

https://dspace.org/

Machine learning applications in string theory

Craven, Jessica Rachel

2022
https://hdl.handle.net/10539/35468

Downloaded from a DSpace repository. https://dspace.org/

Machine learning applications in string theory

Jessica Rachel Craven

under the supervision of Professor Vishnu Jejjala

A dissertation submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in fulfillment of the requirements for the degree of Master of Science.

31 March 2022, Johannesburg.

Contents

Declaration v

Abstract vi

Acknowledgements vii

Included publications viii

List of Figures x

List of Tables xi

1 Introduction 1

1.1 Machine learning . . . 1

1.1.1 Feedforward neural networks . . . 2

1.1.2 Restricted Boltzmann machines . . . 4

1.1.3 Machine learning in physics . . . 5

1.2 Knot theory . . . 6

1.2.1 Reidemeister moves . . . 7

1.2.2 Braids and braidwords . . . 8

1.2.3 The Jones Polynomial . . . 9

1.2.4 The hyperbolic volume . . . 11

1.2.5 The slice genus and the Poincaré conjectures . . . 15

1.3 Chern–Simons theory and knots . . . 16

1.3.1 The Jones polynomial in Chern–Simons theory . . . 17

1.4 The SYK model . . . 20

1.4.1 Black holes and holography . . . 21

1.4.2 The SYK model . . . 22

1.4.3 Out of time ordered correlators and chaos . . . 25

1.5 Structure of dissertation . . . 26

2 Disentangling a deep learned volume formula 27

2.1 Introduction . . . 27

2.2 Chern–Simons theory . . . 31

2.2.1 Knot invariants . . . 31
2.2.2 Analytic continuation . . . 34
2.3 Machine learning . . . 37
2.3.1 Neural networks . . . 38
2.3.2 Layer-wise relevance propagation . . . 38
2.3.3 Strategies for analyzing the network . . . 40
2.4 An approximation formula for the hyperbolic volume . . . 41
2.4.1 Interpretable deep learning . . . 41
2.4.2 Implications in Chern–Simons theory . . . 47
2.5 Discussion . . . 50

Appendices **53**
2.A Basic neural network setup . . . 53
2.B Khovanov homology, Jones coefficients, and volume . . . 54
2.B.1 Khovanov homology . . . 54
2.B.2 Power law behavior of coefficients . . . 56
2.C Normalizations, orientation, and framing . . . 57
2.D t-distributed stochastic neighbor embedding of knot invariants . . . 61
2.D.1 t-distributed stochastic neighbor embedding . . . 61
2.D.2 Volume and Chern–Simons invariant . . . 61
2.E Other experiments . . . 62
2.E.1 The HOMFLY-PT and Khovanov polynomials . . . 63
2.E.2 Representations of $J_2(K;q)$ . . . 64
2.E.3 Small networks . . . 65
2.E.4 Symbolic regression . . . 66

3 Learning knot invariants across dimensions **67**
3.1 Introduction . . . 67
3.2 Khovanov homology and the s-invariant . . . 71
3.2.1 Khovanov homology . . . 71
3.2.2 Lee homology and Rasmussen's s-invariant . . . 73
3.2.3 Rasmussen's s-invariant and the slice genus . . . 74
3.3 Gauge theory and knot invariants . . . 75
3.3.1 3d Chern–Simons theory . . . 76
3.3.2 4d $\mathcal{N} = 4$ super-Yang–Mills . . . 77
3.3.3 5d maximal super-Yang–Mills . . . 78
3.3.4 6d (0,2) theory . . . 79
3.3.5 Rasmussen's s-invariant and gauge theory . . . 79
3.4 Experiments . . . 82
3.4.1 Data generation . . . 83
3.4.2 What does the data look like? . . . 84

3.4.3 Khovanov polynomials 85
3.4.4 Jones polynomials 91
3.4.5 Interdependence of results 92
3.5 Discussion 95

Appendices **98**
3.A Implementation 98

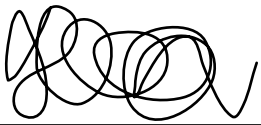
4 Learning holographic quantum chaos **100**
4.1 Introduction 100
4.2 Motivation 101
4.3 Numeric background 101
4.4 Experiments and results 103

Appendices **104**
4.A Stochastic Reconfiguration 104
4.B SYK Hamiltonian 105

5 Conclusion **107**

Declaration

I declare that I have made significant contributions to the publications included in this dissertation and that in sections where the work is unpublished, the contents are my own, original work. This dissertation is being submitted for the Degree of Master of Science at the University of Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.



(Signature of candidate)

31st day of March 2022 at Johannesburg

Abstract

We demonstrate the effectiveness of machine learning techniques as research tools in formal theory. In knot theory (and consequently the related quantum field theories, such as Chern–Simons theory), we use artificial neural networks as tools to identify relationships in large datasets. Specifically, we are able to predict the slice genus and s-invariant of knots from the Jones polynomial. Further, we show how interpretable machine learning techniques can be used to derive new analytic formulae from machine learning results. These interpretation techniques allow us to reverse engineer a neural network to predict the hyperbolic volume from evaluations of the Jones polynomial. We also demonstrate how Restricted Boltzmann Machines can be used to study computationally difficult problems, such as computing out of time ordered correlators of the SYK model. We discuss applications of this technique to numerically mapping out the phase space of chaos.

Acknowledgements

Thank you to my supervisor, Vishnu Jejjala, for your endless patience. I am grateful for the time you gave, not only to share your knowledge of physics with me, but also to teach me what being a physicist is all about.

Thank you as well to Arjun Kar and Mark Hughes for all of the useful physics and math knowledge that you imparted to me during our collaborations.

In Chapter 2, we are grateful to Onkar Parrikar for prior collaboration in machine learning aspects of knot theory. Figure 2.1 was obtained in this joint work. We thank Nathan Dunfield, Sergei Gukov, Yang-Hui He, Onkar Parrikar, and Edward Witten for discussions and comments on Chapter 2. We thank Dror Bar-Natan for computing HOMFLY-PT polynomials at 15 crossings and for the `dbnsymb` LaTeX symbol package [1].

In Chapter 3, we thank Dror Bar-Natan, Seungwon Kim, Scott Morrison, Dirk Schütz, Jackson Switzer, and Edward Witten for helpful discussions during the preparation of the manuscript.

In Chapter 4, we thank Yukai Wu, Luming Duan, and Dong-Ling Deng for helpful discussions on computing OTOCs using neural network states, as well as for sharing their numerical techniques.

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

Included publications

J. Craven, V. Jejjala and A. Kar, *Disentangling a Deep Learned Volume Formula*, *JHEP* **06** (2021) 040, [2012.03955]

Author contributions: Craven - $33\frac{1}{3}\%$, Jejjala - $33\frac{1}{3}\%$, Kar - $33\frac{1}{3}\%$.

J. Craven, M. Hughes, V. Jejjala and A. Kar, *Learning knot invariants across dimensions*, *arXiv preprint arXiv:2112.00016* (2021)

Author contributions: Craven - 25%, Hughes - 25%, Jejjala - 25%, Kar - 25%.

List of Figures

1.1 Neural network structure 2
1.2 Common activation functions 3
1.3 Trefoil and figure eight knots 7
1.4 Representations of the unknot 7
1.5 Reidemeister moves 8
1.6 A braid 8
1.7 Trefoil and figure eight braids 9
1.8 Figure eight knot, divided 12
1.9 Drawing edges on strands 12
1.10 Edges on strands, part 2 12
1.11 Edges for figure eight 13
1.12 Polyhedra for figure-8 knot 13
1.13 Figure-8 polyhedra, no bigons 13
1.14 Gluing rules for figure-8 knot 14
1.15 Trefoil knot in a manifold 18
1.16 Switching marked points of the trefoil 19
1.17 One loop diagram for SYK two-point function 24
1.18 SYK melon diagram 24
1.19 $\mathcal{O}(J^4)$ contributions to SYK two-point function 24

2.1 Quadratic relationship between $n = 2$ and $n = 3$ colored Jones polynomials at roots of unity. 31
2.2 An example of a neural network architecture 39
2.3 Layer-wise relevance propagation for neural networks predicting the hyperbolic volume from Jones polynomials. 42
2.4 Best fits for hyperbolic volumes as functions of the Jones polynomial compared to neural network predictions. 44
2.5 Fit parameters for $V(J_2(K; e^{3\pi i/4}))$ as a function of crossing number. . . . 45
2.6 Fits for hyperbolic volume as a function of $|J_2(K; -1)|$ for alternating and general knots. 45
2.7 Mean absolute errors as a function of x when optimizing the ansatz $V(K) = a \log(|J_2(K; e^{ix})| + b) + c$. 46

2.9 Modified, monotonic, volume conjecture sequence for the figure-eight knot. 52
2.A.1 Results for training a neural network to learn the hyperbolic volume from the Jones polynomial. 54
2.B.1 The absolute values of the (first 8) Jones polynomial coefficients as a function of the hyperbolic volume. 56
2.D.1 t-SNE plot of 2000 Jones polynomials, colored by volume. 62
2.D.2 t-SNE plot of 2000 Jones polynomials, colored by Chern–Simons invariant. 63
2.E.1 Eigenvalues of the weight matrix and the bias vector for a small neural network learning the hyperbolic volume from the Jones polynomial. 66

3.2.1 Two smoothings of a crossing. 71
3.4.1 Relationships in knot dataset (s, g, σ). 85
3.4.2 Accuracy of predicting s-invariant and slice genus from Khovanov polynomial for decreasing training fraction 86
3.4.3 Accuracy of predicting s-invariant and slice genus from Jones polynomial for decreasing training fraction 92
3.4.4 Interdependence of s-invariant and slice genus predictions. 93
3.4.5 Interdependence of s-invariant and signature predictions. 94

4.4.1 Computing an OTOC for an $N = 4$ SYK Hamiltonian 103

List of Tables

2.1 LRP results and relevant roots for neural networks predicting the hyperbolic volume from the Jones polynomial. 43
2.B.1 Fits for the mean absolute value of the coefficients as a function of the volumes. 57

3.4.1 Accuracy of neural network predictions for s-invariant and slice genus for $\mathrm{Kh}(q, -q^n)$ 87
3.4.2 Most common $\mathrm{Kh}(q, -q^{-2})$ in the dataset 88
3.4.3 The Khovanov homology of $K = 9_{20}$ 89
3.4.4 Knight and pawn moves for $\alpha q^m t^n (1 + tq^4)$ and $q^s(q + q^{-1})$. 90
3.4.5 A tabular representation of $\mathrm{Kh}(q, tq^{-2})$. 90
3.4.6 Learning s and g from the Jones polynomial evaluations. 92

Chapter 1

Introduction

Science fiction has become reality and artificial intelligence is an ubiquitous force in our society. We have created computational structures that can learn for themselves and set them loose on all manner of applications, including banking, crime prevention, and medicine. Physics is not immune to this trend and machine learning techniques show great promise as research tools, particularly in string theory. Machine learning gives us a novel way to explore physics and mathematics. Large, previously impermeable datasets, like the Calibu–Yau datasets and the extensive tables of knot invariants, have been successfully tackled using artificial intelligence. Additionally, neural network architectures often provide for a faster representation of computationally difficult problems.

This dissertation involves an intersection of machine learning, mathematics, and physics. The introduction is organised to provide a basic background of the relevant features of each field and to explain the intersection between them. Section 1.1 reviews the basics of machine learning and reviews how string theorists, and other physicists, have recently utilized machine learning in their research. Section 2.2 gives an overview of the mathematical objects called knots and summarizes the relevant features of knots. Section 1.3 reviews Chern–Simons theory and briefly discusses its connection to knots. Section 1.4 studies the physics of the SYK model and discusses our motivations for studying it.

1.1 Machine learning

A machine learning algorithm is an algorithm which uses training data to optimize an approximate mapping between input and output data. Machine learning has been applied to a huge variety of tasks with great success, including identifying suspects when a crime has been committed [4], medical diagnoses [5], natural language processing [6], and playing games like chess and Go [7; 8]. Artificial intelligence has

remarkably even been used to discover a collection of the most fundamental equations of physics, as cataloged in Feynman's Lectures on Physics [9].

In recent years, both mathematicians and physicists have started applying machine learning to very large data sets. These data sets are not hugely useful in their raw format, as we are unable to identify important relationships in the data ourselves. Machine learning is a powerful tool which can be used to identify and model relationships that humans are unable to perceive.

Although useful for identifying mappings, machine learning techniques are often not interpretable, meaning that we cannot always use machine learning results to draw satisfying conclusions or explanations. Throughout this work, the focus is on using machine learning as a first step in identifying new relationships in data. The next step is to employ interpretable machine learning techniques to extract meaningful conclusions from the identified relationships.

1.1.1 Feedforward neural networks

One machine learning architecture used in this work is a feedforward neural network. This system was originally inspired by the structure and functioning of brains, hence the term "neural". Neural networks consist of a series of layers, each containing individual nodes (also called neurons). The first layer is always an input layer and each node corresponds to an element of the input feature vector. The final layer is the output layer, which will output the prediction of the neural network. All of the layers between the input and output layers are called "hidden layers". A neural network is typically fully connected. This means that every node in layer i is constructed as a combination of every node in layer $i-1$. A simple example of a neural network is shown in Figure 1.1. This network consists of an input vector of length three, two hidden layers with five nodes each, and a single output node.

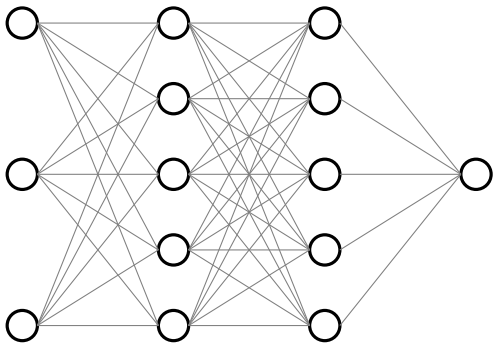


Figure 1.1: A simple neural network with two hidden layers.

The neural network has a set of parameters which are used to calculate each layer as a combination of the previous layer. These parameters are called weights

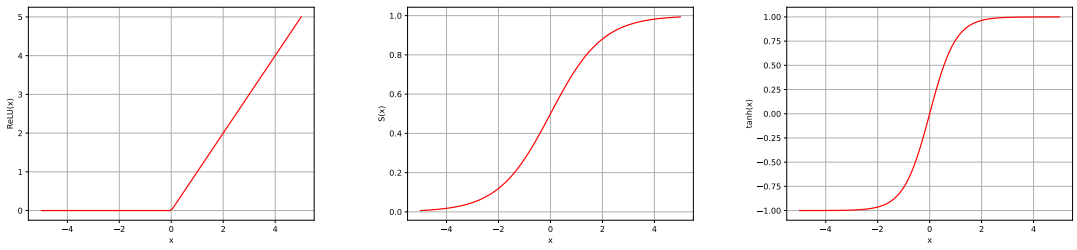


Figure 1.2: Some common activation functions: ReLU, logistic sigmoid, and hyperbolic tangent

and biases. Between the input layer and the first hidden layer in the example above, there is a 5×3 weight matrix and a 5×1 bias vector. The second layer is then calculated as

$$\mathbf{v}_2 = \sigma\Big(W_1 \cdot \mathbf{v}_1 + b_1\Big), \tag{1.1.1}$$

where σ is a chosen activation function. The activation function is used to introduce non-linearity into the neural network. The current standard is the ReLU (rectified linear unit) activation function, which is defined as

$$\sigma(x) = \begin{cases} 0 & x \leq 0 \\ x & x > 0. \end{cases}$$

Since each layer is computed in this way from the previous layer, the neural network is essentially a function that maps the input to the predicted output as:

$$f_\theta(\mathbf{v}_{in}) = \sigma\left(\ldots\sigma\Big(W_2 \cdot \big(\sigma(W_1 \cdot \mathbf{v}_i n + b_1)\big) + b_2\Big)\right). \tag{1.1.2}$$

The terminology f_θ implies that the neural network is approximating some true function f which relates the input and output using the parameters, which are denoted by θ.

In order to train the network, and then to test the accuracy of its predictions, we divide the data set into a training set and a testing set. The training set is used to tune the parameters of the neural network to match the predicted output to the true output as closely as possible. The network is then provided the testing data, which has not yet been encountered, to test the accuracy of the network's predictions. A standard split is $80-90\%$ training data, and $10-20\%$ testing data. In Chapters 2 and 3, we achieve high accuracy when training on $1-10\%$ of the data.

The network is trained by tuning the parameters (the weights and biases) in a way that minimizes a chosen loss function. One simple loss function is the mean squared error, defined as

$$\text{MSE} = \frac{1}{n}\sum_i |y_{\text{true},i} - y_{\text{pred},i}|^2. \tag{1.1.3}$$

The parameters in the neural network are initialized to some random numbers. The training data set is then passed through the network, and a set of predicted output values, $\mathbf{y}_{\text{pred}}$ is generated. The loss function is then computed for the predicted output. The neural network improves its predictions by adjusting its parameters to minimize the loss. The parameters are adjusted through optimization algorithms, typically gradient descent algorithms (and their modern successors, such as the Adam algorithm). To minimize the loss function, the training algorithm approximates the gradient of the loss function with respect to the network's parameters. The number of training examples passed through the network and the number of passes made through the training set are determined by the batch size and the epochs, respectively. The batch size refers to how many training examples are used in one iteration. The network parameters are adjusted after each batch. The epochs determine how many times the total training data set is passed through the neural network in training.

1.1.2 Restricted Boltzmann machines

A restricted Boltzmann machine (RBM) is, structurally, a simple neural network with two layers. The two layers are the visible (input) layer and the hidden layer. However, RBMs are often trained and used quite differently to neural networks. While RBMs can be used for regression and classification tasks, they are commonly used as generative models. This means they are used to find representations of data, like autoencoders. The training process is unsupervised and, as the name suggests, involves an interpretation in terms of some energy. Although our usage of RBMs, following [10], uses a supervised training algorithm called Stochastic Reconfiguration (SR), we will briefly summarize the typical training algorithm used on RBMs: the contrastive divergence algorithm. This discussion follows [11]. For details on the SR algorithm, please refer to Chapter 4.

In an RBM, the visible and hidden nodes are all binary-valued. We can imagine feeding the RBM a flattened black and white image, where each pixel is valued either 0 or 1. The nodes in the hidden layer are called "feature detectors". We assign to the joint configuration of nodes, $(\mathbf{v}, \mathbf{h})$, an energy

$$E(\mathbf{v}, \mathbf{h}) = \sum_{i \in \text{visible}} a_i v_i - \sum_{j \in \text{hidden}} b_j h_j - \sum_{i,j} v_i h_j w_{ij}, \tag{1.1.4}$$

where the a_i are the biases of the visible layer, b_j are the biases of the hidden layer, and w_{ij} are the weight matrix elements. Every possible pair of visible and hidden nodes is assigned a probability given by

$$p(\mathbf{v}, \mathbf{h}) = \frac{1}{Z} e^{-E(\mathbf{v},\mathbf{h})}, \tag{1.1.5}$$

where Z is the usual partition function in the canonical ensemble; $Z = \sum_{\mathbf{v},\mathbf{h}} e^{-E(\mathbf{v},\mathbf{h})}$. The probability that the network assigns to a visible vector, $\mathbf{v}$ is given by tracing out the hidden units,

$$p(\mathbf{v}) = \frac{1}{Z} \sum_{\mathbf{h}} e^{-E(\mathbf{v},\mathbf{h})}. \tag{1.1.6}$$

We would like to maximize the probability that the network assigns to a training image. To accomplish this, we adjust the weights and biases to minimize the energy of the training image and raise the energy of the other images. Taking the derivative of the log probability of $\mathbf{v}$ with respect to the weights, we get

$$\frac{\partial \log p(\mathbf{v})}{\partial w_{ij}} = \langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_{\text{model}}. \tag{1.1.7}$$

The first term is an expectation value calculated with respect to the distribution of the data, while the second is calculated with respect to the model. This formula can be used as an update formula for the weights, but computing expectation values with respect to the model is difficult. Instead, to train the RBM, we first set all of the visible units according to a training image. The hidden units are computed using

$$p(h_j = 1|\mathbf{v}) = \sigma(b_j + \sum_i v_i w_{ij}), \tag{1.1.8}$$

where $\sigma(x)$ is the logistic sigmoid. We then move backwards and attempt to reconstruct the visible layer by setting each v_i to 1 according to

$$p(v_i = 1|\mathbf{h}) = \sigma(a_i + \sum_j h_j w_{ij}). \tag{1.1.9}$$

The weights are then updated according to

$$\Delta w_{ij} = \epsilon \left(\langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_{\text{reconstruction}} \right), \tag{1.1.10}$$

where ϵ is the learning rate.

1.1.3 Machine learning in physics

Physicists have recently been using machine learning in their research, especially in the fields of particle physics, astrophysics, and string theory. For a full review of the string theory applications, see [12]. The success of machine learning in these many physics/string theory applications is motivation to apply machine learning

techniques to other problems in physics.

