greater range of Hodge numbers so will provide new challenges for a NN. Studying the CICY4 data is an important step to establishing the viability of a NN as a tool but is sparsely investigated in the literature. Learning the Hodge data for CICY4s is briefly mentioned in [26] where $h^{1,1}$ for a small portion of the data is learnt using a generic NN with promising results. Another study [28] used a NN to predict with high accuracy if a CICY4 is elliptically fibred. This is a question of great importance to F-theory compactifications but is a binary query so not a rigorous test of a NN's viability as a tool. This study presents a complete evaluation of the performance of NNs on the Hodge data of CICY4s. We will attempt to create a NN for each of the four Hodge numbers the is capable of predicting a Hodge number of an unseen CICY4 configuration matrix. The approach taken will differ slightly to the literature where the performance against the entire data set is often used to tune the NNs. Here, we will consider a fraction of the CICY4 data to be known and test against the entire data set only after the development phase of the

NNs is complete. By working this way, the viability of NNs as a tool for incomplete data sets is better evaluated. Using the entire data set in the development is best suited to probing for unknown underlying structure.

It should be noted that this work will not be a definitive investigation since the techniques used are far from what is considered cutting edge today. In addition to this, computing resources were limited so the hyperparameters of the final NNs can likely be optimised further. The accuracies obtain in this study for the prediction of Hodge numbers can almost certainly be surpassed. For these reasons, this work should be considered as an exploration of what the intersection of string theory and machine learning may yield in the future. This goal motivates a presentation of the work that puts some focus on how NNs were built and applied to the CICY4s as well as the results of their application.

5.1 CICY4 Data

The CICY4 Hodge data has been computed in full in [21, 22] and is available from: www-thphys.physics.ox.ac.uk/projects/CalabiYau/Cicy4folds. The data includes the configuration matrix entries, the Euler number, and the Hodge numbers. The list numbers 921, 497 CICYs but 15, 813 of them have block diagonal configuration matrices. This means that they they describe product manifolds. These are considered redundant since to have complex dimension 4 they must be either T^8, $T^2 \times CY_3$, $T^4 \times K3$, or $K3 \times K3$ and so their Hodge data follows from Calabi-Yau manifolds of lower dimension. With these excluded we are left with a data set of 905, 684 configuration matrix-Hodge number pairs for each Hodge number.

The distribution of the CICY4 Hodge data is plotted in Figure 5.1. The Euler number is included also as it helps in realising the topological variety in the data set. The Hodge numbers are generally clustered at the lower end and fall off for higher Hodge numbers. We see that $h^{3,1}$ and $h^{2,2}$ are likely going to be the most difficult for a NN to learn as they have a large range of possible outputs, 235 and 304 respectively compared to 23 and 31 for $h^{1,1}$ and $h^{2,1}$ respectively. The mean values for the Hodge numbers and their maximal and minimal values are

$$\left\langle h^{1,1} \right\rangle = 10.1^{24}_{1}, \quad \left\langle h^{2,1} \right\rangle = 0.817^{33}_{0}, \quad \left\langle h^{3,1} \right\rangle = 39.6^{426}_{20}, \quad \left\langle h^{2,2} \right\rangle = 241^{1752}_{204}.$$

The configuration matrix is represented in the data set with a two-dimensional array that contains the polynomial degrees e.g. for three polynomials in $(\mathbb{CP}^3)^2 \times \mathbb{CP}^5$ the configuration matrix may look like `{{0,1,1},{1,0,1},{1,3,2}}`. We note that the largest fourfold configuration matrix, not including the column for the dimensions of the projective spaces, is 16×20. The NN requires that all the inputs are of the same dimension so that the input layer can be initialised appropriately. To ensure this we make all our configuration matrices into 16×20 matrices by padding with zeros. We then flatten the 16×20 matrices into a one-dimensional array with 320 entries. We are then left with four sets, one for each Hodge number, of input-output pairs that for the Hodge number h are of the form

$$D := \{d_1, \ldots, d_{320}\} \longrightarrow h, \tag{5.1.1}$$

where $d_i \in \mathbb{Z} : d_i \in [0, 6]$.