The following gives some representative examples of machine learning applications in string theory and is certainly not intended to be exhaustive. The use of data science techniques in string theory was introduced in 2017, when [13], [14], [15], and [16] discussed the application of machine learning techniques to study the string landscape. The authors of [17], [18], and [19] applied machine learning techniques to studying properties of the Calabi–Yau data sets. Further studies of Calabi–Yaus are contained in [20], [21], [22], [23]. Calabi–Yau metrics have also been a popular machine learning topic and are studied in [24], [25], [26], [27]. String standard models are studied in [28] using both supervised and unsupervised learning techniques. In [29], the authors used equation learning techniques to to study Calabi–Yau manifolds. In [30] machine learning is used to compute line bundle cohomologies. The authors of [31] use decision trees to predict non-Higgsable gauge groups. In [32], Generative Adversarial Networks (GANs) were used to generate superpotentials. In [33], autencoder neural networks were used to search for promising string models. In [34], Genetic Algorithms are employed to study high-dimensional Minimal Supersymmetric Standard Models (pMSSMs). In [35], [36], [37] the authors study the AdS/CFT correspondence using neural networks, and later RBMs.

Another area where machine learning techniques have been applied extensively is in condensed matter systems. Many-body quantum systems are very difficult to solve numerically, due to exponential complexity of the systems. In [10], the authors propose an RBM representation for many-body wave functions. Reinforcement learning is then used to find the ground state, or unitarily time evolved state, of the system. In Chapter 4, we explore the application of these techniques to the AdS/CFT correspondence via the SYK model (see Section 1.4 for an introduction to the SYK model).

1.2 Knot theory

A knot is an embedding of a circle into S^3. Knots are commonly represented as knot diagrams, which are projections of the three dimensional objects onto two dimensions. Knots are of interest in many areas, including physics (particularly in Chern–Simons theory) as well as chemistry and biology (to study knotted DNA molecules), see [38]. We will introduce the relevant properties of knots by considering some of the simplest nontrivial examples: the trefoil knot and the figure eight knot, shown in Figure 1.3. The figure eight knot is hyperbolic (see Section 1.2.4), while the trefoil knot is not.

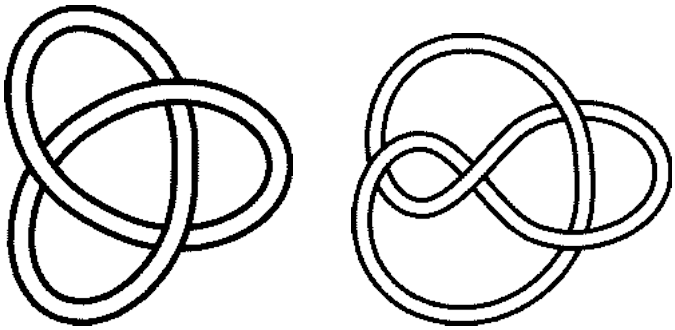


Figure 1.3: Left: the trefoil knot, 3_1. Right: the figure eight knot, 4_1. Figures taken from KnotPlot [39].

Knots are labelled by their positions in the tabulation of knots. For instance, the trefoil knot is also called 3_1. This means that it has three crossings and that it is the first of the three crossing knots (the crossing number of a knot is the minimal number of crossings in a diagram). Knots are classified by their invariants: functions of the knot which are the same no matter how the knot diagram is drawn. Knot invariants serve to distinguish between knots - if two knots are equivalent, then their invariants are the same (note that the converse is not necessarily true). In the original tabulation of knots, the knots were not all known to be distinct. In fact, this tabulation contained knots that were identical, drawn in different ways. One example of this phenomenon is shown in Figure 1.4. This highlights the importance of knot invariants - they allow us to truly distinguish between knots.

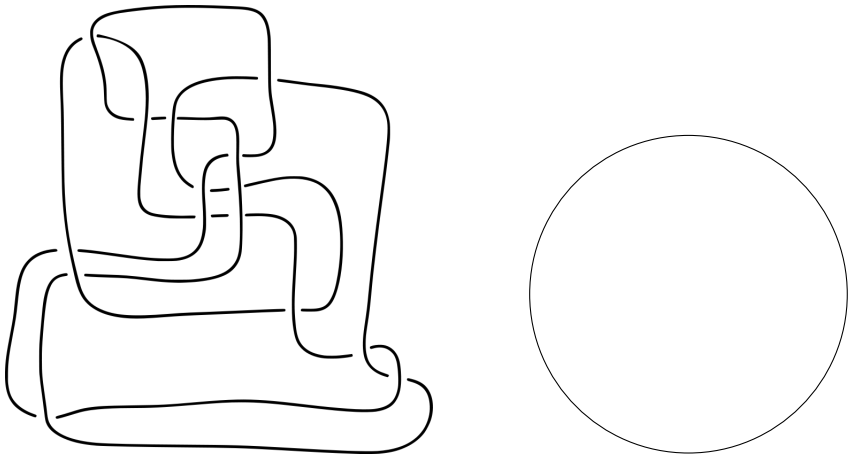


Figure 1.4: Left: the Thistlethwaite unknot. Figure taken from Wikimedia. Right: The unknot. These diagrams appear to represent different knots, but they are actually the same knot. This illustrates the importance of having invariants to identify knots.

1.2.1 Reidemeister moves

Reidemeister moves are a set of moves we can perform on a knot to change the relations between the crossings. Notably, while these moves do change the way

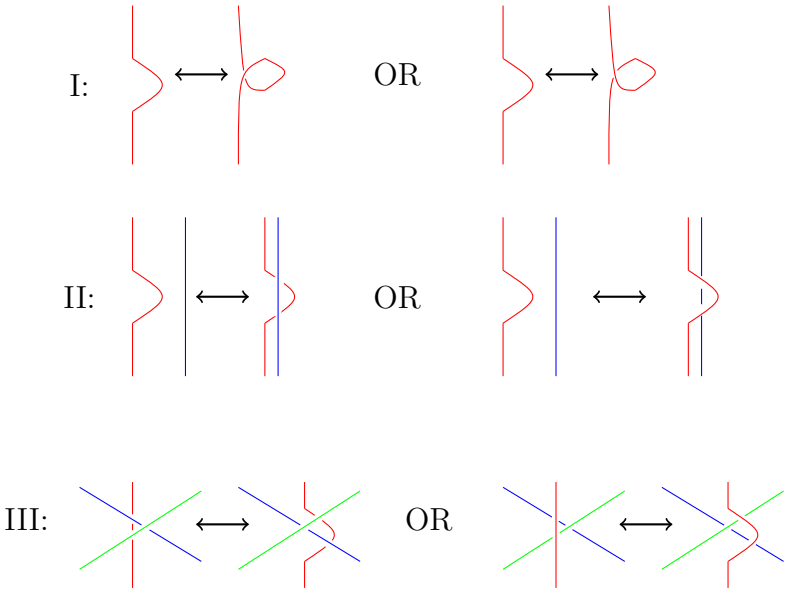


Figure 1.5: The three Reidemeister moves. These are ambient isotopies - they change crossing relations, but do not change the actual knot.

the knot is drawn (its projection), it does not change the knot itself. The moves are shown in Figure 1.5. Reidemeister moves are ambient isotopies - this refers to moving the strands of the knot through space, without letting it pass through itself. The first move allows us to add or remove a twist from the diagram. The second move allows us to add or remove two crossings. The third move allows us to slide a strand from one side of a crossing to the other.

1.2.2 Braids and braidwords

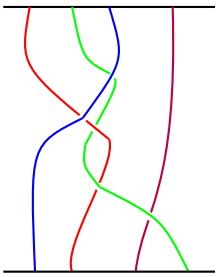


Figure 1.6: An example of a braid.

All knots can be represented as braids - a set of strings which are attached to horizontal "bars" at the top and bottom (see Figure 1.6). We can form the closure of a braid by gluing together the top and bottom bars of the braid. Imagining that there is an axis coming out of the page, in the center of the closure, a closed braid representation is one for which we can traverse the closure while always going clockwise around the axis.

The braid index of a knot is an invariant corresponding to the least number of strings in a closed-braid representation of the knot.

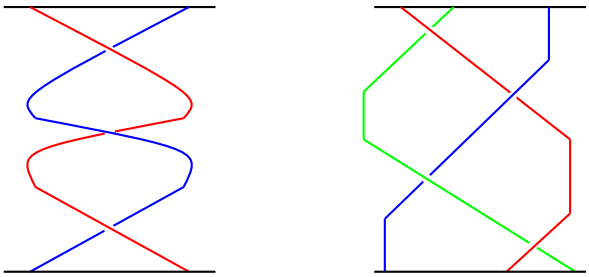


Figure 1.7: Left: Braid of the trefoil knot. Right: Braid of the figure-eight knot.

A braidword gives a description of a braid. As an example, we will compute the braidwords for the trefoil and the figure-eight knot (Figure 1.7). The general rule is that if string i, entering a crossing from the left, goes over string $i+1$, entering from the right, then we append σ_i to the braidword. Otherwise, we append σ_i^{-1}. Looking at the braids in Figure 1.7, we see that the braidword for the trefoil knot is $\sigma_1\sigma_1\sigma_1$, while the braidword for the figure-eight knot is $\sigma_1\sigma_2^{-1}\sigma_1\sigma_2^{-1}$.

The braid index is the number of strands required to form a representation of the knot. For instance, the trefoil knot has braid index $n = 2$ and the figure eight knot has braid index $n = 3$. Applying σ_i and σ_i^{-1} on the n strands defines a group; the braid group B_n.

1.2.3 The Jones Polynomial

A major knot invariant of interest to us is the Jones polynomial. The Jones polynomial was discovered by Jones in [40]. It is a Laurent polynomial with integer coefficients. The Kauffman bracket allows us to compute the Jones polynomial graphically, using the three rules:

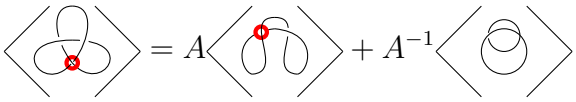
$$\left\langle \bigcirc \right\rangle = 1 \tag{1.2.1a}$$

$$\left\langle \times \right\rangle = A\left\langle \asymp \right\rangle + A^{-1}\left\langle) (\right\rangle \tag{1.2.1b}$$

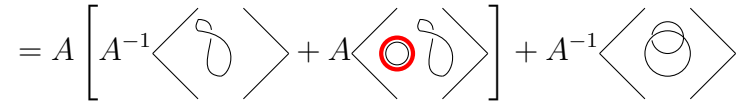
$$\left\langle K \cup \bigcirc \right\rangle = (-A^2 - A^{-2})\left\langle K \right\rangle. \tag{1.2.1c}$$

The first rule states that the Kauffman bracket of the unknot is 1. (The unknot is the trivial knot. It has zero crossings). The second rule gives us a way to undo crossings in a knot diagram. The third rule shows how to separate a diagram containing a knot disjoint from the unknot. Applying these rules recursively, until no brackets remain, gives a Laurent polynomial in A for the knot. We write this polynomial

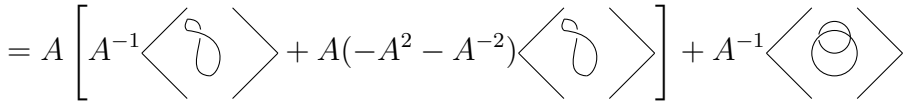
in the variable q by taking $A = q^{1/4}$. As an example, we will compute the Jones polynomial of the trefoil knot, which is the simplest nontrivial knot. The first step is to undo one of the crossings using the first rule in (1.2.1).



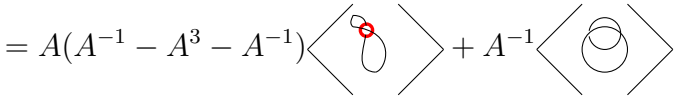
In the next step, we undo another crossing, circled in the first term on the right hand side of (1.2.3).

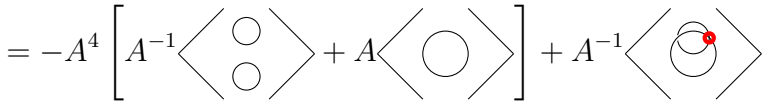


Next, we apply rule (1.2.1c) to remove the disjoint unknot in the Kauffman bracket.

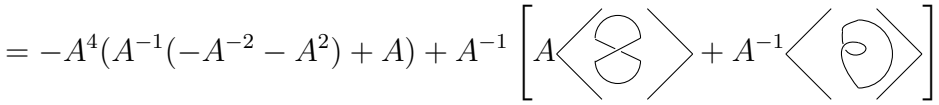


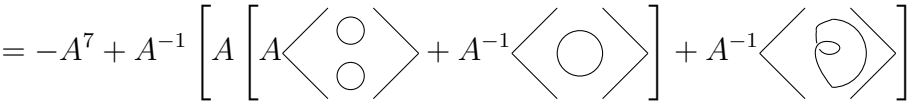
Now we target the final crossing in the first term and resolve the remaining unknots.

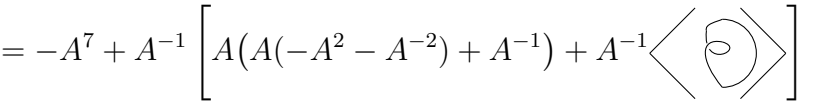


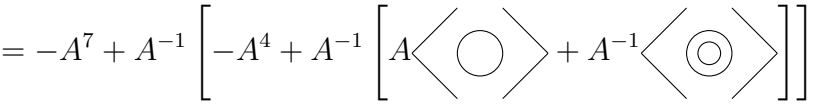


The first term's brackets have been resolved, so we move on to calculating the Kauffman bracket of the second term, focusing first on the crossing on the right.



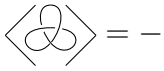






We have completed our calculation of the Kauffman bracket, as no brackets remain in the equation.

$$= -A^7 + A^{-1}\left[-A^4 + A^{-1}\left(A + A^{-1}(-A^2 - A^{-2})\right)\right]$$


$$= -A^7 - A^3 - A^{-5} \tag{1.2.2}$$

To calculate the Jones polynomial, we use the equation

$$J_2(K;q) = (q^{3/4})^{w(K)}\langle K\rangle, \tag{1.2.3}$$

where $w(K)$ is the writhe of the knot. We calculate this by summing the signs of all the crossings in a knot. The sign of a crossing is defined to be 1 if the overpass slants from bottom left to top right, or if it slants from top left to bottom right. It is -1 otherwise. For the trefoil knot, we see that $w(k) = 3$. Applying this equation and using the substitution $A = q^{1/4}$, we get

$$J_2(3_1, q) = -q + q^2 + q^3. \tag{1.2.4}$$

1.2.4 The hyperbolic volume

In Chapter 2, we are particularly interested in hyperbolic knots. Hyperbolic knots are those whose complement, $S^3 \setminus K$, admit a hyperbolic structure (it is a space of constant negative curvature). Here, $S^3 \setminus K$ is shorthand for $S^3 \setminus N(K)$, where $N(K)$ is an open tubular neighborhood of K in S^3. The vast majority of tabulated knots are hyperbolic: of the $1,701,936$ knots up to sixteen crossings, only 33 are not hyperbolic. The hyperbolic volume of a knot is defined as the volume of the knot complement. We will illustrate how to calculate the hyperbolic volume using the simplest example, that of the figure eight knot (see Figure 1.3, right). The basic procedure is to decompose the knot complement into two ideal polyhedra (polyhedra with the vertices removed). One polyhedron is constructed from the top of the knot diagram, and the other from underneath the knot diagram. The volume can then be calculated as the combined volumes of the ideal polyhedra. This procedure was initially invented by Thurston in [41]. Here we follow the discussion in [42].

The first step is to imagine that the knot diagram is dividing the space it sits in into sections, as illustrated in Figure 1.8. Each of these sections will become faces of the polyhedra that we are constructing. We will have one polyhedron sitting on top of the knot and one sitting below the knot. We can see from the figure that the faces meet at the crossings of the knot these will become edges of the polyhedra in our new construction. Imagine the edge as a vertical line that goes from the top strand of the crossing, down to the bottom strand of the crossing, as illustrated in Figure 1.9, left. We can draw each edge four times, as shown on the right of the Figure 1.9.

We notice from Figure 1.9 above that the endpoints of an edge, which are the vertices of our polyhedron, are on the knot. This is not good, since we are trying to

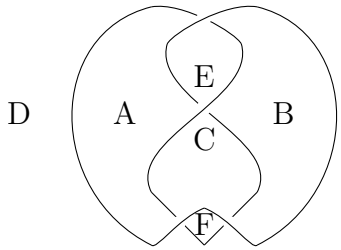


Figure 1.8: The figure-8 knot divides the space into six sections

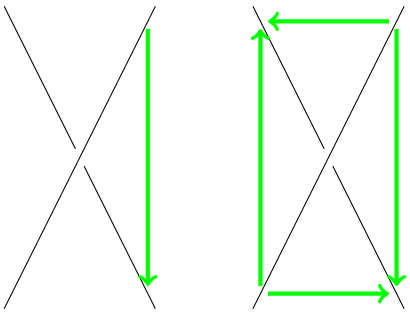


Figure 1.9: (Left) An edge of the polyhedron, drawn from top to bottom strand of the crossing. (Right) Four copies of the edge at a crossing.

construct the knot complement. We want to remove the vertices, so that the knot is not part of our new manifold. We do this by shrinking the strands of the knot down to points. For the top polyhedron, imagine sliding the tails of the edges in the right of Figure 1.9 along the top strand of the crossing, until all four endpoints are at the crossing. For the bottom polyhedron, imagine sliding the heads of the edges along the bottom strand of the crossing. The result is shown in Figure 1.10.

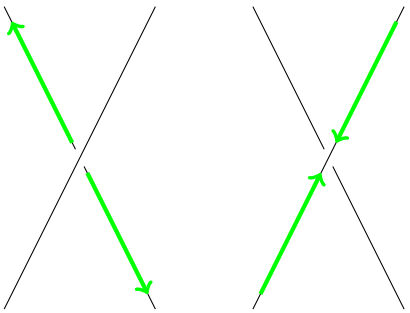


Figure 1.10: Moving the edges along the knot strands.

Applying this procedure to the figure eight knot, we get Figure 1.11 (pictured for the top polyhedron, the process for the bottom is similar). Now, we need to remove the knot by shrinking the strands to a point. We can pretend that we are stretching out the heads of each arrow in Figure 1.11 until it hits another arrowhead. This gives us the two polyhedra shown in Figure 1.12.

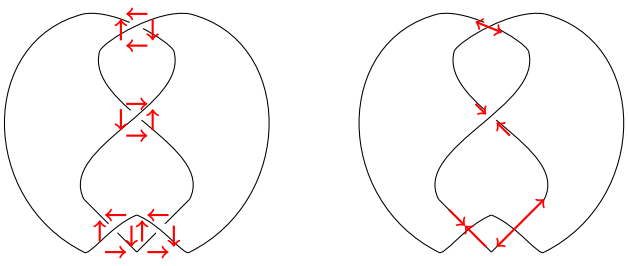


Figure 1.11: Drawing in the edges for the top polyhedron of the figure-8 knot

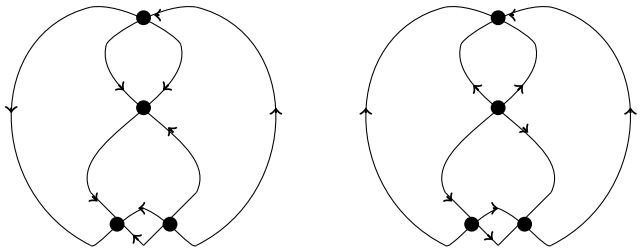


Figure 1.12: Polyhedra for the figure-8 knot.

In Figure 1.12, we see that there are a two faces that have only two edges: these are called bigons, and they can be removed. When we remove them, the polyhedra can be drawn as in Figure 1.13.

For the remainder of this calculation, we will be working with hyperbolic geometry. Some basic facts on hyperbolic geometries are listed below. The three dimensional hyperbolic manifold is defined as

$$\mathbb{H}^3 = \{(x+iy,t) \in \mathbb{C} \times \mathbb{R} | t > 0\}, \tag{1.2.5}$$

with metric

$$ds^2 = \frac{dx^2 + dy^2 + dt^2}{t^2}. \tag{1.2.6}$$

This metric is the Wick rotated metric of AdS_3 in the Poincar patch. The geodesics in hyperbolic space are vertical lines and semi-circles orthogonal to the boundary, $\partial\mathbb{H}^3 = \mathbb{C} \cup \{\infty\}$. The geodesic planes are vertical planes and hemispheres centered

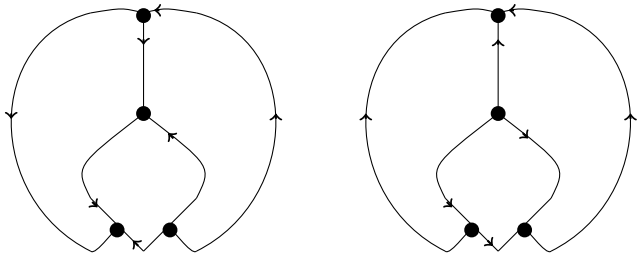


Figure 1.13: Polyhedra with bigons removed.

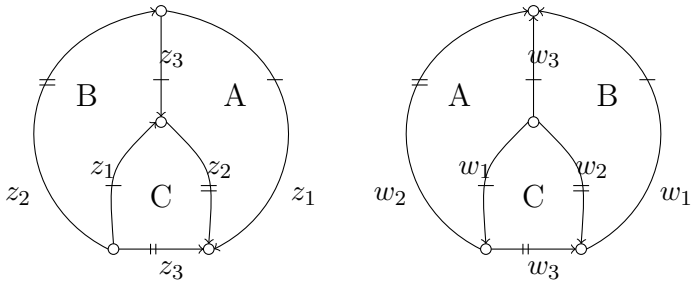


Figure 1.14: How to glue together the two tetrahedra that make up the figure-8 knot complement.

on $\mathbb{C}$.

An ideal tetrahedron is a tetrahedron where all four vertices are on $\partial\mathbb{H}^3$ and where all faces and edges are geodesics. To calculate the hyperbolic volume of our knot complement, we need to decompose it into ideal tetrahedra (completed above), determine a consistent way to glue together the edges of the tetrahedra, and lastly compute a hyperbolic structure on the knot complement.

To find the edge gluing equations, we need to define the idea of an edge invariant. An edge invariant $z(e)$ of an edge e is the complex number obtained by applying the unique isometry of $\mathbb{H}^3$ which takes the vertices of e to 0 and ∞, while taking the other two vertices of the tetrahedron to 1 and $z(e)$. We note, without proof, the following important properties of the edge invariants of an ideal tetrahedron. Suppose T has an edge e_1, with the vertices of e_1 at 0 and ∞, and the remaining two vertices at 1 and $z(e_1)$. Then, the edge opposite to e_1 has edge invariant $z(e_1)$, e_2 has edge invariant $z(e_2) = \frac{1}{1-z(e_1)}$, and e_3 has edge invariant $z(e_3) = \frac{z(e_1)-1}{z(e_1)}$.

For the knot complement to admit a hyperbolic structure, we need to make sure that for every edge e, the following equations are obeyed (where the products/sums run over the edges being glued to e):

$$\prod z(e_i) = 1 \quad \text{and} \quad \sum \arg(z(e_i)) = 2\pi. \tag{1.2.7}$$

For the figure eight knot, we identify the edges as shown in Figure 1.14. This gives the matching equations

$$z_1^2 z_3 w_1^2 w_3 = 1, \tag{1.2.8}$$

and

$$z_2^2 z_3^2 w_2^2 w_3 = 1. \tag{1.2.9}$$

We can use the first equation to solve for z_1 in terms of w_1. We can then impose the restriction that $\mathrm{Im}(z_1)$ and $\mathrm{Im}(w_1)$ must be positive. This gives us, as one solution,

$$z = w = \sqrt[3]{-1} = -\frac{1}{2} + \frac{\sqrt{3}}{2}i. \tag{1.2.10}$$

To prove that the structure on the knot complement is complete, we need to solve what are known as the completeness equations. We omit this section, but it fixes z and w to be precisely those given above.

The last step is to calculate the hyperbolic volume. The volume of an ideal tetrahedron is given by

$$\mathrm{Vol}(T) = \Lambda(\arg(z_1)) + \Lambda(\arg(z_2)) + \Lambda(\arg(z_3)), \tag{1.2.11}$$

where $\Lambda(z)$ is the Lobachebsky function, defined as

$$\Lambda(\theta) = -\int_0^\theta \log(2\sin u)du. \tag{1.2.12}$$

The hyperbolic volume of the knot complement is given by the sum of the volumes of the top and bottom tetrahedra, so

$$\mathrm{Vol}(4_1) = \mathrm{Vol}(T_1) + \mathrm{Vol}(T_2) = 6\Lambda(\pi/3) \approx 2.02988. \tag{1.2.13}$$

1.2.5 The slice genus and the Poincaré conjectures

The Poincaré conjecture states that every simply connected, closed 3-manifold is homeomorphic to the 3-sphere. One of the millennium prize problems, this conjecture was proven nearly a century after it was first stated, by Perelman [43; 44; 45]. The generalized Poincaré conjecture states that every n-manifold which is homotopy equivalent to the n-sphere is isomorphic to the n-sphere. This conjecture has been proven or disproven in various contexts, except for the case in four dimensions - this version of the conjecture states that a homotopy 4-sphere is diffeomorphic to the 4-sphere. This conjecture is known as the smooth 4-dimensional Poincaré conjecture.

If we define a knot $K \in S^3 = \partial D^4$, then that knot is slice if it bounds a disc in the four-dimensional space. There is a knot invariant, called the slice genus, which determines whether a knot is slice. This invariant is the smallest integer g such that K is the boundary of a (connected, orientable) 2-manifold S of genus g embedded in D^4. Since a disk has $g = 0$, if a knot is slice, it has $g = 0$. For any $g > 0$, the knot is not slice. It has been proposed (see [46], [47]) that slice knots could be used to disprove the smooth 4-dimensional Poincaré conjecture. Consider a homotopy 4-sphere, W, and a knot $K \in S^3 = \partial(W \setminus B^4)$. Suppose that K bounds a disk in $W \setminus B^4$, but that it is not slice in B^4. Then W would be homeomorphic, but not diffeomorphic to S^4 - this would provide a counterexample to the conjecture. In [48], the authors propose that this strategy could be pursued by generating knots which

bound smooth discs in homotopy 4-balls. If one can prove that any of these knots are slice, then that knot will provide a counterexample to the smooth 4-dimensional Poincaré conjecture.

In Chapter 3, we employ machine learning techniques to try to learn the slice genus of a knot.

1.3 Chern–Simons theory and knots

Chern–Simons theory is an example of a topological quantum field theory - it is defined by a Lagrangian which does not require a choice of metric on the manifold. This means that any quantities we calculate will be topological invariants - much like the knot invariants introduced in the previous section. Our guiding principle in choosing our Lagrangian is that we want a generally covariant theory - a quantum field theory defined without any *a priori* choice of a metric on the 3-manifold M. In three dimensions, one choice is the Chern–Simons three form,

$$S_{CS} = \frac{k}{4\pi}\int_M \mathrm{Tr}\left(A\wedge dA + \frac{2}{3}A\wedge A\wedge A\right). \tag{1.3.1}$$

Here, A is a Lie algebra valued connection and the trace is taken over the gauge indices. The Lagrangian can also be written where the Lorentz indices on A are made explicit:

$$S_{CS} = \frac{k}{8\pi}\int_M \epsilon^{ijk}\mathrm{Tr}\left(A_i\left(\partial_j A_k - \partial_k A_j\right) + \frac{2}{3}A_i\left[A_j, A_k\right]\right). \tag{1.3.2}$$

For the theory to be consistent, we require that the quantity $e^{iS_{CS}}$ be gauge invariant (since gauge symmetries are just redundancies in our description, we require physical observations to be invariant under gauge transformation). If we consider a gauge transformation

$$A_\mu \to g^{-1}A_\mu g + g^{-1}\partial_\mu g, \tag{1.3.3}$$

then the change in the action is

$$\Delta S_{CS} = \frac{k}{4}\int d^3x\epsilon^{\mu\nu\rho}\left[\partial_\mu\mathrm{Tr}\left((\partial_\nu g)g^{-1}A_\rho\right) + \frac{1}{3}\mathrm{Tr}\left(g^{-1}\partial_\mu g g^{-1}\partial_\nu g g^{-1}\partial_\rho g\right)\right]. \tag{1.3.4}$$

The first term is a total derivative, which can be made to drop out, and the second term is $2kw(k)$, where $w(k)\in\mathbb{Z}$ is the winding number. We then see that for $e^{iS_{CS}}$ to be invariant, k must be an integer value. This is known as the level of the theory.

Chern–Simons theory is a useful model in quantum field theory because, as a topological field theory, many difficult quantum field theory problems become more easily manageable. In addition, Chern–Simons theory has found applications in

many corners of physics. For instance, in condensed matter physics, Chern–Simons theory can be used to study the fractional quantum Hall effect [49]. In the quantum Hall effect, the number of ground states depends on the topology of the manifold - there are m^g ground states on a surface with genus g, where m is the number of ground states on the torus. This "topological order" of the system is characterized by Chern–Simons theory. Chern–Simons theory also, historically, gives us the first example of holography: the three–dimensional Chern–Simons theory is dual to a two–dimensional Conformal Field Theory (CFT) (to be exact, Chern–Simons theory on a 3-manifold M is dual to a rational CFT or WZW model on ∂M). Further examples of applications of Chern–Simons theory include topological string theories and theories of three–dimensional gravity [50].

To completely define our quantum field theory, we need to make a choice of gauge invariant observables. Since the theory is topological. We use Wilson lines because they do not require a choice of metric, and because it is these observables that lead us to a definition of the Jones polynomial. The Wilson line is defined as

$$W_R(C) = \mathrm{Tr}_R P \exp \oint_C dx^i A_i \,, \tag{1.3.5}$$

where C is an oriented closed curve (for example, a knot) and R is an irreducible representation of the gauge group G (of which A is a connection). P denotes path ordering. In a quantum field theory, the observables of interest are correlation functions. Since Chern–Simons theory is topological, the observables are all metric independent, meaning that for operators $\mathcal{O}_i$,

$$\frac{\delta}{\delta g_{\mu\nu}} \langle \mathcal{O}_1 \dots \mathcal{O}_n \rangle = 0. \tag{1.3.6}$$

Witten showed in [51] that the Jones polynomial of a knot can be computed as the vacuum expectation value of the Wilson loop operator along that knot.

1.3.1 The Jones polynomial in Chern–Simons theory

The major claim of Witten's paper [51] is that the Jones polynomial of a knot can be calculated as the unknot normalized vacuum expectation value of a Wilson loop operator, as in (1.3.5):

$$J_2(K;q) = \frac{\int D\mathcal{A} \exp(i\mathcal{S}_{CS}) W_R(K)}{\int D\mathcal{A} \exp(i\mathcal{S}_{CS})}, \tag{1.3.7}$$

where R needs to be the fundamental representation of $SU(2)$. In this section, we will demonstrate how these path integrals can be calculated. We follow Witten's original work in [51], as well as the discussion in [52].

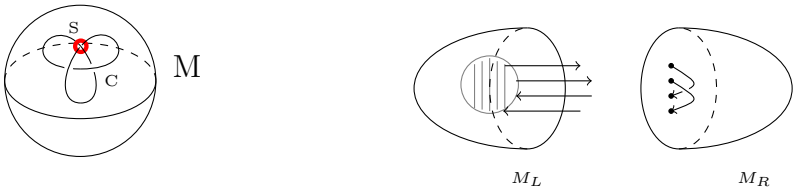


Figure 1.15: Left: The trefoil knot embedded in a sphere. Right: Isolating a crossing and splitting the sphere into two parts.

Suppose that we have a knot K embedded in a manifold, M. As an example, we would eventually like to consider $M = S^3$ with a gauge group $SU(2)$ and the knot $K = 3_1$. We start by surrounding one crossing with a sphere and "cutting it out" of the manifold. The manifold M is the connected sum of the manifolds M_L and M_R, who both have boundary S^2. This situation is depicted in Figure 1.15. We write

$$M_L = S^2 \cup B_R, \tag{1.3.8}$$

$$M_R = M \setminus B_R, \tag{1.3.9}$$

where B_R is a 3−ball bounded by S^2. M_R contains our isolated crossing, and M_L contains the rest of the knot. The points where K intersect S^2 are called the "marked points". We define two two-dimensional Hilbert spaces, $\mathcal{H}_L$ and $\mathcal{H}_L$, which are associated to ∂M_L and ∂M_R, respectively. The path integral on M_L determines a vector χ in $\mathcal{H}_L$, while the path integral on M_R determines a vector ψ in $\mathcal{H}_R$. The total path integral gives the pairing

$$Z(K) = (\chi, \psi). \tag{1.3.10}$$

We know that in a two-dimensional vector space, any three vectors are linearly dependent. This means that for three vectors ψ, ψ_1, and ψ_2, we have:

$$\alpha\psi + \beta\psi_1 + \gamma\psi_3 = 0. \tag{1.3.11}$$

We apply this to our current problem by generating two more manifolds M'_R and M''_R, which are the same as M_R, but with different braids connecting the marked points. This is demonstrated in Figure 1.16.

We may deduce from the above that

$$\alpha(\chi, \psi) + \beta(\chi, \psi_1) + \gamma(\chi, \psi_2) = 0, \tag{1.3.12}$$

which implies the relationship for the path integrals:

$$\alpha Z(K) + \beta Z(K_2) + \gamma Z(K_3) = 0, \tag{1.3.13}$$

where $Z(K_2)$ is the path integral on M'_R. and $Z(K_3)$ is the path integral on M''_R. We are now arriving at something similar to the skein relations from Section 2.2.

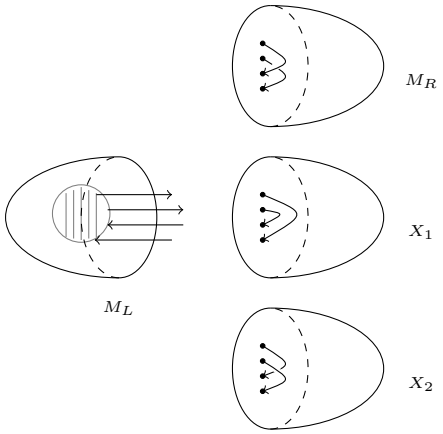


Figure 1.16: Switching marked points to generate the new manifolds.

We define on M_R a diffeomorphism B which will be used to generate M'_R and M''_R as

$$\psi_1 = B\psi, \tag{1.3.14}$$

$$\psi_2 = B^2\psi. \tag{1.3.15}$$

The action of B is to swap marked points. Unfortunately, B does not preserve the framing of a knot[1] - it introduces a twist in this framing. This means that every time we apply B, we introduce an error into our calculation. To undo this error, we apply a Dehn twist, which is defined as

$$D\psi = \exp\left(\frac{-3\pi i}{2(2+k)}\right) \tag{1.3.16}$$

for $N = 2$. We now want to use B to create an equation that looks like (1.3.13). We know that B acts on a two-dimensional vector space, so its characteristic equation is given by

$$B^2 - \mathrm{Tr}(B)B + \mathrm{det}B = 0. \tag{1.3.17}$$

Multiplying through by ψ, we get

$$\psi_2 - \mathrm{Tr}(B)\psi_1 + \mathrm{det}B\psi = 0. \tag{1.3.18}$$

We can then identify $\alpha = \mathrm{det}B$, $\beta = -D\mathrm{Tr}B$, and $\gamma = D^2$ (applying the appropriate number of Dehn twists). Witten uses results from conformal field theory to give the eigenvalues of B as

[1]The framing is a vector field defined on K such that $v(p) \notin T_p of K$ for all $p \in K$. We start with standard framing, which is a unit vector field orthogonal to the plane of projection of the link, oriented towards the plane of projection.

$$\lambda_1 = \exp\left(\frac{i\pi(-N+1)}{N(N+k)}\right), \quad \lambda_2 = \exp\left(\frac{i\pi(N+1)}{N(N+k)}\right). \tag{1.3.19}$$

We can then compute $\alpha = \lambda_1 + \lambda_2$, $\beta = D(\lambda_1 - \lambda_2)$, and $\gamma = D^2$ (specializing to the case where $G = SU(N)$ and $N = 2$), resulting in

$$\alpha = -\exp\left(\frac{\pi i}{2+k}\right), \tag{1.3.20}$$

$$\beta = \exp\left(\frac{-2\pi i}{(2+k)}\right) - 1, \tag{1.3.21}$$

$$\gamma = \exp\left(\frac{-3\pi i}{2+k}\right). \tag{1.3.22}$$

If we introduce the variable q as $q = \exp(2\pi i/(2+K))$, we can write the Skein relation as

$$-qZ(K) + (q^{1/2} - q^{-1/2})Z(K_1) + q^{-1}Z(K_2) = 0. \tag{1.3.23}$$

Finally, we need the normalization

$$\langle 0_1 \rangle = \frac{q - q^{-1}}{q^{1/2} - q^{-1/2}}. \tag{1.3.24}$$

We can then compute the Jones polynomial of a knot as

$$J_2(K, q) = \frac{Z(K)}{\langle 0_1 \rangle}. \tag{1.3.25}$$

Iteratively applying the procedure outlined above, we get the usual formula for the trefoil knot, as in Section 1.2.3. More details on Chern–Simons theory are discussed in Chapter 2.

1.4 The SYK model

The Sachdev–Ye–Kitaev (SYK) model is a model of Majorana fermions in $0 + 1$ dimensions. This model is useful in studying the near horizon region of near extremal black holes. In Section 1.4.1, we will introduce some important ideas of black holes and of the AdS/CFT correspondence to motivate interest in the SYK model. In Section 1.4.2, we will define the model and explore its solutions. In Section 1.4.3 we introduce the notion of a theory being "maximally chaotic" and apply these ideas to the SYK model.

1.4.1 Black holes and holography

Black holes are perhaps one of the most famous objects in physics. They arise, in the simplest case, as solutions to the vacuum Einstein equations,

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 0, \tag{1.4.1}$$

in spherically symmetric spacetimes (corresponding to the Schwarzchild metric). There are other black hole solutions which possess properties like angular momentum, spin, and charge. The most famous property of black holes is that, classically, any object which passes its event horizon will be unable to leave again. However, as Hawking discovered [53] when quantum mechanical effects are considered, black holes do in fact emit particles. This means that black holes are actually thermal objects, with entropy which scales like the area of the black hole. The "no hair" theorem states that (in four dimensions) the mass M, electric charge Q, and angular momentum J uniquely determine the metric of a black hole. If two quantum systems with the same M, Q, and J were to collapse into black holes, they would have identical spacetimes, according to the no hair theorem. However, this violates the conservation of information that we expect in quantum mechanics: an external observer can not recover information about the initial states. This is the famous black hole information paradox.

The study of black holes is connected to quantum field theories, including the SYK model, via the AdS/CFT correspondence [54], [55], [56]. This correspondence describes a duality between gravitational theories in $d+1$ Anti de Sitter (AdS) space, and quantum field theories called Conformal Field Theories (CFTs) in d dimensions. CFTs are quantum field theories which possess, in addition to the usual Poincaré invariance, an invariance under conformal transformations. (Conformal transformations of the coordinates leave the metric invariant up to a scale. In other words, conformal transformations leave angles invariant.) There is a dictionary which maps quantities in the gravity theory onto quantities in the CFT. This is extremely useful, especially since the strongly coupled CFT is dual to the weakly coupled gravity theory. It means that we may calculate quantities in the weakly coupled gravity theory and then map them back to the CFT, giving us results at strong coupling which we may not have been able to calculate otherwise.

The SYK model becomes a CFT in the infrared (IR) limit (see Section 1.4.2 for more details). SYK is a $0+1$-dimensional field theory, which means that it is dual to AdS_2; which is the geometry of the near-horizon region of extremal black holes (black holes with the minimal mass that is compatible with their charge and angular momentum). As an example of an extremal black hole, consider the Reisner–Nordström black hole solution [57]. It has energy $E = M - \frac{Q}{l_P}$. The thermodynamic

description of black holes requires that $E \gg T_H$, where T_H is the energy of a quanta emitted from the black hole:

$$T_H = \frac{1}{2\pi}\left(\frac{2}{l_P Q^3}E\right)^{1/2} + O(E^{3/2}). \tag{1.4.2}$$

When

$$E \sim \frac{1}{l_P Q^3}, \tag{1.4.3}$$

then $T_H \sim E$, so the thermodynamic description does not make sense. The energy (1.4.3) is known as the mass gap - it is the gap between the ground states and the next excited states of the black hole. Because the black hole has a mass gap, the SYK model essentially describes the ground states of the black hole.

1.4.2 The SYK model

This section follows the review in [57]. The SYK model describes a system of N Majorana fermions (charge neutral, spin-$\frac{1}{2}$ particles) with all-to-all interactions. The Hamiltonian is given by

$$H = \frac{1}{4!}\sum_{i,j,k,l=1}^{N} J_{ijkl}\psi_i\psi_j\psi_k\psi_l, \tag{1.4.4}$$

where the ψ_i obey the anticommutation relation

$$\{\psi_i, \psi_j\} = \delta_{ij}. \tag{1.4.5}$$

The couplings J_{ijkl} are random, and are drawn according to a normal distribution with zero mean and variance $\sigma = \sqrt{(3)!}J/N^{\frac{3}{2}}$. In what follows, we will restrict to $N = 2K$. We want to build representations of the Clifford algebra, (1.4.5). These representations should also be Hermitian ($\psi_i = \psi_i^\dagger$). This can be accomplished by defining the operator

$$c_i = \frac{1}{\sqrt{2}}(\psi_{2i} - i\psi_{2i+1}), \;\; i = i, \ldots, K. \tag{1.4.6}$$

These operators have the usual interpretation as annihilation operators. By applying the conjugate creation operators to the vacuum,

$$(c_1^\dagger)^{n_1}\ldots(c_k^\dagger)^{n_k}\left|0\right\rangle; \;\; n_k = 0, 1, \tag{1.4.7}$$

we can build our desired representations of the Clifford algebra. For numerical purposes, it is useful to define the ψ_is with the following recursion relation:

$$\psi_i^{(K)} = \psi_i^{(K-1)} \otimes \sigma_z, \;\; i = 1, \ldots, N-2 \tag{1.4.8a}$$

$$\psi_{N-1}^{(K)} = \frac{1}{\sqrt{2}} I_{2K-1} \otimes \sigma_x, \tag{1.4.8b}$$

$$\psi_N^{(K)} = \frac{1}{\sqrt{2}} I_{(2K-1)} \otimes \sigma_y. \tag{1.4.8c}$$

Here, the σ_i are the usual Pauli matrices. To illustrate how this computation works, we compute the first two levels of the recursion relation. To start with, we choose the initial matrices

$$\psi_1^{(1)} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \tag{1.4.9a}$$

$$\psi_2^{(1)} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{1.4.9b}$$

This is the $K = 1$ level, so the above equation gives us $\psi_{N-1}^{(K)}$ and $\psi_N^{(K)}$. We can then move on to the next step, $K = 2$. First, we use (1.4.8a) to calculate $\psi_1^{(2)} = \psi_1^{(1)} \otimes \sigma_z$, $\psi_2^{(2)} = \psi_2^{(1)} \otimes \sigma_z$. Then, we use (1.4.8b) to calculate $\psi_3^{(2)} = \frac{1}{\sqrt{2}} I_3 \otimes \sigma_x$. Finally, we can use (1.4.8c) to calculate $\psi_4^{(2)} = \frac{1}{\sqrt{2}} I_3 \otimes \sigma_y$.