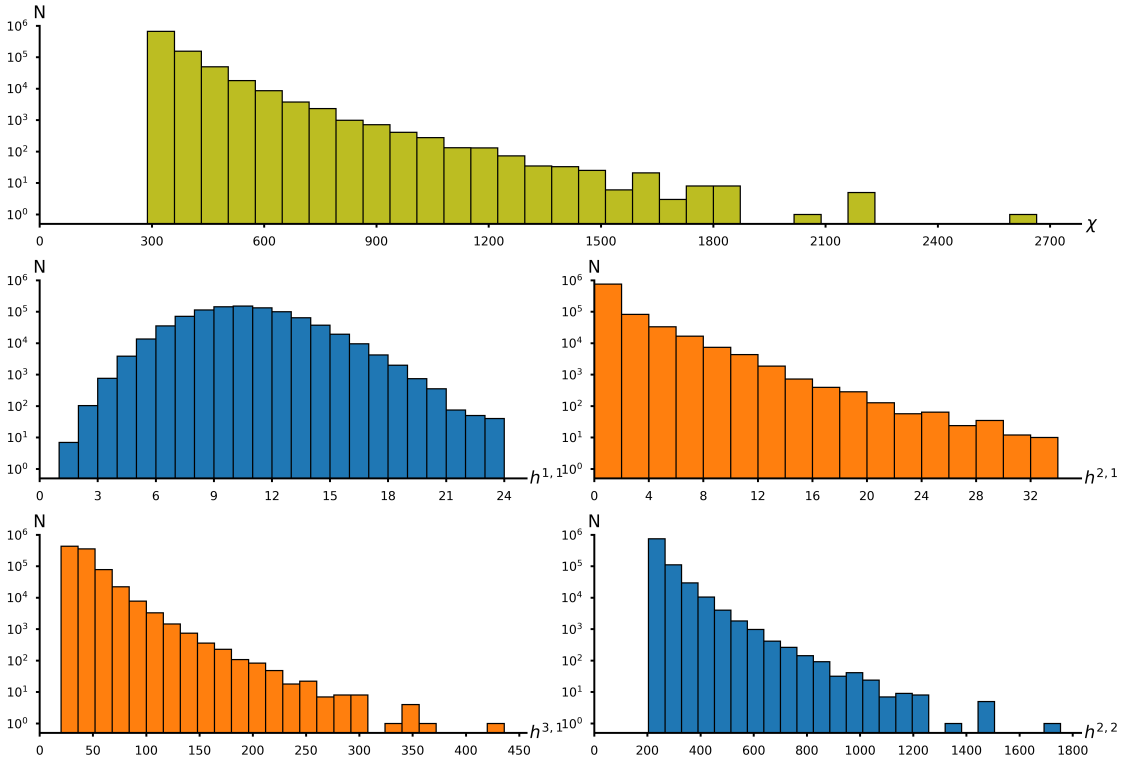


Figure 5.1: Distribution of Euler number (top), $h^{1,1}$ (middle left), $h^{2,1}$ (middle right), $h^{3,1}$ (bottom left), and $h^{2,2}$ (bottom left) in the CICY4 data computed in [22].

5.2 Creating and Training Neural Networks

To create NNs capable of predicting Hodge numbers we use the `Tensorflow` library for `Python`. To interface with this package we use the `Keras` library. This is a high-level interface that allows NNs to be built in layers. We will use the `Sequential` model where each layer of a feedforward fully-connected NN is specified. All code used is original and written in `Python` using these libraries.

The difficulty in creating an effective NN is in choosing choosing the optimal hyperparamters. The hyperparameters we will consider are: the number of layers, the number of neurons in each layer, the activation functions, and the dropout rates. Finding the optimal hyperparameters is a task of trial and error which is often automated by employing HPOs. We will choose the hyperparamters by hand as we would like a familiarity with the NNs and an understanding of their development process. In addition to this, we lack the computing resources to properly apply HPOs to NNs for all four Hodge numbers.

In order to create the NNs by hand, we divide the CICY4 data as follows

$$\text{training set} : \text{validation set} : \text{stoppping set} : \text{test set}. \tag{5.2.1}$$

The union training set $\cup$ validation set $\cup$ stopping set is the seen data. The performance against the stopping set is evaluated after each training epoch and used to dictate when early stopping is triggered. The validation set is used to tune the hyperparamters of the NN between trainings. Dividing the usual role of the validation set between a stopping set and a smaller validation set ensures the the NN does not become too specialised to any one set of data.

To create the NNs we consider 10% of the CICY4 data, chosen at random, to be known, simulating a scenario where we are presented with an incomplete data set. The remaining 90% of the data is then allocated to the test set and we must divide the 10% among the the seen data sets. We make the split of the seen data into training, validation, and stopping as $70 : 15 : 15$. Note that after the hyperparameter tuning is complete the validation and training set are merged for the final training and evaluation against the test set.

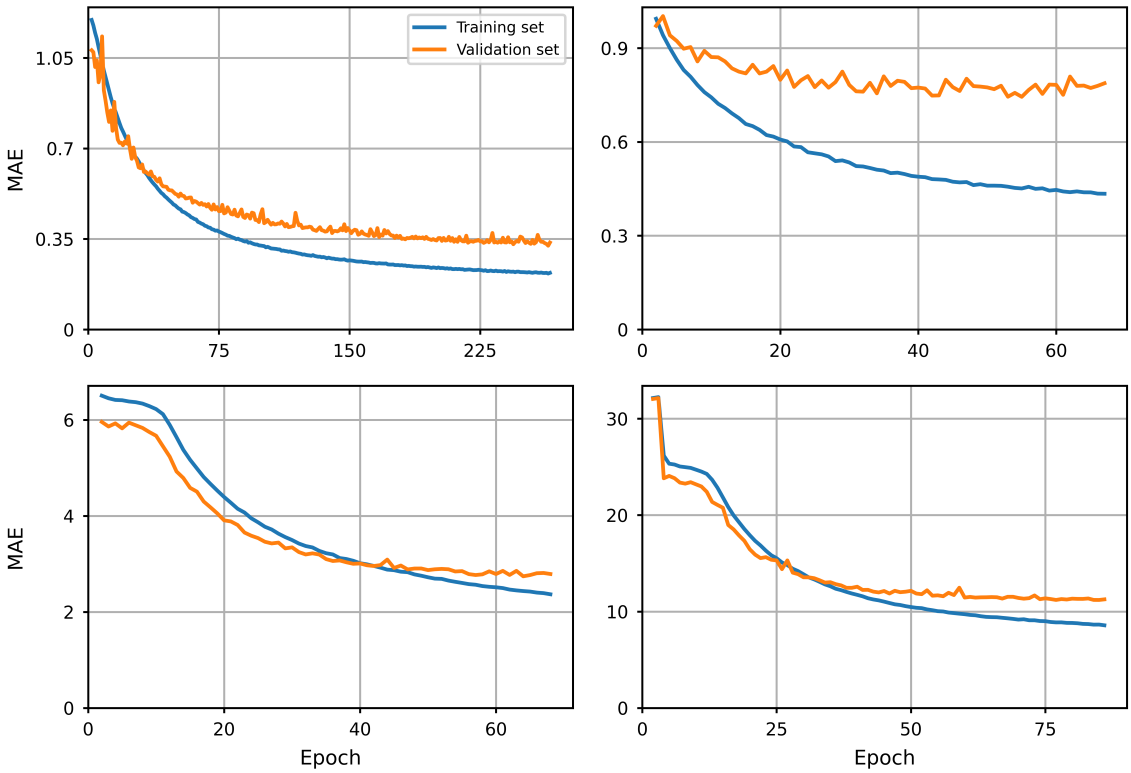


Figure 5.2: Training curves for $h^{1,1}$ (top left), $h^{2,1}$ (top right), $h^{3,1}$ (bottom right), and $h^{2,2}$ (bottom left) showing the mean absolute error of the networks as a function of training epoch.

With the data partitioned in this way, four NNs are created with their hyperparameters each tuned to one of the four Hodge number data sets. The architectures of these NNs are documented in Appendix A. The networks are trained with a mean squared error loss function, using the ADAM optimiser [36] for gradient descent, and until early stopping with a patience of 20 is triggered. Each NN takes approximately 2 hours on a laptop CPU to train.

Before evaluating the performance of the NNs against the Hodge data, we would like to examine the training curves found in Figure 5.2. We note that the fluctuations in the $h^{2,1}$ training curve are due to the use of the ReLU activation function in the network and are not indicative of data set. The curves for $h^{3,1}$ and $h^{2,2}$ appear to level off early in the training before converging to an improved accuracy. This could be suggestive of the maps $D \to h^{3,1}$ and $D \to h^{2,2}$ being more complex than for the other Hodge numbers but could also just be a result of having less distinct configuration matrices per Hodge number in the training set. For $h^{2,2}$, an indication of the former being the case is that during the

creation process the network would often get stuck outputting the average Hodge number for every input. This means that the NN cannot recognise any generalities in the input data and is reverting to outputting the average to minimise the loss function.

5.3 Evaluating Neural Networks

To evaluate the ability of the NNs to predict Hodge numbers we will first look at the accuracies and predicted distributions of the Hodge numbers for the case where the 10% of the data used for creating the NNs is taken as seen.

5.3.1 Prediction Accuracies

The final performance of the NNs can be seen in Table 1. By accuracy, we mean the likelihood that the predicted Hodge number rounded to the nearest integer is correct.

For $h^{1,1}$, the NN is able to predict the Hodge number of the unseen test set configuration matrices with an accuracy of 0.82. This is an impressive result but an 18% chance of an incorrect prediction cannot be considered reliable. The NN for $h^{2,1}$ performed similarly with an accuracy reduced approximately in proportion to $h^{2,1}$'s greater range of outputs. The results of these NNs suggests that machine learning is best suited to identifying candidates for conventional computations to produce data sets with desirable properties, rather than creating entire data sets independently.

We should note that there exists a simple relation between the configuration matrix and $h^{1,1}$ for favourable CICY4s. A favourable CICY is one where the second de Rham cohomology class descends one-to-one from the ambient space that is the product of

Table 1: Summary of accuracies and mean squared errors when trained with 10% of the data seen. Errors are obtained by repeating the training and prediction 5 times.

	Accuracy	MAE
$h^{1,1}$	0.820 ± 0.006	0.27 ± 0.01
$h^{2,1}$	0.633 ± 0.012	0.72 ± 0.01
$h^{3,1}$	0.159 ± 0.002	2.76 ± 0.01
$h^{2,2}$	0.036 ± 0.004	11.39 ± 0.19

projective spaces. This means that $b^2 = h^{1,1} \geq m$ for a favourable CICY4 in a product of m projective spaces [25]. For the CICY4 data (which is labelled as favourable or not) we find that 54% are favourable with $h^{1,1} = m$. So to learn $h^{1,1}$ for these CICY4s the NN would need to learn how to recognise the number of rows in the configuration matrix. We need to check if this is happening to establish if the $h^{1,1}$ results can be considered representative of a NN's performance on generic data sets in string theory. To do so the accuracy of the NN in predicting the favourable CICY4s only is evaluated. This yields an accuracy of 0.909 ± 0.005, somewhat higher than for the full test set. This increase in accuracy can be partly attributed to the favourable CICY4s having $h^{1,1} \in [1, 16]$ where the NN performs better. These results suggest that the NN is not directly exploiting the favourable $h^{1,1}$ restriction.

Both $h^{3,1}$ and $h^{2,2}$ perform significantly worse than $h^{1,1}$ and $h^{2,1}$, highlighting the limitations of NNs. That is, if there are too few seen inputs relative to the output dimension, the network will perform poorly. In fact, the $h^{2,2}$ NN would've achieved a higher accuracy ($\sim 10\%$) had it outputted the modal value of 204 for all unseen inputs.

The mean absolute error for the Hodge number predictions is promising. On an incorrect prediction, the NN will often predict a Hodge number relatively close to the true value. This is a useful since it could allow for a subset of configuration matrices where it is likely that one with a desired Hodge number resides to be identified. This reinforces the notion that machine learning may have a role in string theory as a tool that guides computation in landscapes too large for complete classification.

The difference in performance between the NNs appears to approximately scale inversely with the number of possible Hodge numbers. There are too few data points to make the conclusion that none of the Hodge numbers are intrinsically more complex than the others but the data does seem to suggest this. A study of higher dimensional CICYs would address this properly.

5.3.2 Predicted Distributions

To examine the performance of the NNs more closely we look at the predicted Hodge number distributions for the case where 10% of the data is seen. These can be seen in Figure 5.3. Overall, the NNs are able to reproduce the true distributions quite well. This

is due to the networks' tendency to make incorrect predictions near to the true Hodge number.