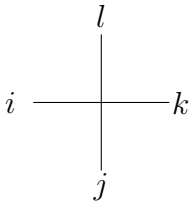
Now that we have defined the model, we can move on to studying the solutions. We do this by performing perturbation theory in the coupling J. We will start with the two-point function,

$$G_{ij}(\tau) = \langle T\{\psi_i(\tau)\psi_j(0)\}\rangle \,. \tag{1.4.10}$$

For $J = 0$, $\psi_i(\tau) = \psi_i(0)$, so

$$G_{ij}^{\text{free}} = \langle T\{\psi_i\psi_j\}\rangle = \frac{1}{2}\delta_{ij}\text{sign}(\tau). \tag{1.4.11}$$

Since the J_{ijkl}s are random, we will calculate Feynman diagrams by averaging over different realizations of the model, with different random couplings. For every member of the ensemble of models, we have an interaction vertex



$$\sim J_{ijkl}. \tag{1.4.12}$$

Consider the one-loop diagram, Figure 1.17. Here, the average is over different realizations of the model. Since it has a single vertex, it is linear in J_{ijkl}, so it evaluates to zero. The next diagram is the so-called melon diagram, Figure 1.18. The dotted line has a unique interpretation. It is not a particle, but an average over all possible couplings. When we compute the average over J of Figure 1.18, we get $3!\frac{J^2}{N^3} G_{ii}^{\text{free}} G_{jj}^{\text{free}} G_{kk}^{\text{free}} \delta_{l_1 l_2}$ (here we used $\langle J_{ijkl} J_{abcd}\rangle_J = 3!\frac{J^2}{N^3}\delta_{ia}\delta_{jb}\delta_{kc}\delta_{ld}$).

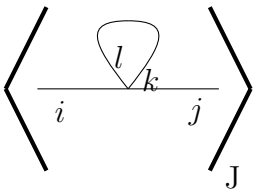


Figure 1.17: The first contribution to the two-point function.

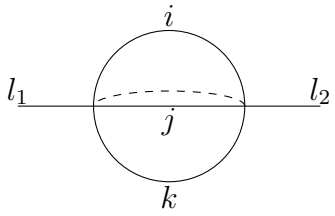


Figure 1.18: The melon diagram

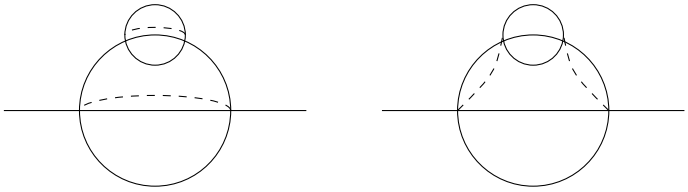


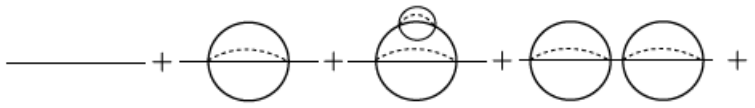
Figure 1.19: Left: $O(J^4)$ diagram with pairings within melons. Right: $O(J^4)$ diagrams with pairings between melons.

Lastly, we will consider possible $O(J^4)$ diagrams. There are two relevant classes of diagrams: ones where the pairings happen inside a melon, and ones where pairings move between melons, as in Figure 1.19. These evaluate to

$$\frac{1}{2}J^4(G^{\text{free}})^6\delta_{k_3k_4}, \tag{1.4.13a}$$

$$\frac{1}{4N^2}J^4(G^{\text{free}})^4G^{\text{free}}_{k_3k_4}. \tag{1.4.13b}$$

We see that 1.4.13b is suppressed by N^2, while 1.4.13a is not suppressed by any powers of N. This extends generally: the only relevant diagrams at large N are those whose vertices are inside a single melon. So, we may write $G(\tau)$ as:

$G(\tau) =$  $+ \ldots$ (1.4.14)

We can reformulate the solution for the two-point function in terms of the two equations:

$$G = (\partial_\tau - \Sigma)^{-1}, \tag{1.4.15a}$$

$$\Sigma(\tau,\tau') = J^2\left(G(\tau,\tau')\right)^3. \tag{1.4.15b}$$

At low energies, these equations can be written as

$$\int d\tau'' G(\tau,\tau'')\Sigma(\tau'',\tau') = -\delta(\tau-\tau'), \tag{1.4.16a}$$

$$\Sigma(\tau,\tau') = J^2 G(\tau,\tau')^3. \tag{1.4.16b}$$

The equations above are invariant under $\tau \to \phi(\tau)$, as long as G and Σ transform as conformal two-point functions with dimensions $1/4$ and $3/4$, respectively. We have seen that with taking an average over ensemble of SYK theories, each of which has a different Hamiltonian due to the different couplings, produces a result that is dual to gravitational physics. This is very different from the usual picture of AdS/CFT, where we have a single CFT dual to gravity. Witten showed in [58] that this averaging may not be necessary, as there are SYK-like models that do not require taking the disorder average.

1.4.3 Out of time ordered correlators and chaos

This section follows discussion in [59]. A system is said to be chaotic if it is highly sensitive to their initial conditions - a slight change in initial conditions can result in a totally different outcome. The commutator

$$[W(t), V(0)] \tag{1.4.17}$$

is used to diagnose chaos in a quantum system. It can measure the impact of perturbations by V on measurements of W at a later time. We say that a quantum system is strongly chaotic if

$$C(t) = -\langle [W(t), V(0)]^2 \rangle \tag{1.4.18}$$

is of order $2\langle VV\rangle\langle WW\rangle$ at late times. We can understand how (1.4.18) describes chaos by considering the classical limit. Suppose we have a one-particle system, where $V = p$ is the momentum of the particle and $W(t) = q(t)$ is the position of the particle. The commutator turns into a Poisson bracket, and we see that

$$\{q(t), p\} = \frac{\partial q(t)}{\partial q(0)}. \tag{1.4.19}$$

This is the classical measure of chaos: how strongly the final configuration of the particle depends on changes in the initial configuration. If the system is strongly chaotic, then similar initial positions will diverge exponentially; $e^{\lambda_L t}$, where λ_L is called the Lyapunov exponent.

Chaos in thermal quantum systems with many degrees of freedom can also be characterized by the Lyapunov exponent. To study chaos in these systems, we use the function

$$F(t) = \mathrm{tr}[yVyW(t)yVyW(t)], \tag{1.4.20}$$

where $y = \frac{1}{Z}e^{-\beta H}$. We see that the V operators are acted on with Euclidean time evolution and the W operators with Lorentzian time evolution. At early times, we have $F(t) \sim e^{\lambda_L t}$.

It was shown in [59] that there is a bound on the chaos of large N thermal quantum systems, given by

$$\lambda_L \leq \frac{2\pi k_B}{\beta\hbar}. \tag{1.4.21}$$

The SYK model is known to saturate this bound, which means that it is a maximally chaotic system. A CFT which saturates the chaos bound signals black hole physics in the bulk spacetime which is dual to that CFT.

1.5 Structure of dissertation

Four chapters follow from this introduction. In Chapter 2, we discuss the application of machine learning to studying the relationship between two knot invariants: the Jones polynomial and the hyperbolic volume. We discuss our strategy for reverse engineering the neural network and deriving a formula which reproduces the network's accuracy. We also provide an interpretation of the results in analytically continued Chern–Simons theory. In Chapter 3, we use neural networks to uncover relationships between polynomial invariants (the Jones and Khovanov polynomials) and the s-invariant and slice genus of knots. We further propose a gauge-theoretic interpretation of these results. In Chapter 4, we present a framework for studying OTOCs of the SYK model using neural network quantum states and discuss the applications of this technique to studying quantum chaos. In Chapter 5 we conclude and discuss the possible future of machine learning as a tool in theoretical physics.

Chapter 2

Disentangling a deep learned volume formula

We present a simple phenomenological formula which approximates the hyperbolic volume of a knot using only a single evaluation of its Jones polynomial at a root of unity. The average error is just 2.86% on the first 1.7 million knots, which represents a large improvement over previous formulas of this kind. To find the approximation formula, we use layer-wise relevance propagation to reverse engineer a black box neural network which achieves a similar average error for the same approximation task when trained on 10% of the total dataset. The particular roots of unity which appear in our analysis cannot be written as $e^{2\pi i/(k+2)}$ with integer k; therefore, the relevant Jones polynomial evaluations are not given by unknot-normalized expectation values of Wilson loop operators in conventional $SU(2)$ Chern–Simons theory with level k. Instead, they correspond to an analytic continuation of such expectation values to fractional level. We briefly review the continuation procedure and comment on the presence of certain Lefschetz thimbles, to which our approximation formula is sensitive, in the analytically continued Chern–Simons integration cycle.

2.1 Introduction

The intersection of mathematical phenomenology and physical mathematics has historically led to profound insights into mathematical structures and the natural world. A modern example is the development of mirror symmetry, first as an observation about the non-uniqueness of Calabi–Yau compactifications of superstring theory [60; 61], then as a statement about mirror pairs of Calabi–Yau manifolds with Hodge numbers interchanged [62] and as a relationship between Gepner models [63]. These dualities among physical theories and the success of mirror symmetry as a tool to solve problems in enumerative geometry [64] supplied an invitation and a

motivation for later rigorous mathematical theorems [65; 66].

Knot theory presents another setting in which physics (quantum field theory) provides a novel window into mathematics. A knot is an embedding of $S^1 \subset S^3$. Because the same knot can be drawn in multiple ways, topological invariants provide labels for identifying a given knot K. Two of these topological invariants are the Jones polynomial, $J_2(K;q)$ [67], and, for hyperbolic knots, the volume of the knot complement, $\mathrm{Vol}(S^3 \setminus K)$ [68]. The Jones polynomial, as we review below, has a physical realization in terms of Wilson loop observables in Chern–Simons gauge theory [51], where it can be generalized to the colored Jones polynomial $J_R(K;q)$, with R a representation of the gauge group. These two topological invariants are related through the volume conjecture [69; 70; 71]:

$$\lim_{n\to\infty} \frac{2\pi \log |J_n(K;\omega_n)|}{n} = \mathrm{Vol}(S^3 \setminus K) \ , \tag{2.1.1}$$

where K is a hyperbolic knot, the color n denotes the n-dimensional irreducible representation of the gauge group $SU(2)$ in which the trace of the Wilson loop is taken, and $\omega_n = e^{2\pi i/n}$. The ordinary Jones polynomial corresponds to taking the fundamental representation $n = 2$.

While the large n limit of the colored Jones polynomial takes center stage in the volume conjecture, it turns out that the ordinary Jones polynomial also conveys some information about the volume. Dunfield initially observed the trend that, for alternating knots up to 13 crossings, a simple linear function of $\log |J_2(K;-1)|$ was approximately proportional to $\mathrm{Vol}(S^3 \setminus K)$ [72]. As noted already in that work, this relationship is mysterious because the coefficients of the linear function are not what one would expect by simply writing down the expression on the left hand side of (2.1.1) with $n = 2$. So, the linear relationship cannot be explained by suggesting that the volume conjecture converges quickly. Indeed, it is known that the left hand side of (2.1.1) is not even monotonic in n for certain knots, and [73] conjectured that convergence is only eventually monotonic.

Subsequently, a neural network predicted the hyperbolic volume from the Jones polynomial with $97.55 \pm 0.10\%$ accuracy for all hyperbolic knots up to 15 crossings using only 10% of the dataset of $313,209$ knots for training [74]. The input to the network was a vector consisting of the maximum and minimum degrees of the Jones polynomial along with its integer coefficients. The network's accuracy is essentially optimal for the following reason: knots with different hyperbolic volumes can have the same Jones polynomial, and when this happens, the volumes typically differ by about 3%. There are a total of 174619 unique Jones polynomials in the dataset. This means that the approximately 3% error of our neural network can be nearly entirely attributed to this phenomenon. This represented a large improvement over

the approximation formula using $J_2(K;-1)$, but introduced many more free parameters (the weights and biases of the neural network) and completely obscured the true functional dependence on $J_2(K;q)$.

This work in experimental mathematics established that there is more information about the volume contained in the full Jones polynomial than in the single evaluation $J_2(K;-1)$.[1] Indeed, training on only 319 knots (0.1% of the dataset) is sufficient to extract this additional information. Moreover, the predictions apply to all hyperbolic knots, not just the alternating ones. The neural network, however, is a black box that learns semantics without knowing any syntax. While machine learning successfully identifies relationships between topological invariants of hyperbolic knots, we do not have an analytic understanding of why this occurs or how the machine learns the relationship in the first place. The aim of this chapter is to determine which aspects of the input are most important for the neural network's considerations and to use this to deduce a simpler functional relationship between the Jones polynomial and the hyperbolic volume that depends on only an $O(1)$ number of free parameters. In other words, we seek an approximation formula which similarly outperforms the formula based on $J_2(K;-1)$, but which does not rely on the complicated structure of a neural network.

The main result obtained in this work is an approximation formula for the hyperbolic volume of a knot in terms of a single evaluation of its Jones polynomial.

$$V_{3/4}(K) = 6.20 \log\left(|J_2(K;e^{3\pi i/4})| + 6.77\right) - 0.94\ . \tag{2.1.2}$$

This formula is numerical in nature, and we have no sharp error bounds, but it achieves an accuracy of more than 97% on the first 1.7 million hyperbolic knots. The phase $e^{3\pi i/4}$ can be adjusted to some degree to obtain similar formulas which perform almost as well; we explore these alternatives. We obtain (2.1.2) and its analogues by reverse engineering the behavior of the aforementioned neural network, a task which is generally considered difficult in the machine learning community. In our case, it is possible due to the simplicity of the underlying neural network architecture and the power of layer-wise relevance propagation, a network analysis technique which we review.

Since the phase $e^{3\pi i/4}$ cannot be written as $e^{2\pi i/(k+2)}$ for integer k, (2.1.2) suggests that we should consider analytically continued Chern–Simons theory, which was previously explored as a possible route to understand the volume conjecture [76].

[1] We refer loosely to the information contained in certain invariants, as all of our results are numerical and therefore "probably approximately correct" [75].

Specifically, the evaluations of the Jones polynomial that are relevant for us correspond to fractional Chern–Simons levels $k = \frac{2}{3}$ and $k = \frac{1}{2}$, which must be understood in the context of the analytically continued theory due to the integer level restriction normally enforced by gauge invariance. We provide a review of the main points of the analytically continued theory, as our results have some speculative bearing on which Lefschetz thimbles should contribute to the path integral at various values of n and k. Our interpretations pass several sanity checks when compared with the main lessons of [76], and we use our results to formulate an alternative version of the volume conjecture.

Though we do not have a complete explanation for why (2.1.2) predicts the volume so well, its implications are intriguing. It points to the existence of a sort of quantum/semiclassical duality between $SU(2)$ and $SL(2,\mathbb{C})$ Chern–Simons theory, since some simple numerical coefficients are enough to transform a strong coupling object (the Jones polynomial at small k) into a weak coupling one (the hyperbolic volume). This is reminiscent of the shift $k \to k+2$ induced by the one-loop correction in the $SU(2)$ Chern–Simons path integral [51], which transforms a semiclassical approximation into a more quantum result via a single $O(1)$ parameter. We can explore this phenomenon in a preliminary way by checking whether J_2 contains any information about J_n for $n > 2$ and some $O(1)$ phases like the ones appearing in the volume conjecture. Computing $11,921$ colored Jones polynomials in the adjoint representation of $SU(2)$, we notice from Figure 2.1 that

$$|J_3(K;\omega_3)| \sim J_2(K;\omega_2)^2 \ . \tag{2.1.3}$$

Comparing to the volume conjecture (2.1.1), this suggests that the ordinary Jones polynomial has some knowledge of the behavior of higher colors and in particular the $n \to \infty$ limit.[2]

The organization of this chapter is as follows. In Section 2.2, we review relevant aspects of Chern–Simons theory, including its relationship to knot invariants and its analytic continuation. In Section 2.3, we describe the machine learning methods we employ, particularly neural networks and layer-wise relevance propagation. In Section 2.4, we analyze the structure of the neural network and, using layer-wise relevance propagation, deduce a relationship between evaluations of the Jones polynomial of a hyperbolic knot and the volume of its knot complement. We also comment on the implications of our numerics for the presence of certain integration cycles in the analytically continued Chern–Simons path integral as a function of coupling. In Section 3.5, we discuss the results we obtain and propose some future

[2]This analysis was performed in collaboration with Onkar Parrikar.

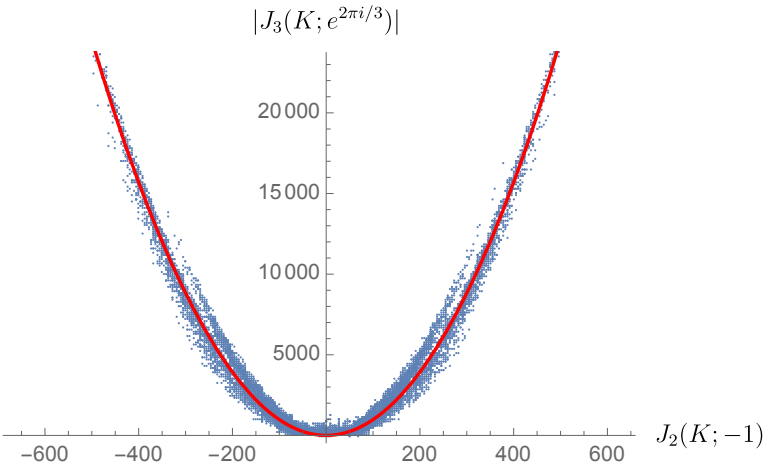


Figure 2.1: The $n = 2$ and $n = 3$ colored Jones polynomials evaluated at roots of unity show a quadratic relation.

directions. We also provide several appendices which outline details about the implementation of our machine learning algorithms (Appendix 2.A), results relating the scaling of the Jones polynomial coefficients with the hyperbolic volume (Appendix 3.2), details concerning the various normalizations of the Jones polynomial in the mathematics and physics literature (Appendix 2.C), a data analysis of knot invariants using t-distributed stochastic neighbor embedding (Appendix 2.D), and finally, an overview of related experiments (Appendix 2.E).

2.2 Chern–Simons theory

We review several aspects of Chern–Simons theory, its relation to the colored Jones polynomials and hyperbolic volumes of knots, and its analytic continuation away from integer level.

2.2.1 Knot invariants

Chern–Simons theory is a three-dimensional topological quantum field theory which provides a unifying language for the knot invariants with which we will be concerned [51; 71]. The Chern–Simons function is defined, using a connection (or gauge field) A on an $SU(2)$-bundle E over a three manifold M, as

$$W(A) = \frac{1}{4\pi} \int_M \mathrm{Tr} \left[A \wedge dA + \frac{2}{3} A \wedge A \wedge A \right] \ . \tag{2.2.1}$$

The trace is taken in the fundamental representation. The path integral of Chern–Simons gauge theory is then given in the compact case by

$$Z(M) = \int_{\mathcal{U}} [\mathcal{D}A] \ \exp(ikW(A)) \ , \tag{2.2.2}$$

where $\mathcal{U}$ is the space of $SU(2)$ gauge fields modulo gauge transformations. The coupling k is integer-quantized, $k \in \mathbb{Z}$, to ensure gauge invariance. For $SU(2)$ Chern–Simons theory on $M = S^3$, it was shown in [51] that the expectation value of a Wilson loop operator along the knot, defined by

$$U_R(K) = \mathrm{Tr}_R \, \mathcal{P} \exp\left(- \oint_K A \right) \ , \tag{2.2.3}$$

is related to the colored Jones polynomials $J_R(K;q)$ of a knot $K \subset S^3$ evaluated at $q = e^{2\pi i/(k+2)}$. In our work, we will be interested in evaluations of the Jones polynomial (where the representation R is the fundamental one) away from this particular root of unity, and indeed away from all roots expressible as $e^{2\pi i/(k+2)}$ for some $k \in \mathbb{Z}$. Strictly speaking, these evaluations of the Jones polynomial are not provided by the usual formulation using the path integral of Chern–Simons theory. However, evaluation at arbitrary phases $e^{2\pi i/(k+2)}$ for $k \in \mathbb{R}$ can be achieved by making use of the analytic continuation machinery developed in [76].