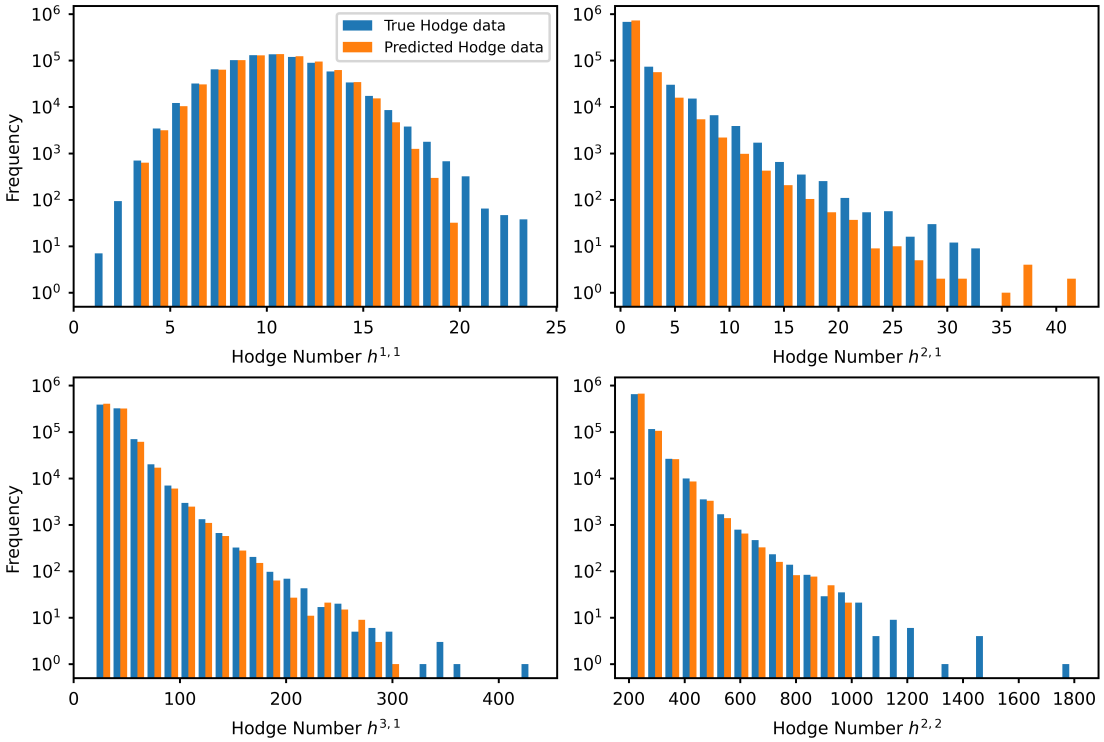


Figure 5.3: Predicted and true distributions for $h^{1,1}$ (top left), $h^{2,1}$ (top right), $h^{3,1}$ (bottom left), and $h^{2,2}$ (bottom right) with 10% seen data.

The predicted distributions for $h^{1,1}$, $h^{3,1}$, and $h^{2,2}$ share a similar relationship with their respective true distributions. They match with the true distribution well near the modal Hodge number but fail to predict the Hodge numbers on the tails. For the $h^{2,1}$ distribution the discrepancy between predicted and true is spread evenly throughout the Hodge numbers, interestingly with some Hodge numbers beyond the range the NN has seen being predicted. The reason for this difference is that the $h^{2,1}$ NN is the only one to use ReLU activation functions, the others use sigmoids. When sigmoid activation functions are used, the inputs to the output layer are restricted to $(0, 1)$ so that the weights in the output layer have to forgo significant performance on the most common Hodge numbers just to ensure it can output extremal values. The ReLU can output any non-negative number so that the output layer can produce extremal values from its inputs without needing significant adjustment to the weights. This highlights the requirement that a NN is purpose built for a problem. If the desirable Hodge numbers are at the tails

ends of the seen data, then the hyperparameters of the NN must be chosen to maximise performance against those numbers.

5.3.3 Learning Curves

To understand how the performance changes with the fraction of the CICY4 data that is seen by the NNs we look at the learning curves. These can be seen in Figure 5.4. Note that it would be informative to extend the seen data tested beyond 30% but time constraints prevented this as the data point at 30% took approximately a day on a laptop CPU to compute.

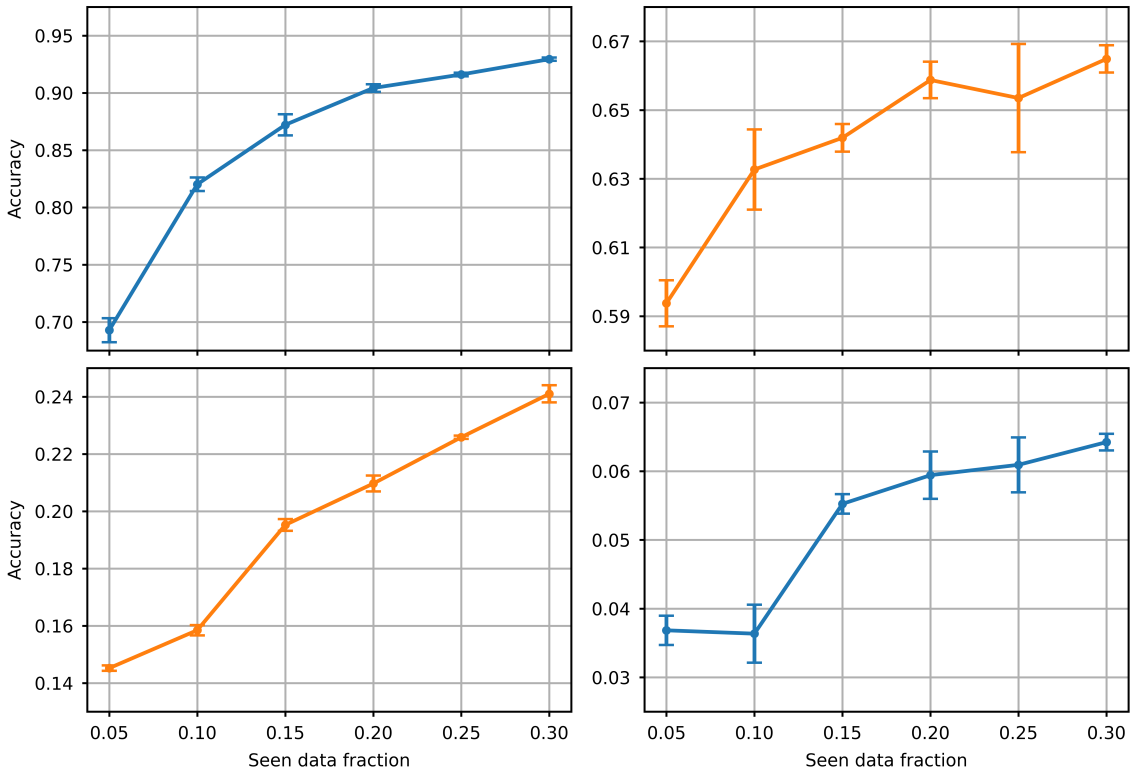


Figure 5.4: Learning curves for $h^{1,1}$ (top left), $h^{2,1}$ (top right), $h^{3,1}$ (bottom left), and $h^{2,2}$ (bottom right) showing the accuracy of the NNs as a functions of the fraction of data seen. Error bars are obtained by repeating the training and prediction 5 times.