We will also be interested in a complex-valued Chern–Simons function $W(\mathcal{A})$ obtained from an $SL(2,\mathbb{C})$-bundle $E_{\mathbb{C}}$ over M with connection $\mathcal{A}$. In the non-compact $SL(2,\mathbb{C})$ case, there are two couplings, and the path integral is

$$\mathcal{Z}(M) = \int_{\mathcal{U}_{\mathbb{C}}} [\mathcal{D}\mathcal{A}][\mathcal{D}\overline{\mathcal{A}}] \ \exp\left[\frac{it}{2} W(\mathcal{A}) + \frac{i\widetilde{t}}{2} W(\overline{\mathcal{A}}) \right] \ , \tag{2.2.4}$$

where $\mathcal{U}_{\mathbb{C}}$ is the space of $SL(2,\mathbb{C})$ gauge fields modulo gauge transformations, and the complex couplings $t = \ell + is$ and $\widetilde{t} = \ell - is$ obey $s \in \mathbb{C}$ and $\ell \in \mathbb{Z}$. The coupling ℓ, which multiplies the real part $\mathrm{Re}(W)$, is integer-quantized for the same reason as k was in the compact case. On the other hand, $\mathrm{Im}(W)$ is a well-defined complex number even under arbitrary gauge transformations, so s can (in principle) take any complex value.[3] There is a particularly interesting saddle point (flat connection) which contributes to $\mathcal{Z}(M)$ in the case where M admits a complete hyperbolic metric. Such manifolds are relevant for us because, as explained by Thurston [41], most knots K admit a complete hyperbolic metric on their complements $S^3 \setminus K$. The knot complement $S^3 \setminus K$ is the three manifold obtained by drilling out a tubular neighborhood around the knot $K \subset S^3$. This complement is topologically distinct from S^3, and the knot shrinks to a torus cusp in the complete hyperbolic metric.[4] For such hyperbolic M, there exists a "geometric" flat $SL(2,\mathbb{C})$ connection $\mathcal{A}_-$[5]

[3] Usually one takes $s \in \mathbb{R}$ to ensure that the argument of the path integral is bounded, but via analytic continuation one can obtain something sensible for more general s [76].

[4] We emphasize that this complete hyperbolic metric on $S^3 \setminus K$ does not descend somehow from a metric on S^3. It exists due to an ideal tetrahedral decomposition of $S^3 \setminus K$, where each tetrahedron has an embedding in $\mathbb{H}^3$, and the embeddings can be glued together in a consistent way.

[5] We choose this notation to match (5.61) of [76], and comment further in footnote 20.

for which $\mathrm{Im}(W(\mathcal{A}_-))$ is related to the volume of the hyperbolic metric on M via $\mathrm{Im}(W(\mathcal{A}_-)) = \mathrm{Vol}(M)/2\pi$.[6] Thus, $SL(2,\mathbb{C})$ Chern–Simons theory is intimately related to the hyperbolic volumes of three manifolds, as this quantity makes a saddle point contribution to the path integral.

One of the primary motivations for both this work and the work of [76] was the so-called volume conjecture [69], which relates the colored Jones invariants and the hyperbolic volume of three manifolds (which, by the Mostow–Prasad rigidity theorem, is a topological invariant of any hyperbolic M). As written in the introduction and reproduced here, the volume conjecture states [69; 70; 71]

$$\lim_{n\to\infty} \frac{2\pi \log |J_n(K; e^{2\pi i/n})|}{n} = \mathrm{Vol}(S^3 \setminus K) \ , \tag{2.2.5}$$

where n is the n-dimensional irreducible representation of $SU(2)$. Thus, the natural unifying language for the volume conjecture is $SU(2)$ and $SL(2,\mathbb{C})$ Chern–Simons theory, because the knot invariants appearing on both the left and right hand sides are quantities which appear in the calculation of the Chern–Simons path integral.[7]

In writing the above relationship, we must be careful to specify precisely what we mean by the colored Jones invariants $J_n(K;q)$. We choose a normalization of the colored Jones invariants so that they are Laurent polynomials in the variable q, and that the unknot 0_1 obeys $J_n(0_1;q) = 1$. As explained clearly in Section 2.5.3 of [76], this choice implies that the colored Jones polynomials are reproduced by the ratio of two Chern–Simons path integrals. The numerator of this ratio has a single Wilson loop insertion along the knot $K \subset S^3$, and the denominator has a Wilson loop along an unknot $0_1 \subset S^3$. Explicitly, we have

$$J_n(K; q = e^{2\pi i/(k+2)}) = \frac{\int_{\mathcal{U}} [\mathcal{D}A] \ U_n(K) \exp(ikW(A))}{\int_{\mathcal{U}} [\mathcal{D}A] \ U_n(0_1) \exp(ikW(A))} \ , \tag{2.2.6}$$

and the manifold implicit in $W(A)$ is $M = S^3$. It is for this reason that the statement of the volume conjecture only involves evaluation of the n-colored Jones polynomial at integer $k = n-2$, whereas to understand the conjecture in Chern–Simons theory it is necessary to analytically continue the path integral away from integer values of k. Without normalization, the path integral yields an expression which vanishes at $q = e^{2\pi i/n}$, and this vanishing is removed by dividing by the unknot expectation. In short, we can either measure how fast the path integral vanishes by computing a derivative with respect to k (which relies on the analytic structure of the function),

[6] The real part $\mathrm{Re}(W(\mathcal{A}_-))$ of the geometric connection is proportional to a number which is often called the "Chern–Simons invariant" of M.

[7] This combination of invariants, the colored Jones polynomial and hyperbolic volume, was explored in the context of entanglement entropy of Euclidean path integral states in [77; 78].

or we can explicitly divide by a function which vanishes equally as fast to obtain a finite ratio.[8]

2.2.2 Analytic continuation

With these conventions in hand, we will provide a brief review of the main results in [76], essentially to introduce the language, as we will make some speculative comments on the relationship between our numerical results and the analytic techniques in [76]. To analytically continue $SU(2)$ Chern–Simons theory on $S^3 \setminus K$ away from integer values of k, [76] instructs us to first rewrite the $SU(2)$ path integral over A as an $SL(2,\mathbb{C})$ path integral over $\mathcal{A}$ restricted to a real integration cycle $\mathcal{C}_{\mathbb{R}}$ in the space of complex-valued connections modulo gauge transformations $\mathcal{U}_{\mathbb{C}}$. Analytic continuation occurs then by lifting $\mathcal{C}_{\mathbb{R}}$ to a cycle $\mathcal{C}$ in the universal cover $\widehat{\mathcal{U}}_{\mathbb{C}}$,[9] where the Chern–Simons function $W(\mathcal{A})$ is well-defined as a complex-valued function. (Recall that large gauge transformations can modify the real part of $W(\mathcal{A})$ by a multiple of 2π.) In the presence of a Wilson loop along a knot, the relevant path integral looks like

$$\int_{\mathcal{C}\subset\widehat{\mathcal{U}}_{\mathbb{C}}} [\mathcal{D}\mathcal{A}] \int [\mathcal{D}\rho] \ \exp(\mathcal{I}) \ , \qquad \mathcal{I} = ikW(\mathcal{A}) + iI_n(\rho,\mathcal{A}) \ . \tag{2.2.7}$$

We have added an additional term that depends on a field ρ, which is associated with the knot itself. The introduction of this field along with its action I_n is a way to absorb the Wilson loop $U_n(K)$ into the exponential, and makes use of the Borel–Weil–Bott theorem. We will not provide a discussion of this point, and simply refer interested readers to [76]; we will only refer to the total exponential argument $\mathcal{I}$ from now on. We will just make one important remark concerning I_n: when evaluated on a flat connection, I_n is proportional to $n-1$ [76]. Therefore, up to an overall rescaling, $\mathcal{I}$ depends only on the combination

$$\gamma \equiv \frac{n-1}{k} \ , \tag{2.2.8}$$

when evaluated on a flat connection.[10] If we wish to understand the volume conjecture, $\gamma = 1$ is held fixed in the semiclassical limit $n, k \to \infty$. When quoting values for γ, we have in mind the ratio (2.2.6); in the bare path integral setup of [76], we

[8] This subtle point, while discussed in this language in [79], is implicit throughout [76] (also see footnote 7 in [79], and note that our $J_n(K;q)$ would be written as $J_n(K;q)/J_n(0_1;q)$ in the notation of [79]).

[9] This space can also be thought of as the space of complex-valued gauge fields modulo topologically trivial, or "small," gauge transformations.

[10] This differs from the definition of γ in [76], where n/k was used instead. In that discussion, the semiclassical limit was of much greater importance, and in that limit our definition becomes equal to n/k. However, as we are working at n and k of $O(1)$, we will keep the exact ratio implied by the value of $\mathcal{I}$ on a flat connection.

would need to move slightly away from $\gamma = 1$.[11]

The cycle $\mathcal{C}$ must be extended using the machinery of Morse theory; this extension guarantees that the Ward identities will hold. Morse theory on $\widehat{\mathcal{U}}_{\mathbb{C}}$, specifically with $\mathrm{Re}(\mathcal{I})$ as a Morse function, plays a key rôle in this extension and in the definition of other integration cycles. Analytic continuation away from a discrete set of points (integer values of k, in this case) is not unique, and this corresponds to an ambiguity in lifting $\mathcal{C}_{\mathbb{R}}$ to $\mathcal{C}$. The relatively natural resolution in this situation is to ask that the path integral should have no exponentially growing contributions as $k \to \infty$ with fixed n.[12] This is equivalent to requiring that the colored Jones polynomials, as defined in the mathematical literature, and the ratio of Chern–Simons path integrals (2.2.6) should hold for more general q after replacing the path integrals with their analytic continuations.[13]

Once the cycle $\mathcal{C}$ has been defined, we must vary the value of γ from zero to our desired point. We begin at zero since we have defined a sort of boundary condition on $\mathcal{C}$ at $k \to \infty$ with fixed n by selecting the relatively natural analytic continuation just described. As we vary γ, we must track the behavior of $\mathcal{C}$. It may seem like there is nothing to keep track of, but in fact there are subtle Stokes phenomena which must be taken into account, as we will now briefly explain. The cycle $\mathcal{C}$ has a decomposition in terms of so-called Lefschetz thimbles $\mathcal{J}_\sigma$, which are cycles defined using Morse theory that each pass through precisely one critical point σ of $\mathcal{I}$, and are defined so that the path integral along them always converges:

$$\mathcal{C} = \sum_\sigma \mathfrak{n}_\sigma \mathcal{J}_\sigma \ . \tag{2.2.9}$$

These thimbles can intuitively be visualized as downward flow lines coming from a critical point. Since our Morse function is $\mathrm{Re}(\mathcal{I})$, the path integral decreases exponentially when following a downward flow line: it is for this reason that convergence is guaranteed.[14]

[11] Actually, even with the ratio of path integrals, we need to move away from exactly $\gamma = 1$. We will continue to write $\gamma = 1$ as the relevant point for the volume conjecture in the semiclassical limit, but the true value for integer n and k is more like $\gamma = \frac{n-1}{n-2} > 1$.

[12] As mentioned in [76], it is not quite clear how to enforce this condition on $\mathcal{C}$ for general knots, but we will not need the details.

[13] Conventions are not uniform across the mathematical literature, and the relevance of framing is often unmentioned. See Appendix 2.C for further discussion of the alignment between the path integral ratio (2.2.6) and the mathematical definitions of J_2.

[14] The downward flow equations of Morse theory on the space of complex-valued gauge fields, which define the Lefschetz thimbles in the infinite-dimensional path integral setting, behave similarly to the finite-dimensional case due to their elliptic nature.

When crossing certain loci in the complex γ plane, known as Stokes lines, the decomposition of $\mathcal{C}$ in terms of Lefschetz thimbles may be required to change in order to both preserve the cycle $\mathcal{C}$ locally and ensure convergence of the path integral. Functionally, the coefficients $\mathfrak{n}_\sigma$ change in exactly the right way to compensate for a similar jumping in the Lefschetz thimbles $\mathcal{J}_\sigma$ themselves. This jumping occurs for a cycle $\mathcal{J}_\sigma$ when there is a downward flow from σ that ends at another critical point rather than flowing to $-\infty$. Thus, recalling that critical points σ of $\mathcal{I}$ on $\widehat{\mathcal{U}}_{\mathbb{C}}$ are flat $SL(2,\mathbb{C})$ connections on $S^3 \setminus K$ with a prescribed monodromy around K due to the Wilson loop, Stokes phenomena can lead to the addition of complex $SL(2,\mathbb{C})$ connections to the analytically continued $SU(2)$ path integral, even though we begin with an integration contour that includes only real $SU(2)$-valued connections.

As γ is varied, two flat connections can become coincident, which leads to a singularity in the moduli space of flat connections. Studying such singularities is necessary to understand the Stokes phenomena involved, as there is a trivial solution of the Morse theory flow equations when two flat connections are coincident. Indeed, the existence of finite-dimensional local models of such singularities allows one to understand some Stokes phenomena in Chern–Simons theory without dealing with the full geometry of $\widehat{\mathcal{U}}_{\mathbb{C}}$. The point we emphasize here is that, at least for these Stokes phenomena in particular, we may analyze Stokes curves purely as a function of γ rather than some more complicated parameter, since the flow is trivial and the only relevant evaluations of $\mathcal{I}$ are on flat connections.

As an explanation for the volume conjecture, we should find that Stokes phenomena in Chern–Simons theory can lead to exponential growth of the path integral. The final crucial detail which leads to exponentially growing contributions is as follows. When passing from the space of $SL(2,\mathbb{C})$ gauge fields modulo gauge transformations to its universal cover, each flat connection in $\mathcal{U}_{\mathbb{C}}$ is lifted to an infinite family of flat connections in $\widehat{\mathcal{U}}_{\mathbb{C}}$ which differ only in the real parts of their Chern–Simons functions W by a multiple of 2π.[15] As γ is varied, degenerate pairs of these lifted critical points can separate (say at γ_1), and subsequently recombine (at γ_2) into new degenerate pairs of two critical points which were not paired initially.

If the Lefschetz thimbles associated to such a pair of critical points are added to the integration cycle due to Stokes phenomena at or before γ_1, their contributions will have changed drastically by the time γ_2 is reached. Namely, the thimbles are now associated with one critical point each from two newly recombined pairs, and

[15] There is actually a similar issue which arises for the field ρ appearing in the action I_n, if we wish to analytically continue in n as well as k. If we do not continue in n, this issue modifies the way that downward flow conserves $\mathrm{Im}(\mathcal{I})$. But again, we refer interested readers to [76].

this gives in the path integral a difference of phases which vanishes only for $k \in \mathbb{Z}$. The prefactor of this phase difference can be exponentially large in k, and so the total contribution may diverge exponentially for non-integer k. Indeed, schematically the path integral will have a semiclassical term of the form

$$e^{ikW(\mathcal{A})}(1 - e^{2\pi ik}), \tag{2.2.10}$$

so the pair of critical points will exactly cancel for $k \in \mathbb{Z}$, and diverge exponentially in k for $\mathrm{Im}(W(\mathcal{A})) < 0$ and non-integer real $k > 0$. Therefore, it is the combination of the lifting of critical points to $\widehat{\mathcal{U}}_{\mathbb{C}}$, their splitting and recombination as a function of γ, and Stokes phenomena which can lead to the situation predicted by the volume conjecture: an $SL(2, \mathbb{C})$ flat connection contributes an exponentially growing term to the asymptotic behavior ($\gamma \to 1$, $k \to \infty$) of an $SU(2)$ Chern–Simons path integral.

We return to these ideas in Section 2.4.2, where they become relevant in light of our approximation formula (2.1.2) and generalizations thereof.

2.3 Machine learning

In this work, we build upon the findings of [74], often by employing deep learning [80] (and other machine learning techniques) to decipher the relationships between knot invariants. Neural networks have been recently employed in knot theory to calculate invariants like the slice genus [81] and to solve computationally complex problems like unknot recognition [82]. Indeed, it is known that a neural network of suitable size can approximate any function [83; 84]. Our dataset (which matches that of [74]) consists of the Jones polynomial, hyperbolic volume, and other knot invariants for all $1,701,913$ hyperbolic knots with 16 or fewer crossings [85], tabulated using a combination of the Knot Atlas database [86] and the `SnapPy` program [87].

In [74], a neural network was used to demonstrate that there is a relationship between the Jones polynomial and the hyperbolic volume of a knot. This was initially achieved with a fully connected two hidden layer network with 100 neurons in each layer. Experiments with the Jones polynomial evaluations (see Section 2.4) were initially performed with a network with two hidden layers 50 neurons wide. Later experiments (see Appendix 2.E for details) found that the network could predict the volumes with roughly 96% accuracy with a two hidden layer network only 5 neurons wide. 2.A.1 The robust performance of these small neural networks is compelling evidence that a simple approximate function exists which maps the Jones polynomial to the hyperbolic volume. Ideally, we would like to go beyond demonstrating the existence of a relationship and actually write down a function that faithfully models

the relationship. Unfortunately, the neural network is not much help here. Though it essentially optimizes a function that fits the data, the function is computed via the multiplication and addition of matrices; even in the case of a small network, the multiplication of 5×5 matrices still produces functions which are difficult to interpret.

Before describing our strategies for extracting simpler expressions for the relevant correlations learned by the neural networks, we review the deep learning ideas involved.

2.3.1 Neural networks

A neural network is a function f_θ which approximates the relationship between a vector of input features $\mathbf{v}_{\text{in}}$ and some output $\mathbf{v}_{\text{out}}$. The network f_θ is an approximation of the true relationship $A : \mathbf{v}_{\text{in}} \to \mathbf{v}_{\text{out}}$. In our case, the input vectors are (vectors based on) the Jones polynomials and the outputs are the corresponding hyperbolic volumes. The dataset is divided into training and testing sets. The network uses the training data to adjust the parameters θ (the weights and biases) to approximate the relationship A as closely as possible. To do this, a loss function is chosen and minimized in the space of parameters.In the architecture used in this work (see Figure 2.2), the network is built out of n hidden layers which perform matrix multiplication by a weight matrix W_θ^m, addition of a bias vector $\mathbf{b}_\theta^m$, and element-wise application of the activation function σ. The network can be written as

$$f_\theta(\mathbf{v}_{\text{in}}) = L_\theta^n \left(\sigma \left(\dots L_\theta^2 \left(\sigma \left(L_\theta^1(\mathbf{v}_{\text{in}})\right)\right) \dots\right)\right) \ , \qquad L_\theta^m(\mathbf{v}) = W_\theta^m \cdot \mathbf{v} + \mathbf{b}_\theta^m \ . \quad (2.3.1)$$

The values of hidden layer neurons after applying the activation function are often called the activations $\mathbf{a}^m = \sigma(W_\theta^m \cdot \mathbf{a}^{m-1} + \mathbf{b}_\theta^m)$, with $\mathbf{a}^0 \equiv \mathbf{v}_{\text{in}}$. In this work, we use the Rectified Linear Unit (ReLU) activation function, which is $\sigma(x) = x\Theta(x)$, where $\Theta(x)$ is the Heaviside step function. The loss function is minimized on the training data by using the backpropagation algorithm. This algorithm computes gradients of the loss function for each training data point and adjusts the parameters layer by layer in the network. Once this training is complete, f_θ is applied to the previously unseen testing set to see how well it approximates A. A specific discussion of the neural network architecture used in this work is included in Appendix 2.A.

2.3.2 Layer-wise relevance propagation

Layer-wise relevance propagation (LRP) is a technique which attempts to explain neural network predictions by calculating a relevance score for each input feature [88]. This is a useful tool when attempting to derive an analytic function:

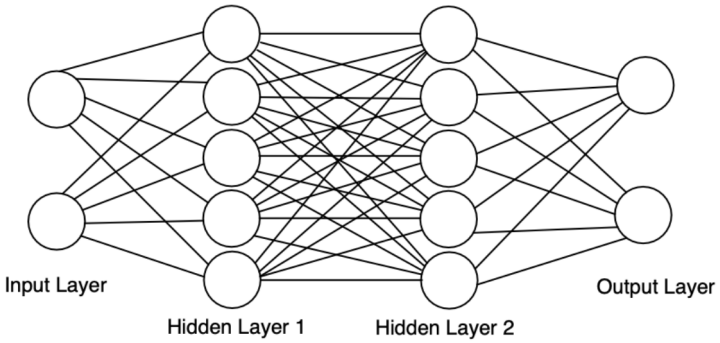


Figure 2.2: An example of a two hidden layer fully connected neural network architecture. Each hidden layer represents matrix multiplication by a weight vector, the addition of a bias vector, and the element-wise application of an activation function, which introduces non-linearity into the function. In this work we used the Rectified Linear Unit (ReLU) activation function.

we can determine which input variables typically carry the most importance when predicting the output. This allows us to hypothesize how to weight input variables in our function and perhaps reduce the complexity of the problem by eliminating redundant variables.