As expected, a larger training set improves the performance of the NNs. The learning curve of the $h^{1,1}$ NN is the only one to show significant improvements in performance, reaching an accuracy of 0.93 with 30% of the data seen. Considering that the $h^{1,1}$ NN can likely be optimised further, this suggests that it may be possible to learn $h^{1,1}$ near perfectly. Achieving near perfect accuracy would make machine learning a much more

powerful tool for model building and open up the possibility of using it to reveal hidden relations in string theory data sets. Although the NNs for other Hodge numbers do improve with more seen data, it is not by enough to elevate machine learning beyond a guiding role in model building. Take the $h^{2,1}$ NN for example, increasing the accuracy from 0.59 to 0.66 by increasing the seen data from 5% to 30% does not change how we would use the predictions from the NNs.

A general remark from the learning curves is that it is important to maximise the amount of seen data available. The best way to do this is to have more of the data set computed but when this is not possible there are alternative ways. The seen data can be enlarged in a process called data augmentation. For CICYs this would mean including all permutations of the configuration matrix in with the original seen data. Such techniques are not used in this work but a successful application of them to CICY3s can be seen in [9].

Conclusion

In this dissertation we gave a review of topics from complex geometry necessary for understanding the Calabi-Yau manifold. The Calabi-Yau manifolds was then introduced and we saw how it enters the field of string theory through the compactification of the heterotic string. We then described how Calabi-Yau manifolds can be constructed as complete intersections of polynomials in projective spaces. The notion of the Calabi-Yau landscape was explored briefly and a machine learning approach to it was motivated. The basic principles behind NNs were then explained.

After presenting many concepts important to understanding machine learning and the Calabi-Yau landscape, we undertook an investigation into the applicability of NNs to the CICY4 Hodge data, something that is absent in the literature. The aim being to further establish the role of machine learning in string theory. We found mixed success with accuracies for a NN that has seen 10% of the CICY4s as high as 0.82 for $h^{1,1}$ and as low as 0.04 for $h^{2,2}$. The results suggest that machine learning techniques can be used to guide model building efforts by identifying promising regions in the landscape of string theory vacua. The success of the NN in $h^{1,1}$ prediction indicates that machine learning may even be applied to make conjectures about data sets in string theory.

Further work on the application of machine learning to the CICY4 Hodge data would require making use of more advanced techniques such as data augmentation, utilisation of HPOs, or experimentation with convolutional NNS. A recent study [20] was able to model a NN on Google's Inception model [50] that could predict $h^{1,1}$ of the CICY3s with an accuracy of 0.97 after having seen 30% of the data. Application of such a NN to the CICY4 data may yield interesting results. In a more general sense, it is important that we experiment with machine learning in the landscape of string theory vacua so that its role in string theory can be better established. By doing this, we might hope that string theory will thrive in the age of big data.

References

[1] Steven Abel and John Rizos. "Genetic Algorithms and the Search for Viable String Vacua". In: *JHEP* 08 (2014), p. 010. DOI: 10.1007/JHEP08(2014)010. arXiv: 1404.7359 [hep-th].

[2] Lara B. Anderson et al. "Yukawa Couplings in Heterotic Compactification". In: *Commun. Math. Phys.* 297 (2010), pp. 95–127. DOI: 10.1007/s00220-010-1033-8. arXiv: 0904.2186 [hep-th].

[3] Victor V. Batyrev and Lev A. Borisov. "On Calabi-Yau complete intersections in toric varieties". In: *Proceedings of the International Conference* (Trento). Vol. 1. De Gruyter, 1994, pp. 39–65. DOI: 10.1515/9783110814736.

[4] Arnaud Beauville. *Complex Algebraic Surfaces*. 2nd ed. London Mathematical Society Student Texts. Cambridge University Press, 1996. DOI: 10.1017/CBO9780511623936.

[5] K. Becker, M. Becker, and J.H. Schwarz. *String theory and M-theory: A modern introduction*. Cambridge University Press, 2006. DOI: 10.1017/CBO9780511816086.

[6] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006.

[7] R. Bott and L.W. Tu. *Differential Forms in Algebraic Topology*. Graduate Texts in Mathematics. Springer New York, 1995. DOI: 10.1007/978-1-4757-3951-0.

[8] Vincent Bouchard. "Lectures on complex geometry, Calabi-Yau manifolds and toric geometry". In: *arXiv e-prints* (2007). arXiv: hep-th/0702063 [hep-th].

[9] Kieran Bull et al. "Machine Learning CICY Threefolds". In: *Phys. Lett. B* 785 (2018), pp. 65–72. DOI: 10.1016/j.physletb.2018.08.008. arXiv: 1806.03121 [hep-th].

[10] Kieran Bull et al. “Getting CICY High”. In: *Phys. Lett. B* 795 (2019), pp. 700–706. DOI: 10.1016/j.physletb.2019.06.067. arXiv: 1903.03113 [hep-th].

[11] Eugenio Calabi. “The Space of Kähler metrics”. In: *Proceedings of the International Congress of Mathematicians* (Amsterdam). Vol. 2. Erven P. Noordhoff N.V., Groningen, North-Holland Publishing Co. Amsterdam 1957, 1954, pp. 206–207.

[12] P. Candelas. “Lectures on Complex Manifolds”. In: *1987 Spring School on Superstrings* (Trieste). Vol. 1. Singapore: World Scientific, 1987, pp. 1–88.

[13] P. Candelas, C.A. Lutken, and R. Schimmrigk. “Complete Intersection Calabi-Yau Manifolds. 2. Three Generation Manifolds”. In: *Nucl. Phys. B* 306 (1988), p. 113. DOI: 10.1016/0550-3213(88)90173-3.

[14] P. Candelas et al. “Vacuum Configurations for Superstrings”. In: *Nucl. Phys. B* 258 (1985), pp. 46–74. DOI: 10.1016/0550-3213(85)90602-9.

[15] P. Candelas et al. “Complete Intersection Calabi-Yau Manifolds”. In: *Nucl. Phys. B* 298 (1988), p. 493. DOI: 10.1016/0550-3213(88)90352-5.

[16] Jonathan Carifio et al. “Machine Learning in the String Landscape”. In: *JHEP* 09 (2017), p. 157. DOI: 10.1007/JHEP09(2017)157. arXiv: 1707.00655 [hep-th].

[17] Wei-Liang Chow. “On Compact Complex Analytic Varieties”. In: *American Journal of Mathematics* 71.4 (1949), pp. 893–914. DOI: 10.2307/2372375.

[18] Charles F. Doran and Ursula A. Whitcher. “From Polygons to String Theory”. In: *Mathematics Magazine* 85.5 (2012), pp. 343–359. DOI: 10.4169/math.mag.85.5.343.

[19] Tohru Eguchi, Peter B. Gilkey, and Andrew J. Hanson. “Gravitation, Gauge Theories and Differential Geometry”. In: *Phys. Rept.* 66 (1980), p. 213. DOI: 10.1016/0370-1573(80)90130-1.

[20] Harold Erbin and Riccardo Finotello. “Inception Neural Network for Complete Intersection Calabi-Yau 3-folds”. In: *arXiv e-prints* (2020). arXiv: 2007.13379 [hep-th].

[21] James Gray, Alexander S. Haupt, and Andre Lukas. "All Complete Intersection Calabi-Yau Four-Folds". In: *JHEP* 07 (2013), p. 070. DOI: 10.1007/JHEP07(2013)070. arXiv: 1303.1832 [hep-th].

[22] James Gray, Alexander S. Haupt, and Andre Lukas. "Topological Invariants and Fibration Structure of Complete Intersection Calabi-Yau Four-Folds". In: *JHEP* 09 (2014), p. 093. DOI: 10.1007/JHEP09(2014)093. arXiv: 1405.2073 [hep-th].

[23] Michael B. Green, John H. Schwarz, and Edward Witten. *Superstring Theory Vol. 2: 25th Anniversary Edition*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 2012. DOI: 10.1017/CBO9781139248570.

[24] Paul S. Green, Tristan Hubsch, and Carsten A. Lutken. "All Hodge Numbers of All Complete Intersection Calabi-Yau Manifolds". In: *Class. Quant. Grav.* 6 (1989), pp. 105–124. DOI: 10.1088/0264-9381/6/2/006.

[25] Paul Green and Tristan Hubsch. "Calabi-yau Manifolds as Complete Intersections in Products of Complex Projective Spaces". In: *Commun. Math. Phys.* 109 (1987), p. 99. DOI: 10.1007/BF01205673.

[26] Yang-Hui He. "Deep-Learning the Landscape". In: *arXiv e-prints* (2017). arXiv: 1706.02714 [hep-th].

[27] Yang-Hui He. "The Calabi-Yau Landscape: from Geometry, to Physics, to Machine-Learning". In: *arXiv e-prints* (2018). arXiv: 1812.02893 [hep-th].

[28] Yang-Hui He and Seung-Joo Lee. "Distinguishing elliptic fibrations with AI". In: *Phys. Lett. B* 798 (2019), p. 134889. DOI: 10.1016/j.physletb.2019.134889. arXiv: 1904.08530 [hep-th].

[29] Yang-Hui He and Andre Lukas. "Machine Learning Calabi-Yau Four-folds". In: *arXiv e-prints* (2020). arXiv: 2009.02544 [hep-th].

[30] R. Hirzebruch. *Topological methods in algebraic geometry*. 3rd ed. Grundlehren der mathematischen Wissenschaften. Springer, 1982. DOI: 10.1007/978-3-662-30697-0.

[31] Petr Horava and Edward Witten. "Heterotic and type I string dynamics from eleven-dimensions". In: *Nucl. Phys. B* 460 (1996), pp. 506–524. DOI: 10.1016/0550-3213(95)00621-4. arXiv: hep-th/9510209.

[32] Tristan Hubsch. "Calabi-yau Manifolds: Motivations and Constructions". In: *Commun. Math. Phys.* 108 (1987), p. 291. DOI: 10.1007/BF01210616.

[33] Tristan. Hübsch. *Calabi-Yau Manifolds: A Bestiary for Physicists*. World Scientific, 1992. DOI: 10.1142/1410.

[34] Frank Hutter, Lars Kotthoff, and Joaquin Vanschoren, eds. *Automatic Machine Learning: Methods, Systems, Challenges*. Springer, 2019. DOI: 10.1007/978-3-030-05318-5.

[35] C.J. Isham. *Modern Differential Geometry for Physicists*. World Scientific lecture notes in physics: Volume 61. World Scientific, 1999. DOI: 10.1142/3867.

[36] Diederik P. Kingma and Jimmy Ba. "Adam: A Method for Stochastic Optimization". In: *arXiv e-prints* (2014). arXiv: 1412.6980 [cs.LG].

[37] A. Klemm et al. "Calabi-Yau fourfolds for M theory and F theory compactifications". In: *Nucl. Phys. B* 518 (1998), pp. 515–574. DOI: 10.1016/S0550-3213(97)00798-0. arXiv: hep-th/9701023.