LRP propagates backwards through a neural network, starting at the output layer. The LRP algorithm redistributes the relevance scores from the current layer into the previous layer, employing a conservation property. Denote the activation of a neuron i in layer m by a_i^m. Suppose we have all the relevance scores for the current layer, and want to determine the scores R_j^{m-1} in the previous layer. The most basic LRP rule calculates these relevance scores using the formula [89]

$$R_j^{m-1} = \sum_k \frac{a_j^{m-1} W_{jk}^m + N_{m-1}^{-1} b_k^m}{\sum_l a_l^{m-1} W_{lk}^m + b_k^m} R_k^m \ , \tag{2.3.2}$$

where the m^{th} layer has N_m neurons. The subscripts on the weights W and biases b here denote matrix and vector indices. The numerator is the activation of the j^{th} neuron in layer $m-1$, multiplied by the weight matrix element connecting that neuron to neuron k in layer m to model how much neuron j contributes to the relevance of neuron k. This fraction is then multiplied by the relevance of neuron k in the layer m. Once the input layer (layer zero) is reached, the propagation is terminated, and the result is a list of relevance scores for the input variables. The sum in the denominator runs over all of the neurons in layer $m-1$, plus the bias in layer m: it imposes the conservation property because we begin with $R^n = 1$ at the output layer and always preserve

$$\sum_k R_k^m = 1 \ . \tag{2.3.3}$$

This methodology was originally proposed in a classifier problem; we have adapted it to the case of regression.

2.3.3 Strategies for analyzing the network

Armed with the evidence that our function exists, how do we proceed? The input is too complicated for educated guesswork or traditional curve fitting techniques, as our encoding of the Jones polynomial is a 16-vector.[16] Our approach involves performing experiments through which we can probe how the neural network makes its predictions. As stated previously, a neural network's inner workings are largely inaccessible. Despite this, by studying a network's success when faced with a transformed or truncated form of the Jones polynomial, we can begin to update our idea of what the eventual function might look like. There are three main ways that we accomplish this.

The first type of experiment is training a neural network on some truncated or mutated form of the input data. For instance, inspired by the volume-ish theorem [90], we could create a new input vector containing just the degrees of the polynomial and the first and last coefficients: $(p_{\min}, p_{\max}, c_1, c_{-1})$. If the neural network was then still able to learn the volume, with comparable accuracy to the original experiment, then perhaps our function could be built from these four numbers. It turns out that the neural network did not perform well with this input. Another example of this method is detailed in Section 2.4.

The second strategy is taking a neural network which has already been trained and feeding it altered input data. For instance, if we train a network on our original 16-vectors and give it an input vector where certain elements have been zeroed out or shifted by some constant, can it still predict the volume? This allows us to probe our function for, among other things, redundant variables and invariance under translations. We could, for example, shift pairs of input variables by a constant and record the network's ability to predict the volume. These experiments are inspired by [91], where symbolic regression and machine learning are approached with a physicist's methodology.[17]

Together with these two experiments, we use LRP to understand the relevance of the various input features in making a prediction. As we reviewed above, LRP

[16] In [74], we provided the degrees of the Jones polynomial as inputs. As we explicate in Appendix 2.E.2, the neural network performs just as well with the coefficients alone. This is the 16-vector to which we refer.

[17] We explore symbolic regression techniques further in Appendix 2.E.4. We obtain 96.56% accuracy, but the formulæ are not obviously interpretable.

is an algorithm which uses the neuron activations, along with the weight and bias matrices, to map the activations in the final network layer back onto the input layer. This technique is successfully implemented in Section 2.4 to further reduce the number of input variables needed to predict the volume, eventually yielding formulas with just a handful of numerical parameters and a single nonlinearity.

2.4 An approximation formula for the hyperbolic volume

Our goal is to employ the machine learning techniques reviewed in Section 2.3 to determine what particular property of the Jones polynomial was exploited by the neural network in [74] to compute the hyperbolic volume. Inspired by the observations of [76] concerning the analytic continuation of Chern–Simons theory that we briefly summarized in Section 2.2, we approach this task by evaluating the Jones polynomial at various roots of unity. Indeed, the Jones polynomial is determined by its values at roots of unity, so we lose no essential information if we include enough such evaluations. We use techniques from interpretable deep learning to reverse engineer neural networks in Section 2.4.1 and comment on the implications for the path integral of analytically continued Chern–Simons theory in Section 2.4.2.

2.4.1 Interpretable deep learning

We begin by training a neural network composed of two fully connected hidden layers with 50 neurons each on the Jones polynomial evaluated at $e^{2\pi ip/(r+2)}$, for integers $r \in [3, 20]$ and $p \in [0, r+2]$. Complex conjugates are omitted from this tabulation since Laurent polynomials obey $J_2(K; \overline{q}) = \overline{J_2(K; q)}$. The input data (which includes all hyperbolic knots up to and including 15 crossings) is represented as a vector where the entries at position $2p$ and $2p+1$ correspond to the real and complex parts of the p^{th} evaluation. Layer-wise relevance propagation (LRP) [88; 89] is used to identify which evaluations of the Jones polynomial are important in calculating the volume.[18]

The LRP results are easily interpreted graphically, as demonstrated in Figure 2.3. We can immediately see which columns are relevant in making a prediction of the volume. The final column in Table 2.1 shows the results when training the neural network on just those evaluations which LRP determines are relevant. Comparing the values in the final column to the original error in the second column, we see that there is no real reduction in performance. LRP successfully selects the relevant

[18] Layer-wise relevance propagation has not been widely applied in the physics context, but see [92] for an example.

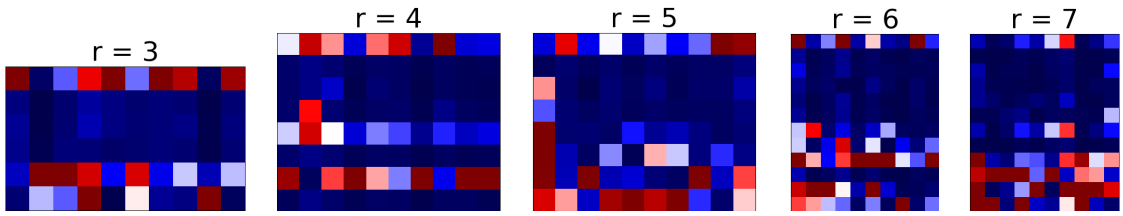


Figure 2.3: Each grid shows the LRP results for 10 neural network inputs. The grid labels $r \in [3, 7]$ determine the form of the input vector. Each column represents a single input, corresponding to the real and imaginary parts of the knot's Jones polynomial evaluated at the phases $e^{2\pi ip/(r+2)}$ with $p \in \mathbb{Z}$ and $0 \leq 2p \leq r + 2$. Dark blue squares correspond to the smallest magnitude for the relevance score, and red to the highest. So, in a given column, the red squares represent evaluations which were highly relevant for the neural network to predict that knot's volume. We also see that the same evaluations are often relevant across all 10 knots in the displayed set; these correspond to mostly red rows.

input features, and in some cases performance is actually improved by pruning irrelevant inputs.

The results in Table 2.1 were each obtained over one run, so fluctuations in the error are expected. To test the stability of the results, we trained the neural network 20 times, using $J_2(K; e^{4\pi i/5})$ (the relevant root for $r = 3$) as input. Averaged over 20 runs, the error was $3.71 \pm 0.06\%$. With this relatively small standard deviation, we conclude that network performance is stable over multiple runs.

In the first line of Table 2.1, the neural network is learning to predict the volume from two numbers: the real and imaginary parts of $J_2(K; e^{4\pi i/5})$. If we represent the input as a magnitude and a phase rather than as a real and imaginary part, the performance of the neural network is unchanged. In fact, it turns out that if we drop the phase and work only with the magnitude, the performance remains unchanged; the phase does not matter. This means that the network is predicting the volume (to $96.29 \pm 0.06\%$ accuracy) with just one number: $|J_2(K; e^{4\pi i/5})|$. Another promising candidate is $|J_2(K; e^{3\pi i/4})|$, which was also determined by relevance propagation to be important in several rows of Table 2.1. Plotting either $|J_2(K; e^{4\pi i/5})|$ or $|J_2(K; e^{3\pi i/4})|$ against the volume (Figure 2.4), we find a graph which looks like $a \log(x + b) + c$ for some constant $O(1)$ parameters a, $b > 0$, and c. Thus, we may be able to obtain a reasonably compact analytic expression by performing simple regression on such an ansatz.

We use the curve fitting and optimization procedures built into `Mathematica`. The dataset now includes all hyperbolic knots up to and including 16 crossings. For

r	Error	Relevant roots	Fractional levels	Error (relevant roots)
3	3.48%	$e^{4\pi i/5}$	$\frac{1}{2}$	3.8%
4	6.66%	-1	0	6.78%
5	3.48%	$e^{6\pi i/7}$	$\frac{1}{3}$	3.38%
6	2.94%	$e^{3\pi i/4}$, -1	$\frac{2}{3}$, 0	3%
7	5.37%	$e^{8\pi i/9}$	$\frac{1}{4}$	5.32%
8	2.50%	$e^{3\pi i/5}$, $e^{4\pi i/5}$, -1	$\frac{4}{3}$, $\frac{1}{2}$, 0	2.5%
9	2.74%	$e^{8\pi i/11}$, $e^{10\pi i/11}$	$\frac{3}{4}$, $\frac{1}{5}$	2.85%
10	3.51%	$e^{2\pi i/3}$,$e^{5\pi i/6}$, -1	1, $\frac{2}{5}$, 0	4.39%
11	2.51%	$e^{8\pi i/13}$, $e^{10\pi i/13}$, $e^{12\pi i/13}$	$\frac{5}{4}$, $\frac{3}{5}$, $\frac{1}{6}$	2.44%
12	2.39%	$e^{5\pi i/7}$, $e^{6\pi i/7}$, -1	$\frac{4}{5}$, $\frac{1}{3}$, 0	2.75%
13	2.52%	$e^{2\pi i/3}$, $e^{4\pi i/5}$, $e^{14\pi i/15}$	1, $\frac{1}{2}$, $\frac{1}{7}$	2.43%
14	2.58%	$e^{3\pi i/4}$, $e^{7\pi i/8}$, -1	$\frac{2}{3}$, $\frac{2}{7}$, 0	2.55%
15	2.38%	$e^{12\pi i/17}$, $e^{14\pi i/17}$, $e^{16\pi i/17}$	$\frac{5}{6}$, $\frac{3}{7}$, $\frac{1}{8}$	2.4%
16	2.57%	$e^{2\pi i/3}$, $e^{7\pi i/9}$, $e^{8\pi i/9}$, -1	1, $\frac{4}{7}$, $\frac{1}{4}$, 0	2.45%
17	2.65%	$e^{14\pi i/19}$, $e^{16\pi i/19}$, $e^{18\pi i/19}$,	$\frac{5}{7}$, $\frac{3}{8}$, $\frac{1}{9}$	2.46%
18	2.49%	$e^{4\pi i/5}$, $e^{9\pi i/10}$, -1	$\frac{1}{2}$, $\frac{2}{9}$, 0	2.52%
19	2.45%	$e^{2\pi i/3}$, $e^{16\pi i/21}$, $e^{6\pi i/7}$, $e^{20\pi i/21}$	1, $\frac{5}{8}$, $\frac{1}{3}$, $\frac{1}{10}$	2.43%
20	2.79%	$e^{8\pi i/11}$, $e^{9\pi i/11}$, $e^{10\pi i/11}$, -1	$\frac{3}{4}$, $\frac{4}{9}$, $\frac{1}{5}$, 0	2.4%

Table 2.1: LRP results when training the neural network on different roots of unity, corresponding to different r. The third column shows the roots relevant to calculating the output, the fourth column shows the corresponding fractional levels k obtained by writing the relevant roots as $e^{2\pi i/(k+2)}$, and the final column the error when training the network on only the relevant roots.

$|J_2(K; e^{3\pi i/4})|$, we find the formula

$$V_{3/4}(K) = 6.20 \log\left(|J_2(K; e^{3\pi i/4})| + 6.77\right) - 0.94 \ , \tag{2.4.1}$$

which predicts the volume with an error of 2.86% (Figure 2.4, left). This result is comparable to some of the best performances in Table 2.1 and performs only slightly worse than the neural network which had access to the full Jones polynomial. Repeating the analysis for $|J_2(K; e^{4\pi i/5})|$, the optimized formula is

$$V_{4/5}(K) = 4.89 \log\left(|J_2(K; e^{4\pi i/5})| + 1.91\right) + 0.63 \ , \tag{2.4.2}$$

which predicts the volume with an error of 3.33% (Figure 2.4, right). Due to the fact that the error increase is small and the parameter reduction is enormous in passing from the black box neural network to the formulas (2.4.1) and (2.4.2), we may conclude that we have more or less successfully reverse engineered the simple function which is learned by the network. The dependence of our approximation

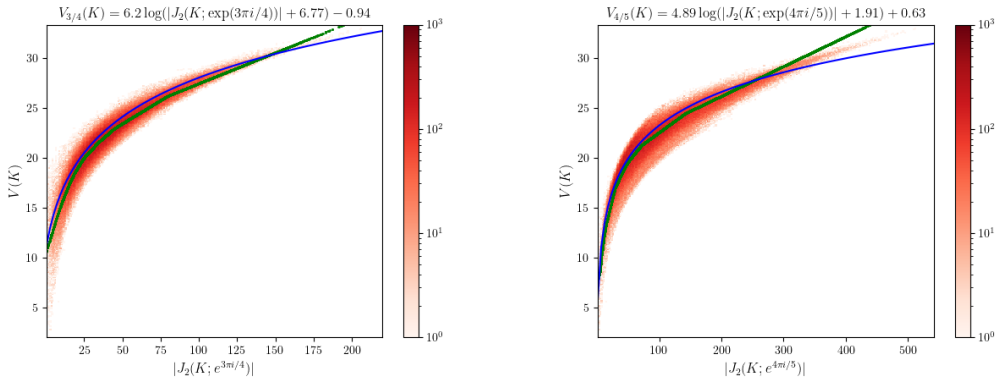


Figure 2.4: (Left) Best fit for the hyperbolic volume as a function of $|J_2(K; e^{3\pi i/4})|$. The fitted function (blue) predicts the volume to 97.14% accuracy and the neural network trained on the magnitude of $J_2(K; e^{3\pi i/4})$ (green) achieves 96.55% accuracy. (Right) Best fit for the hyperbolic volume as a function of $|J_2(K; e^{4\pi i/5})|$. The fitted function (blue) predicts the volume to 96.67% accuracy and the neural network trained on the magnitude of $J_2(K; e^{4\pi i/5})$ (green) achieves 96.30% accuracy. Both plots display (J_2, Vol) data pairs as a density cloud where darker red regions contain more points.

formulas only on the absolute value of a Jones polynomial evaluation aligns nicely with the unchanged performance of the neural network when we drop the maximum and minimum degree information or cyclically permute the polynomial coefficients, as both operations leave the absolute value unchanged (see Appendix 2.E.2 for other experiments with input representation). Moreover, when performing regression on only a subset of the dataset and increasing the size of that subset, the determined coefficients a, b, and c are roughly converging to some constant values (Figure 2.5, left), which gives us confidence that these approximation formulas should perform well even outside of our dataset. As more knots are included in the regression, the error on the total dataset reaches its optimal value smoothly (Figure 2.5, right). For comparison, a neural network trained on $|J_2(K; e^{3\pi i/4})|$ predicts the volumes to 96.55% accuracy, and a neural network trained on $|J_2(K; e^{4\pi i/5})|$ predicts the volumes to 96.30% accuracy. We observe that at large magnitudes for $J_2(K; e^{i\theta})$, both the best fit function and the neural network start to differ from the true volumes. This effect is attributable to the sparsity of data.

In [72], it was noted that for alternating knots up to 13 crossings that $V(K)$ was roughly proportional to $\log|J_2(K; -1)|$. As a comparison to our results, we fit the functional ansatz $a\log(|J_2(K; -1)| + b) + c$ for alternating and non-alternating knots (Figure 2.6). For knots up to and including 16 crossings, the functions predict the volumes of alternating knots to 94.24% accuracy and the volumes of all knots to 93.84% accuracy. While this phase performs worse than the ones found by our layer-wise relevance propagation, it is worthwhile to understand the extent to which

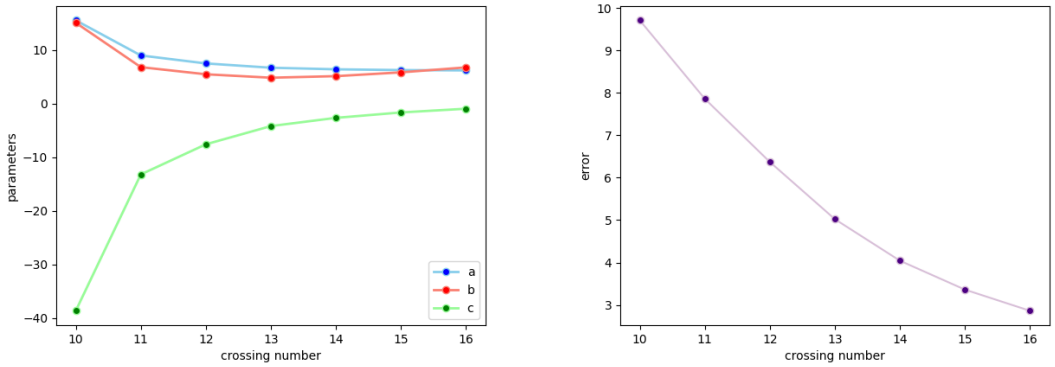


Figure 2.5: Variation of fit parameters with regression data. As more regression data is included, the fit parameters converge to some constant values. This indicates that the approximation formulas should perform well outside of our dataset.

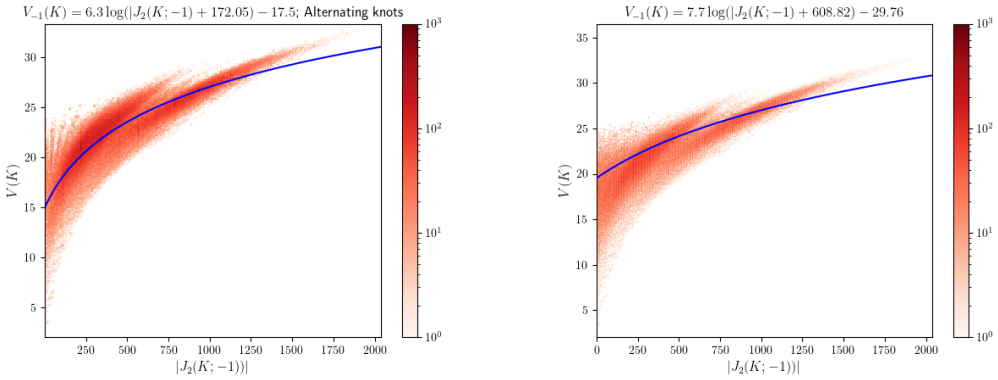


Figure 2.6: Comparison of (2.4.1) with a fit of the same functional ansatz on $J_2(K;-1)$; as in [72]. When applied to alternating knots only (left), the function predicts the volume with a mean error of 5.76%. Fitting the same ansatz to all knots (right) predicts the volume with an error of 6.16%.

we can vary the particular phase used in this approximation ansatz and still retain reasonable performance, since we did find that other roots of unity are relevant for different values of r in Table 2.1.

To explore this possibility, we optimize the ansatz $V(K) = a\log(|J_2(K;e^{ix})|+b)+c$ in `Mathematica` and report the mean absolute error under the constraint $x \in [0,\pi]$ (Figure 2.7) for all knots up to and including 16 crossings.[19] The minimum error of

[19] Amusingly, as is now ubiquitous in high energy theory, the graph evinces the customary dip, ramp, plateau shape [93]. Of course, here these features are not connected to random matrix theory in any clear way. A tentative explanation for the location of the dip makes use of the saddle point form of the path integral (2.2.10), where there is a reasonable spread of this value for different volumes in the region $k \in [0.5, 0.8]$. However, we are far from the semiclassical limit, and cannot really trust (2.2.10).

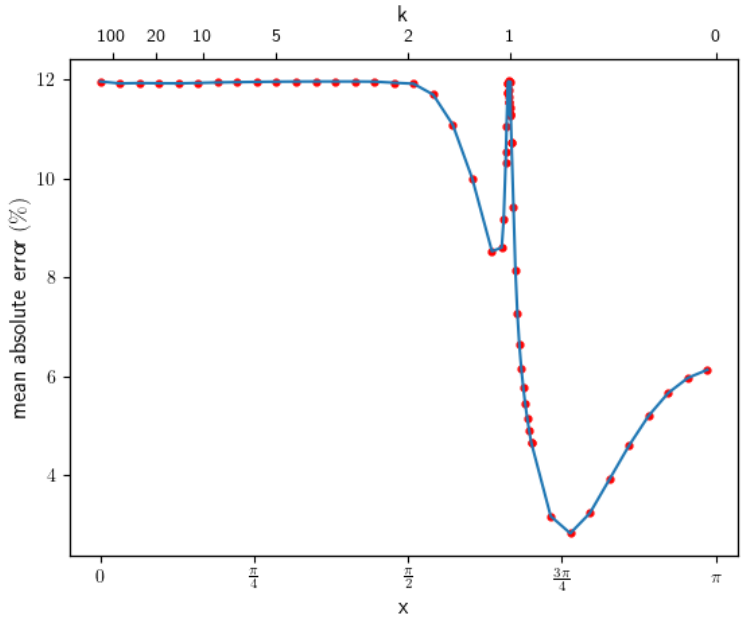


Figure 2.7: Mean absolute errors as a function of x when optimizing the ansatz $V(K) = a\log(|J_2(K;e^{ix})| + b) + c$. We re-optimize the constants a, b, and c for each individual value of x. The minimum error of 2.83% occurs at $x = 2.3$. This is evidence that we can approximate the hyperbolic volume from the Jones polynomial evaluated at a wider variety of fractional levels than just $k = \frac{1}{2}$ or $k = \frac{2}{3}$.