[38] Maximilian Kreuzer and Harald Skarke. "On the classification of reflexive polyhedra". In: *Commun. Math. Phys.* 185 (1997), pp. 495–508. DOI: 10.1007/s002200050100. arXiv: hep-th/9512204.

[39] Maximilian Kreuzer and Harald Skarke. "Complete classification of reflexive polyhedra in four-dimensions". In: *Adv. Theor. Math. Phys.* 4 (2002), pp. 1209–1230. DOI: 10.4310/ATMP.2000.v4.n6.a2. arXiv: hep-th/0002240.

[40] Henry W. Lin, Max Tegmark, and David Rolnick. "Why Does Deep and Cheap Learning Work So Well?" In: *Journal of Statistical Physics* 168.6 (2017), pp. 1223–1247. DOI: 10.1007/s10955-017-1836-5. arXiv: 1608.08225 [cond-mat.dis-nn].

[41] Emanuel Malek. "Advanced topics in string and field theory: Complex manifolds and Calabi-Yau manifolds". 2014.

[42] Tom Michael Mitchell, R. Keller, and S. Kedar-Cabelli. "Explanation-based generalization: A unifying view". In: *Machine Learning* 1 (1986), pp. 47–80. DOI: 10.1023/A:1022691120807.

[43] H. Murayama. "Notes on Clifford Algebra and $Spin(N)$ Representations". 2007.

[44] M. Nakahara. *Geometry, topology and physics*. Graduate Student Series in Physics. CRC Press, 2003. DOI: 10.1201/9781315275826.

[45] Gal Raayoni et al. "The Ramanujan Machine: Automatically Generated Conjectures on Fundamental Constants". In: *arXiv e-prints* (2019). arXiv: 1907.00205 [cs.LG].

[46] Lisa Randall and Raman Sundrum. "An Alternative to compactification". In: *Phys. Rev. Lett.* 83 (1999), pp. 4690–4693. DOI: 10.1103/PhysRevLett.83.4690. arXiv: hep-th/9906064.

[47] Fabian Ruehle. "Data science applications to string theory". In: *Physics Reports* 839 (2020). Data science applications to string theory, pp. 1–117. DOI: https://doi.org/10.1016/j.physrep.2019.09.005.

[48] Friedrich Schöller and Harald Skarke. "All Weight Systems for Calabi–Yau Fourfolds from Reflexive Polyhedra". In: *Commun. Math. Phys.* 372.2 (2019), pp. 657–678. DOI: 10.1007/s00220-019-03331-9. arXiv: 1808.02422 [hep-th].

[49] Nitish Srivastava et al. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: *Journal of Machine Learning Research* 15.56 (2014), pp. 1929–1958.

[50] Christian Szegedy et al. "Going Deeper with Convolutions". In: *arXiv e-prints* (2014). arXiv: 1409.4842 [cs.CV].

[51] Cumrun Vafa. "Evidence for F theory". In: *Nucl. Phys. B* 469 (1996), pp. 403–418. DOI: 10.1016/0550-3213(96)00172-1. arXiv: hep-th/9602022.

[52] Stefan Vandoren. "Lectures on Riemannian Geometry, Part II: Complex Manifolds". In: 2008. arXiv: alg-geom/9412017 [math.AG].

[53] G. Veneziano. "Construction of a crossing - symmetric, Regge behaved amplitude for linearly rising trajectories". In: *Nuovo Cim. A* 57 (1968), pp. 190–197. DOI: 10.1007/BF02824451.

[54] Claire Voisin. *Hodge Theory and Complex Algebraic Geometry I*. Ed. by Leila Schneps. Vol. 1. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2002. DOI: 10.1017/CBO9780511615344.

[55] Timo Weigand. "Lectures on F-theory compactifications and model building". In: *Class. Quant. Grav.* 27 (2010). Ed. by J. Walcher, p. 214004. DOI: 10.1088/0264-9381/27/21/214004. arXiv: 1009.3497 [hep-th].

[56] Shing-Tung Yau. "Calabi's conjecture and some new results in algebraic geometry". In: *Proceedings of the National Academy of Sciences* 74.5 (1977), pp. 1798–1799. DOI: 10.1073/pnas.74.5.1798.

[57] Shing-Tung Yau. "A survey of Calabi-Yau manifolds". In: *Surveys in differential geometry* 13 (2008), pp. 277–318. DOI: 10.4310/SDG.2008.v13.n1.a9.

A Neural Network Architectures

Figures A.1 and A.2 show the NN architectures used in § 5. The NNs are built using the `Sequential` class in `Keras`. Dense (fully-connected) layers are labelled with their activation function and an identifying number if similar layers exist. Dropout layers are labelled with their dropout ratio and an identifying number if similar layers exist.

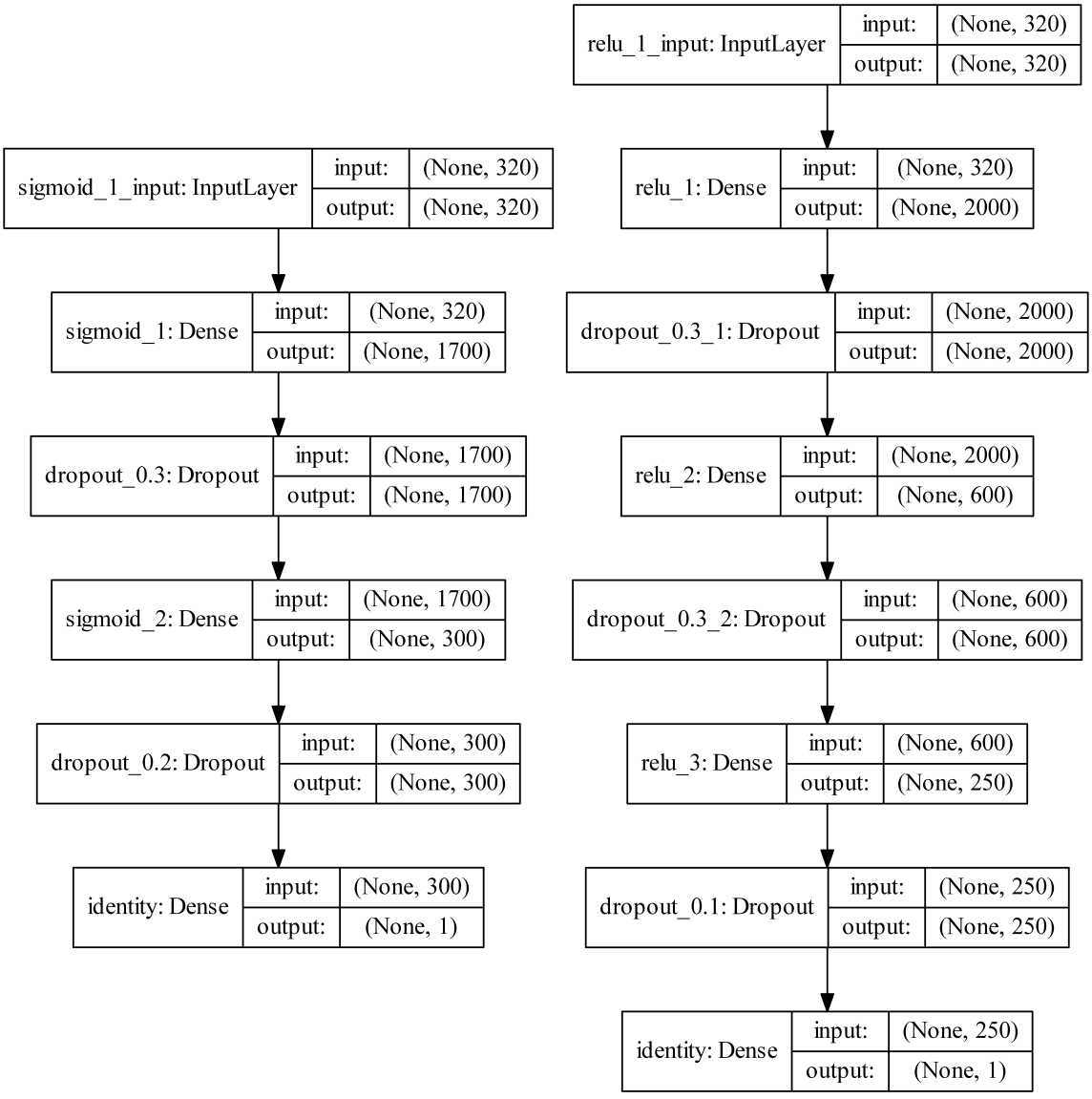


Figure A.1: Neural networks used for $h^{1,1}$ (left) and $h^{2,1}$ prediction (right).

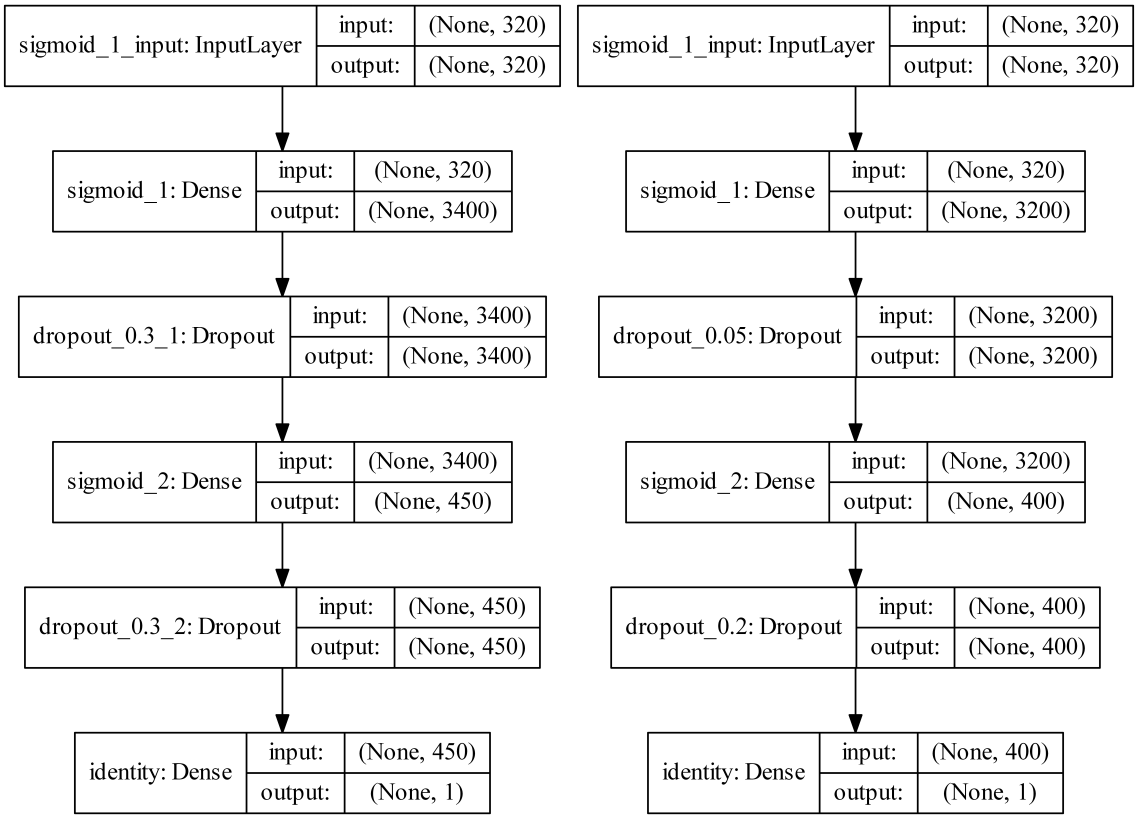


Figure A.2: Neural networks used for $h^{3,1}$ (left) and $h^{2,2}$ prediction (right).