2.83% is realized for $x = 2.3$. We can relate this to the relevant roots in Table 2.1 by noting that this root is similar to $e^{3\pi i/4}$. This corresponds to a fractional level of $\frac{2}{3}$ appearing in the $r = 6$ and $r = 14$ lines of Table 2.1. Notice also that there is a plateau in error rate at around 12% for $x \lesssim \frac{\pi}{2}$. The error of 12% is significant; if we take $V(K) = V_0$, where V_0 is the average volume of all knots up to 16 crossings, the error for this constant approximation function is 11.97%. This represents a latent correlation in the dataset, and if we have an error rate greater than or equal to this then our approximation formula is learning essentially nothing. This error plateau, and other interesting features of Figure 2.7 like the error spike around $k = 1$, will be analyzed in the context of Chern–Simons theory in Section 2.4.2. Before moving on, we note that upon closer inspection of the region around $k = 1$, the single error spike shown in Figure 2.7 is actually resolved into two separate spikes at $k = 1$ and $k = 1.014$ with a shallow dip of roughly 0.7%. We do not have a good understanding of this feature, and we have not ruled out numerical errors as its source, because our `Mathematica` fitting procedures give convergence warnings at these levels. The robust conclusion that we claim here is only that there is a large error spike around $k = 1$ which reaches up to the height of the plateau.

For completeness, we also record the bounding curves of the dataset, computed using our approximation ansatz. The curves

$$f(|J_2(K;e^{3\pi i/4})|) = 6.053\log(33.883 + |J_2(K;e^{3\pi i/4})|) + 0.798\ , \tag{2.4.3}$$

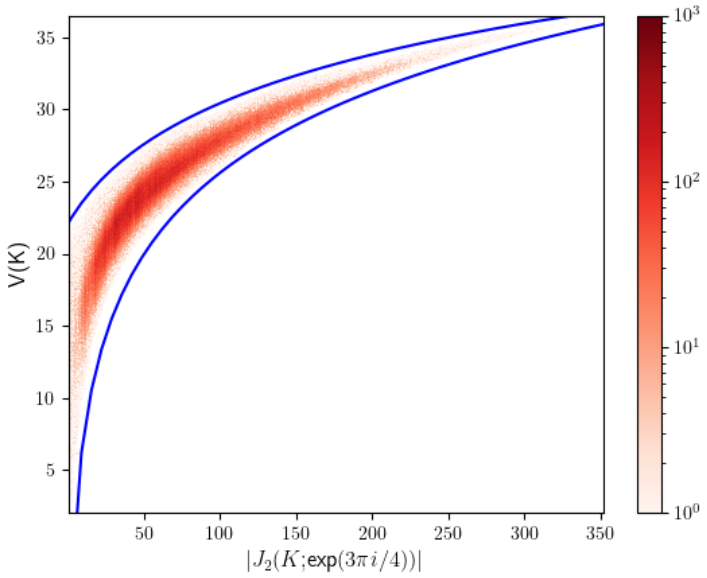


Figure : Bounding curves for the hyperbolic volumes up to and including 16 crossings.

$$g(|J_2(K;e^{3\pi i/4})|) = 8.210\log(0.695 + |J_2(K;e^{3\pi i/4})|) - 12.240\ , \tag{2.4.4}$$

completely enclose the hyperbolic volumes of the knots up to and including 16 crossings (Figure .)

2.4.2 Implications in Chern–Simons theory

Using machine learning and data analysis techniques, we determined that the relevant information about $\mathrm{Vol}(S^3\setminus K)$ contained in $J_2(K;q)$ can be efficiently extracted with a single evaluation of J_2 at a root of unity $q = e^{2\pi i/(k+2)}$, which corresponds to fractional level $k = \frac{1}{2}$ or $k = \frac{2}{3}$. At $k = \frac{2}{3}$, for example, we have $q = e^{3\pi i/4}$, and we find that the simple function (2.1.2) reproduces the hyperbolic volume of K with almost no loss of accuracy compared to the full neural network of [74]. The fractional level $k = \frac{1}{2}$ has been studied in the context of Abelian Chern–Simons theory [50], but has not played a significant rôle in the non-Abelian theory we study here, and level $k = \frac{2}{3}$ (to our knowledge) has hardly been studied at all. Of course, due to gauge invariance, non-Abelian Chern–Simons theory cannot be defined in the naïve way at fractional level, and it is for this reason that the analytic continuation techniques of [76] have an intriguing bearing on our observations.

Our numerical results suggest the Lefschetz thimbles involved in the analytically continued path integral at $\gamma = 2$ or $\gamma = \frac{3}{2}$ are somehow related to the set of thimbles which appear semiclassically at $\gamma = 1$. Perhaps the set of relevant thimbles around $\gamma \in (\frac{3}{2}, 2)$ contains the geometric conjugate $SL(2,\mathbb{C})$ connection that we

expect semiclassically at $\gamma = 1$.[20] This interpretation suggests that the geometric conjugate connection contributes to the path integral for the maximum number of knots in our dataset around $\gamma = \frac{3}{2}$, or $k = \frac{2}{3}$, since that is where the approximation formula performs optimally. This location coincides with the bottom of the dip in Figure 2.7. Similarly, the ramp between $\frac{2}{3} < k < 2$ (ignoring the spike at $k = 1$) is consistent with the interpretation that large fractions of knots are beginning to lose the geometric conjugate connection for $k > \frac{2}{3}$. That the dip begins at $k = 0$ with an error rate which is already fairly low may signal that many knots retain the geometric conjugate connection even as γ becomes large.

We must emphasize that we are not working in the semiclassical limit of large n and k. Instead, we are actually in the opposite limit where both quantities are $O(1)$. Therefore, quantum effects in the path integral are strong, and it is not clear how to isolate the contribution of the critical points. Nevertheless, it seems that our machine learning techniques suggest this is approximately possible with some simple processing of the path integral result. This phenomenon is a bit reminiscent of the ease with which one-loop quantum corrections are incorporated in Chern–Simons theory, through a simple shift $k \to k + 2$ and $n \to n + 1$ in the semiclassical formulas. Here, we have an effect at the level of the analytically continued path integral, where a simple multiplication and shift can absorb quantum corrections to a reasonable degree. It would be interesting to try to make this precise.

It is also instructive to study the error spike at $k = 1$ in Figure 2.7. Readers familiar with the main message of [76] will understand why this feature is in fact expected: at integer values of the Chern–Simons level with $k + 1 \geq n$, the path integral ratio (2.2.6) receives contributions only from $SU(2)$-valued critical points, and can be analyzed using the original techniques in [51]. In other words, the $SU(2)_k$ current algebra has an integrable representation of dimension $n = k + 1$, and so the techniques of [51] are sufficient and no analytic continuation is necessary. Of course, at $n = k + 2$, there is a vanishing of the bare path integral, but the ratio (2.2.6) remains finite and is sensitive to the volume like the analytically continued bare path integral. That this occurs only at the special value $n = k + 2$ and not for any $n < k + 2$ essentially follows from the fact that the bare expectation value of the unknot does not vanish for $0 < n < k + 2$ with integer k.

[20] The geometric conjugate connection is a flat $SL(2, \mathbb{C})$ connection which has $\mathrm{Im}(W) = -\mathrm{Vol}(M)/2\pi$, which is necessary for the factors of i to cancel and yield $e^{k\mathrm{Vol}(M)/2\pi}$ in a saddle point approximation of $|\mathcal{Z}(M)|$ or the analytic continuation of $|Z(M)|$. It is simply the complex conjugate of the geometric connection we mentioned previously. To prevent any ambiguity, we remark that the object called $\mathcal{A}_+$ in (5.61) of [76] is an instance of what we are calling the geometric conjugate connection.

This observation supports our interpretation that the presence of the geometric conjugate connection is responsible for the approximation formula's success. In this case, we cannot expect to learn anything about the hyperbolic volume at $k = 1$ and $n = 2$ beyond latent correlations in the dataset because the geometric conjugate connection does not contribute to the path integral there. Even if it is present in the numerator of (2.2.6), it will cancel with another contribution for integer $k > n - 2$, and so our approximation formula performs as poorly as possible at $k = 1$. However, there are sharp improvements to the approximation formula just below or above this value, where k again becomes fractional and we may find a contribution from the geometric conjugate connection.

Intriguingly, there seems to be a transition around $\gamma = \frac{2}{3}$ (equivalently, $k = \frac{3}{2}$) where the approximation formulas begin to improve over the maximum error, and [76] found that the geometric conjugate connection for the figure-eight knot is added to the path integral by a Stokes phenomenon precisely at this value. The appearance of this maximum error plateau, which roughly matches the error of taking a constant approximation function equal to the average volume in the dataset, is consistent with the interpretation that the approximation formula with $k > \frac{3}{2}$ fails to convey anything useful because the geometric conjugate connection is completely absent from the numerator of (2.2.6), rather than being present and canceling for integer k. Perhaps there is a larger class of knots with a Stokes phenomenon at $\gamma = \frac{2}{3}$ which adds the geometric conjugate connection to the path integral, and there are only a few knots (or perhaps none at all) which receive such a contribution for $k > \frac{3}{2}$.

If we follow the suggestion that the success of our approximation formula is signalling the presence of the geometric conjugate connection for most knots in the dataset, we are led to an interesting observation concerning the volume conjecture. As discussed previously, it is well known that the volume conjecture does not converge monotonically in n [73]. Perhaps the reason for this is that the geometric conjugate connection is missing for early terms in the sequence, and appears at a certain point, after which convergence is monotonic. For example, the first term in the sequence involves $|J_2(K; -1)|$, which corresponds to $k = 0$ and therefore $\gamma = \infty$. While it may be that some knots acquire the geometric conjugate contribution by $\gamma = 1$ and never lose it for any $\gamma > 1$, this may not be the case for all knots. If the geometric conjugate connection is lost by Stokes phenomena at some $\gamma > 1$, the first term in the volume conjecture sequence cannot be expected to be a good approximation. We comment further on this in Section 3.5.

2.5 Discussion

In this work, we have utilized neural networks and layer-wise relevance propagation to extract a very simple function $V(K)$ which predicts with better than 97% accuracy the volume of a hyperbolic knot using only a single evaluation of the Jones polynomial at a root of unity. The existence of such a function was predicted in [74], and prior related observations had also been suggestive [72; 94]. The main problem described at the end of [74] was to determine what simple function was being computed by a neural network, and the roughly equal accuracy of our simple function and the neural network is strong evidence that we have succeeded. We also found excellent alignment between the form of the approximation ansatz and the operations on the inputs (dropping degree information, cyclic coefficient permutations) which left the network performance unchanged.

We commented briefly on the implications of this result for analytically continued Chern–Simons theory in Section 2.4.2. It is clear that there is at least some connection, because the root of unity involved in the calculation is not accessible by the standard definition of the path integral: it corresponds to fractional level k. However, precisely what is gained by studying the path integral far from the semiclassical limit around $\gamma = \frac{3}{2}$ or $\gamma = 2$, and why this should be related to the semiclassical limit near $\gamma = 1$, is not at all clear. Our usual intuition from quantum field theory suggests that the strong quantum effects at small k ought to completely mask the particular value of the action at the geometric conjugate connection. Mysteriously, this does not happen, and (at least, *e.g.*, for large $|J_2(K; e^{3\pi i/4})|$) there is an essentially universal way to extract the saddle point value of this particular connection with high accuracy for any hyperbolic knot. We have found some supporting evidence for our interpretation that the success of the approximation formula signals the presence of the geometric conjugate connection in most knots in our dataset. This evidence involved the spike in error near the integer $k = 1$ as well as the rough matching between the location of the relevant Stokes phenomenon for the figure-eight knot and the critical value of γ where the approximation formula begins to perform well.

An interesting future direction would be to try to derive an inequality, along the lines of the volume-ish theorem [90], using analytically continued Chern–Simons theory. Indeed, the volume-ish theorem should generalize to higher colors, with the upper and lower bounds converging in the infinite n-limit. Deducing an inequality seems quite difficult, as the analysis is very involved for each individual knot [76]. Nevertheless, we may hope to find some unifying theme now that we have a specific location of interest (further along the real γ axis than previously suspected). As a

very first step, one would have to understand how to evaluate the full path integral on the relevant Lefschetz thimbles in order to bound the contributions of other critical points.

We observed in Section 2.4.2 that there could be large discrepancies between the value of γ for the early terms in the volume conjecture and the value of γ at which the geometric conjugate connection is added to the path integral. This motivates a new style of volume conjecture which could be engineered to be immediately monotonic. We simply keep $\gamma \approx 1$ throughout the limit, though this must be done to carefully avoid integer level. By avoiding integer level, we mean that, *e.g.*, for $n = 2$, if we simply solve $\gamma = 1$ we find $k = 1$, and we already argued why this evaluation should yield no nontrivial information about the volume. So we must instead begin at some value like $\gamma = 2$, which would correspond to $k = \frac{1}{2}$. Moreover, we should tune this value with n so that we approach $\gamma = 1$ in the semiclassical limit. All these constraints lead to a version of the volume conjecture where we evaluate the relevant path integrals at a candidate value of γ such as

$$\gamma = \frac{n}{n-1} \ . \tag{2.5.1}$$

This corresponds to level

$$k = \frac{(n-1)^2}{n} \ , \tag{2.5.2}$$

which is certainly always fractional for integer $n \geq 2$ since the parities of the numerator and denominator do not match. With this choice, a monotonic version of the volume conjecture would be (recalling that the prefactor $\frac{2\pi}{n}$ in the usual volume conjecture is really $\frac{2\pi}{k}$ from the Chern–Simons perspective)

$$\lim_{n\to\infty} \frac{2\pi n \log |J_n(K; e^{2\pi i n/(n^2+1)})|}{(n-1)^2} = \mathrm{Vol}(S^3 \setminus K) \ . \tag{2.5.3}$$

This conjecture repairs the non-monotonicity in the volume conjecture sequence for the figure-eight knot (Figure 2.9). Of course, we are not guaranteed that all knots still receive a contribution from the geometric conjugate connection at $\gamma = 2$. Indeed, unfortunately, (2.5.3) just barely fails to be monotonic for the knot 7_2. To formulate a more promising conjecture, we should assume that there is some finite radius around $\gamma = 1$ where all knots still receive such a contribution, and begin the sequence there, which would lead to a different functional dependence than (2.5.1). As noted in [73], eventual monotonicity is important in proving that the limit in the volume conjecture actually exists, so it may be easier to prove that the limit exists in some improved version of (2.5.3).

A less technically complex but more computationally intensive direction involves repeating our machine learning analysis for the $n > 2$ colored Jones polynomials. We

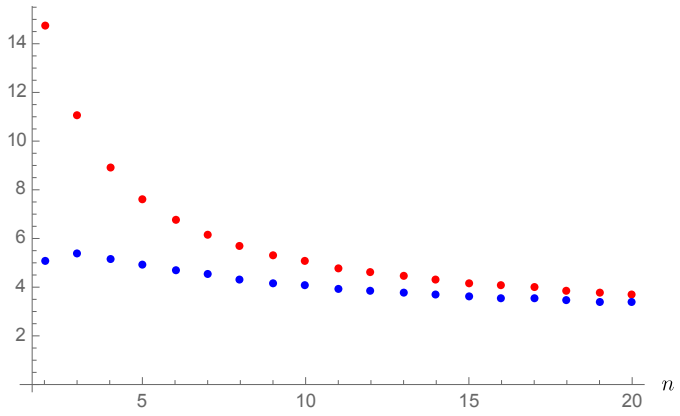


Figure 2.9: The original volume conjecture sequence (left hand side of (2.1.1), blue) for the figure-eight knot, and the modified sequence (left hand side of (2.5.3), red). The non-monotonicity at the beginning of the original sequence is repaired by (2.5.3). Both sequences are computed using Habiro's formula for the colored Jones polynomial of the figure-eight knot [95].

expect to find a similar approximation function, perhaps with different coefficients, which converges quickly to the function which appears in the volume conjecture as n increases. While we attempted (and failed) to formulate a monotonic version of the usual volume conjecture above, there is a more general (and vague) family of conjectures which follow more directly from our numerics:

$$a_n \log(|J_n(K; e^{2\pi i \gamma_n/(n+2\gamma_n-1)})| + b_n) + c_n \approx \mathrm{Vol}(S^3 \setminus K) \ . \tag{2.5.4}$$

The $\approx$ in the above expression means there is a margin for error which monotonically decreases to zero as $n \to \infty$, and furthermore we must have $a_n \to \frac{2\pi}{n}$, $b_n \to 0$, $c_n \to 0$, and $\gamma_n \to 1$ in this limit. We view $a_n - \frac{2\pi}{n}$, b_n, and c_n as "quantum corrections" which account for the large quantum effects in the path integral far from semiclassicality. The quantity γ_n begins around $\gamma_2 = \frac{3}{2}$, according to our numerical results in this work. However, we leave unspecified its precise functional dependence on n, and similarly for the coefficients a_n, b_n, and c_n. The main advantage of (2.5.4) over the simpler (2.5.3) is that we expect (2.5.4) to already be roughly 97% accurate immediately at $n = 2$, whereas for (2.5.3) we do not have such a guarantee. Indeed, we expect that the error in (2.5.4) is actually bounded, whereas (2.5.3) can be arbitrarily wrong at small n, though an improved version of (2.5.3) would still converge monotonically in n. Of course, the price we pay for this immediate accuracy is the introduction of many free parameters in the conjecture.

Appendix

2.A Basic neural network setup

The neural networks used in these experiments were built in `Python 3`, using `GPU-Tensorflow` with a `Keras` wrapper [96]. The neural network has two hidden layers, 100 neurons wide, with ReLU activation functions. The training used an `Adam` optimizer and a mean squared loss function. There were typically 100 training epochs. The network was trained on 10% of the 313,209 knots up to (and including) 15 crossings and tested on the remaining 90%. The code snippet below summarizes the key steps of building and training the neural network.

```
import numpy as np
import tensorflow as tf
from sklearn.model_selection import train_test_split

train_jones, test_jones, train_vols, test_vols =
train_test_split(jones, vols, test_size=0.9)

model = tf.keras.models.Sequential([
tf.keras.layers.Dense(100, activation='relu',
    input_dim=len(train_jones[0])),
tf.keras.layers.Dense(100, activation='relu'),
tf.keras.layers.Dense(1)
])

model.compile(optimizer='adam', loss='mean_squared_error')

model.fit(train_jones, train_vols, epochs=100,
    batch_size=32, verbose=1)
```

Results: Throughout the experiments, different representations/parts of the Jones polynomial were used as input to the neural network. To demonstrate the basic setup, the input data consists of 16-vectors containing the coefficients of the Jones polynomials. The network generally achieves around 2.3% relative error, as shown in Figure 2.A.1.

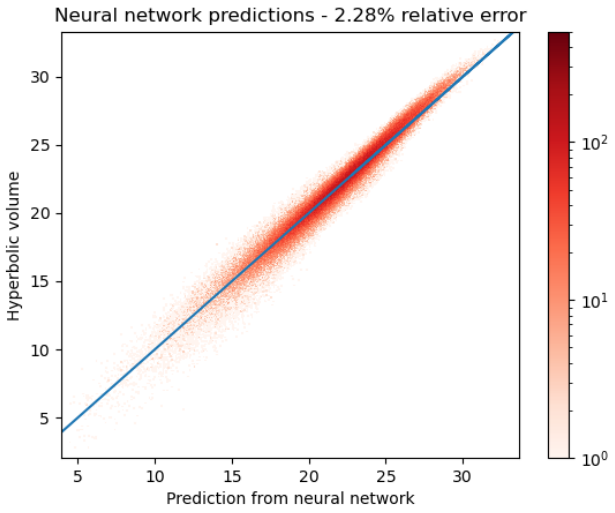


Figure 2.A.1: Results for training the baseline neural network in `Python`. The network has two hidden layers, 100 neurons wide, with ReLU activation functions.

2.B Khovanov homology, Jones coefficients, and volume

2.B.1 Khovanov homology

Machine learning has been applied to string theory for a variety of purposes, beginning in [13; 15; 14; 16]. Many instances in which machine learning has been successfully applied to mathematical physics involve an underlying homological structure [17; 18; 22; 97].[21] In certain cases, analytic formulas inspired by machine learning results have been obtained [102; 103; 104]. Knot theory appears to be no exception to this pattern, as there is an underlying homology theory related to the Jones polynomial known as Khovanov homology [105]. As we will make reference to the homological nature of the Jones polynomial at several points in this supplemental discussion, here we provide a brief explanation of Khovanov homology following the compact treatment in [106].

[21] Such structures appear in string theory fairly often, and more sophisticated machine learning (beyond simple feed-forward network architecture) has also been applied fruitfully to the study of the string landscape. For a small sample, see [16; 98; 99; 100; 101], and also see the review [12] for more complete references in this area.

The Jones polynomial, in its form more familiar to mathematicians, is a Laurent polynomial in a variable q defined by a skein relation[22]

$$\langle \emptyset \rangle = 1, \quad \langle \bigcirc K \rangle = (q + q^{-1}) \langle K \rangle \ , \quad \langle \times \rangle = \langle \asymp \rangle - q \langle) \, (\rangle \ , \quad \hat{J}(K;q) = (-1)^{n_-} q^{n_+ - 2n_-} \langle K \rangle \ , \tag{2.B.1}$$

where $\langle K \rangle$ is known as the Kauffman bracket [107] of the knot K, and $n_\pm$ are the number of right- and left-handed crossings in a diagram of K. The quantity $n_+ - n_-$ is often called the writhe of the knot diagram. We have used the notation $\hat{J}$ because this object is not quite equivalent to our $J_2(K;q)$. It is unnormalized in the following sense:

$$\hat{J}(K;q) = (q + q^{-1}) J_2(K;q^2) \ . \tag{2.B.2}$$

The skein relation for the Kauffman bracket involves two "smoothings" of a given crossing. Of course, in each of these smoothings, the total number of crossings is reduced by one. The recursion terminates when all crossings have been smoothed, so the set of binary strings of length c, $\{0,1\}^c$, is the set of total smoothings for a knot diagram with c crossings.

Khovanov's homology, roughly speaking, begins by assigning a vector space $V_\alpha(K)$ to each string $\alpha \in \{0,1\}^c$. If V is the two-dimensional graded vector space with basis elements $v_\pm$ and $\deg(v_\pm) = \pm 1$, then $V_\alpha(K) \equiv V^{\otimes k}\{r\}$ where the total smoothing α of K results in k closed loops, and $\{r\}$ is the degree shift by r.[23] The height of a string α is defined as the number of ones, $|\alpha| = \sum_i \alpha_i$. The strings with equal height r can be grouped together, and (through some more work) the corresponding vector spaces can be assembled into a doubly-graded homology theory $\mathcal{H}^r(K)$, with Poincaré polynomial

$$Kh(K;t,q) \equiv \sum_r t^r q \dim \mathcal{H}^r(K) \ . \tag{2.B.3}$$

Khovanov proved that this homology theory is a knot invariant, and that its graded Euler characteristic is the unnormalized Jones polynomial

$$\chi_q(K) = Kh(K;-1,q) = \hat{J}(K;q) \ . \tag{2.B.4}$$

Another useful quantity associated with Khovanov homology is the Khovanov rank, given by

$$\mathrm{rank}(\mathcal{H}^\bullet(K)) = \sum_r q \dim \mathcal{H}^r(K) \ . \tag{2.B.5}$$

[22] Knot theorists will notice that this is a rather non-standard skein relation. We will comment on this, and related normalization issues, in Appendix 2.C.

[23] Recall that a graded vector space $W = \oplus_m W_m$ has graded dimension $q \dim W = \sum_m q^m \dim W_m$. Furthermore, the degree shift produces another graded vector space $W\{r\}$ with homogeneous subspaces $W\{r\}_m \equiv W_{m-r}$, so $q \dim W\{r\} = \sum_m q^{m+r} \dim W_m$.

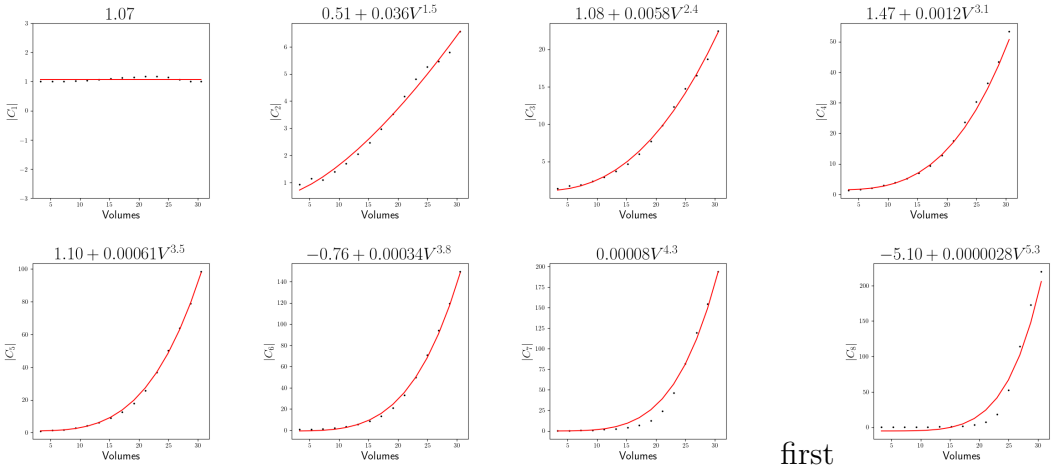


Figure 2.B.1: The absolute values of the (first 8) Jones polynomial coefficients as a function of the hyperbolic volume. The coefficients obey a power law relationship with respect to the volume, where (for c_i the i^{th} coefficient) the constant coefficient decreases (by about an order of magnitude) and the exponent increases as i increases. These power law relationships are symmetric with the last 8 coefficients, as demonstrated in the Table 2.B.1.

The Khovanov rank is correlated with the hyperbolic volume, as noticed in [94], and was compared to neural network prediction techniques in [74].

The main point we wish to convey with this brief summary is that the coefficients of the Jones polynomial are related to dimensions of homology groups, and therefore any success of our machine learning techniques represents another piece of evidence that underlying homological structure is relevant for characterizing physics and mathematics which is machine-learnable.

2.B.2 Power law behavior of coefficients

In this appendix, we study how the coefficients in the Jones polynomial may be individually related to the hyperbolic volume. Through Khovanov homology, we know that each coefficient is measuring the dimension of a certain homology group. Thus, this numerical study is in some sense a generalization of Khovanov's study of patterns in his homology [94], and his observation that coarser properties of his homology are related to the volume (see [74] for more explanation and analysis of this point). Patterns in Jones polynomial coefficients were also studied in [108; 109].

The vectors of coefficients is represented as $\mathbf{J} = (c_1, \ldots, c_8, c_{-8}, \ldots, c_{-1})$, where padding with zeros is done from the center so that, say, c_1 is always the first nonzero coefficient in the polynomial and c_{-1} is always the last nonzero coefficient in the polynomial. The volume is binned into 16 intervals and the mean absolute values of each coefficient is calculated in the bins. Using `Mathematica` curve fitting procedures,

Coefficient	Function	Constant term	V exponent	V coefficient
c_1	1.0662	1.0662	–	–
c_{-1}	1.07462	1.07462	–	–
c_2	$0.506466 + 0.035896V^{1.5}$	0.506466	1.5	0.035896
c_{-2}	$0.709162 + 0.036988V^{1.5}$	0.709162	1.5	0.0369388
c_3	$1.0797 + 0.00578349V^{2.4}$	1.0797	2.4	0.00578349
c_{-3}	$1.1049 + 0.00607516V^{2.4}$	1.1049	2.4	0.00607516
c_4	$1.46782 + 0.0012235V^{3.1}$	1.46782	3.1	0.0012235
c_{-4}	$1.26703 + 0.00188961V^{3}$	1.26703	3	0.00188961
c_5	$1.1018 + 0.0006121V^{3.5}$	1.1081	3.5	0.0006121
c_{-5}	$0.000629931V^{3.5}$	0	3.5	0.000629931
c_6	$-0.758 + 0.000338V^{3.8}$	-0.758	3.8	0.000338
c_{-6}	$0.0003446V^{3.8}$	0	3.8	0.000344
c_7	$0.0000795V^{4.3}$	0	4.3	0.0000795
c_{-7}	$0.0000803V^{4.3}$	0	4.3	0.0000803
c_8	$-5.104 + 0.00000283V^{5.3}$	-5.104	5.3	0.00000283
c_{-8}	$-5.0795 + 0.00000283V^{5.3}$	-5.079	5.3	0.00000283

Table 2.B.1: Fits for the mean absolute value of the coefficients as a function of the volumes.

it was found that the (absolute) coefficients obey power law relationships with respect to the volume; $|c_i| \to aV^b$. For $i \in [1, 8]$, we find that c_i and c_{-i} always obey (nearly) the same power law relationship. Additionally, as i increases, the exponent b increases and a decreases by roughly an order of magnitude. These relationships are given in Table 2.B.1 and Figure 2.B.1.

2.C Normalizations, orientation, and framing

The precise definition of the Jones polynomial can sometimes differ between physicists and mathematicians. Fundamentally, the path integrals which enter (2.2.6) that we have taken to define $J_n(K; q)$ produce invariants of an oriented, framed link in S^3. Actually, since reversing the orientation of a Wilson loop corresponds to conjugation of the associated representation, and representations of $SU(2)$ are real, the $SU(2)$ Chern–Simons path integral defines an invariant of unoriented, framed links in S^3. The writhe factor in (2.B.1) serves to exchange the dependence on framing of the Kauffman bracket for a dependence only on orientation; it is a sort of normalization which removes the extra factors of $q^{3/4}$ that may be introduced by type

I (unframed) Reidemeister moves.[24] This is more or less the same as selecting the 0-framing for all components of a link. So, the path integral formalism incorporates framing changes naturally, whereas the values for the Jones polynomial often quoted in the mathematical literature effectively assume a fixed framing which can always be selected for links in S^3.

An additional detail arises when we make a more direct comparison between the path integrals found in (2.2.6) and the skein relation definition of the Jones polynomial. To begin, we refer to [51], where a skein relation was derived using the path integral directly. In equations (4.22) and (4.23) of [51], the $SU(2)$ skein relation for the path integral was found to be

$$qZ(L_+) - q^{-1}Z(L_-) = (q^{1/2} - q^{-1/2})Z(L_0) \ , \tag{2.C.1}$$

where L_+, L_-, and L_0 represent the same link but with a single crossing modified (if necessary) to be an over-crossing, under-crossing, or uncrossed (away from this crossing, the link is unchanged). Pictorially, (2.C.1) is

$$qZ(\times) - q^{-1}Z(\times) = (q^{1/2} - q^{-1/2})Z(\;)\;) \ . \tag{2.C.2}$$

Other than this relation, the only other piece of information needed to determine the Jones polynomial (as defined in (2.2.6)) is the expectation of the unknot. If we normalize the path integral as in (2.2.6), the unknot has expectation 1.

$$J_2(0_1; q) = 1 \ . \tag{2.C.3}$$

This is equivalent to requiring that the Jones invariants should be multiplicative under connected sum of links, $J_2(L_1 \# L_2; q) = J_2(L_1; q)J_2(L_2; q)$. The skein relation (2.C.1) actually matches exactly with the one written above Remark 2.9 in [110] for the writhe-normalized Jones polynomial $\mathring{V}_L$ (in the notation of [110]).[25] Thus, the ratio in (2.2.6) and $\mathring{V}_L/\mathring{V}_{0_1}$ from [110] will be equivalent up to a power of $q^{3/4}$ associated with a change of framing in the numerator of (2.2.6).

However, other formulations of the skein relation favored by some mathematicians, such as the one used in [111; 112], differ from (2.C.1). The skein relation in [111] is

$$q^{-1}V_{L_+} - qV_{L_-} = (q^{1/2} - q^{-1/2})V_{L_0} \ . \tag{2.C.4}$$

[24] See the discussion around Corollary 2.8 and Remark 2.9 in [110].

[25] There is a typo in equation (6) of [110]. The coefficient on $\mathring{V}_{L_-}$ should be q^{-1}. The correct expression can be derived by multiplying equation (3) of [110] by $q^{-(3/4)w(L_0)}$ and rewriting the coefficients of V_{L_+} and V_{L_-} in terms of $q^{-(3/4)w(L_+)}$ and $q^{-(3/4)w(L_-)}$, respectively. To do so, observe that the writhe obeys $w(L_\pm) = w(L_0) \pm 1$.

It is clear that (2.C.4) is not equivalent to (2.C.1), so which of these determines the "true" Jones polynomial?[26] This difference is not a question of normalization, because the skein relations are all linear, so any normalization obeys the same relation. We will see that the two skein relations are related by a simple, but non-obvious, transformation. Let L' be the mirror of L; in other words, L with all over-crossings changed to under-crossings and vice-versa. Also, let $|L|$ be the number of disjoint circles in L. Then, we have

$$V_{L'}(q) = (-1)^{|L|+1} J_2(L;q) \ . \tag{2.C.5}$$

We begin by noting that $V_{L'}$ obeys a skein relation which is derived by simply switching L_+ and L_- in (2.C.4), since this is the effect of the mirroring operation on a given crossing.

$$q^{-1} V_{L'_-} - q V_{L'_+} = (q^{1/2} - q^{-1/2}) V_{L'_0} \ . \tag{2.C.6}$$

We now observe a few relationships among $|L_+|$, $|L_-|$, and $|L_0|$. Since moving one strand through another does not change the number of components, $|L_+| = |L_-|$. Furthermore, if the crossing in $L_\pm$ is between two different components, then $|L_\pm| = |L_0| + 1$, as in L_0 these two components are joined. Similarly, if the crossing in $L_\pm$ is only one component crossing itself, then $|L_\pm| = |L_0| - 1$, as in L_0 this component is split into two. The upshot is that $|L_0|$ always has opposite parity from $|L_\pm|$, so

$$(-1)^{|L_+|} = (-1)^{|L_-|} = (-1)^{|L_0|+1} \ . \tag{2.C.7}$$

Dividing (2.C.1) by $Z(0_1;k)$ produces an identical skein relation for J_2, and multiplying by $(-1)^{|L_+|+1}$ and using (2.C.7) yields

$$q(-1)^{|L_+|+1} J_2(L_+) - q^{-1}(-1)^{|L_-|+1} J_2(L_-) = -(q^{1/2} - q^{-1/2})(-1)^{|L_0|+1} J_2(L_0) \ . \tag{2.C.8}$$

Interpreted as a skein relation for $(-1)^{|L|+1} J_2(L)$, this equation is equivalent to (2.C.6). Therefore, the quantity $(-1)^{|L|+1} J_2(L)$ obeys the same skein relation as $V_{L'}$, and since they take the same initial value on the unknot, they are equal, proving (2.C.5).

The Jones polynomial is not invariant under the mirroring operation. However, it changes in a predictable way. Since the transformation $q \to q^{-1}$ changes (2.C.4) into (2.C.6), we have

$$V_{L'}(q) = V_L(q^{-1}) \ . \tag{2.C.9}$$

For the purposes of the volume conjecture, however, it does not matter whether we work with L or L'. This is because $S^3 \setminus L$ is homeomorphic to $S^3 \setminus L'$, and since

[26] The relationship between these two skein relations is likely well-known to knot theorists, but we have not been able to find a discussion in the literature, so we provide a self-contained treatment in what follows.

the hyperbolic volume is a topological invariant by the Mostow–Prasad rigidity theorem, their volumes are equal. Furthermore, the evaluations of the colored Jones invariants which appear in the volume conjecture are simply complex conjugated under $q \to q^{-1}$ because q is a root of unity, which leaves unchanged the magnitude $|J_n(e^{2\pi i/n})|$. That being said, when we write numerical results for (2.2.6), we ought to keep in mind that (for example) the right-handed trefoil knot (three over-crossings) obeys

$$J_2(3_1; q) = q^{-1} + q^{-3} - q^{-4} = V_{3_1}(q^{-1}) \ . \tag{2.C.10}$$

So, if our dataset consists of left-handed knots with polynomials derived using (2.C.4), we will write explicit evaluations of (2.2.6) for the corresponding right-handed knots. Interestingly, certain generalizations of the basic volume conjecture (2.1.1) (including the "complex volume conjecture" which involves the Chern–Simons invariant $\mathrm{CS}(S^3 \setminus K)$) are not invariant under an exchange of handedness on one side, which means that whenever one writes something like the complex volume conjecture [113]

$$\lim_{n\to\infty} \frac{2\pi \log J_n(K; e^{2\pi i/n})}{n} = \mathrm{Vol}(S^3 \setminus K) + 2\pi^2 i \mathrm{CS}(S^3 \setminus K) \ , \tag{2.C.11}$$

there is actually a unique skein relation which defines J_n in such a way that (2.C.11) could be true. We have not seen a discussion in the literature of this point, and it would be interesting to determine which of (2.C.4) and (2.C.1) is the correct choice. We suspect the correct relation is the one which comes from Chern–Simons theory, (2.C.1). If we were instead to use the alternate skein relation, we would need to take the complex conjugate of the right hand side in (2.C.11) to have a chance of being correct.

We will explain one final issue related to normalizations. The q^2 in the argument of J_2 on the right hand side of (2.B.2) arises because the skein relation we have written in (2.B.1) matches the rather non-standard choice in [106], which in turn matches [105]. Unlike the more standard relations discussed in [111; 110] that we addressed above, (2.B.1) has the advantage of generating a Laurent polynomial $\hat{J}(K; q)$ on links with any number of components. Therefore, it differs in detail from the path integral ratio (2.2.6), which may require a change of framing to produce a Laurent polynomial. As a consequence of these details, the relationship (2.B.2) should be understood up to a change of framing of the right hand side. Since we will always consider Laurent polynomials in this work, we assume a framing where the left hand side of (2.2.6) is a Laurent polynomial that matches the output of the skein relation in [111].

2.D t-distributed stochastic neighbor embedding of knot invariants

2.D.1 t-distributed stochastic neighbor embedding

t-distributed stochastic neighbor embedding (t-SNE) is a method used to visualize higher-dimensional data in two or three dimensions [114]. It allows one to get a visual understanding of the underlying structure of the data. t-SNE works by preserving the distances between the data points. By centering a Gaussian over the point x_i and measuring the density of all the points j under that Gaussian, the quantity $p_{j|i}$ is calculated. This can be viewed as the conditional probability that i would pick j as a nearest neighbor. In the lower-dimensional space, the analogous quantity $q_{j|i}$ is calculated. Both of these values are symmetrized ($p_{ij} = \frac{1}{2}(p_{i|j} + p_{j|i})$). To ensure that q_{ij} accurately represents p_{ij}, the sum of Kullback–Leibler divergences (relative entropies) over all data points,

$$D_{\mathrm{KL}}(p\|q) = \sum_i \sum_{j \neq i} p_{ij} \log \frac{p_{ij}}{q_{ij}} \ , \tag{2.D.1}$$

is minimized. Thus, the minimization of this quantity with respect to q determines a map from the high dimensional space to the lower dimensional space, which can then be used to project the input. An important parameter which we vary in our application is the perplexity, which quantifies the width of the Gaussians around each data point in terms of the local data point density. Greater perplexity roughly corresponds to considering more points as "neighbors", and extends the Gaussian widths. In this way, the t-SNE algorithm is a sort of constrained optimization of q, and robust results should not depend strongly on the precise form of this constraint.

2.D.2 Volume and Chern–Simons invariant

We used t-SNE to reduce the input vector of Jones polynomial coefficients to a two-dimensional space for visualization purposes (Figure 2.D.1). The t-SNE analysis was performed using `Python`'s `scikit-learn` library [115]. There are two key things to notice in this visualisation. First, the points clearly separate into clusters - this indicates that there is some clustering or hidden structure that separates knots by their Jones polynomial. While this alone could be considered an artifact of the algorithm, we next notice that the points are colored by volume and it is clear that similar colors tend to cluster together. The results here are fairly robust under changes in the internal parameters of the algorithm, like the perplexity, which indicates that the clustering effects we see really do exist in our data.

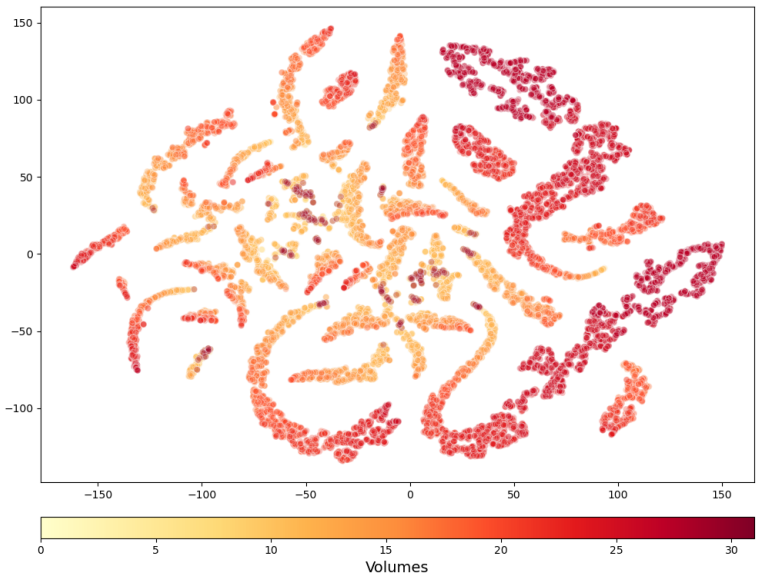


Figure 2.D.1: t-SNE plot of 2000 Jones polynomials, colored by volume (default perplexity of 30).

We emphasize that, even though LRP played the largest rôle in our analysis of the neural network, t-SNE is a purely data-driven technique which is extremely useful for building intuition about which quantities may be machine learnable from others. For instance, Figure 2.D.1 demonstrates noticeable clustering, while Figure 2.D.2 does not. This is evidence that the hyperbolic volume may be machine learnable from the Jones polynomials, while the Chern–Simons invariant may not be. Indeed, this intuition is confirmed in our experiments: none of the neural networks we employed in this work were able to reliably predict the Chern–Simons invariant of a knot from its Jones polynomial.

As a more general lesson, through the creation of plots like Figures 2.D.1 and 2.D.2, we can visualize data without any choice of learning framework. This technique may have promising applications in other areas of experimental mathematics or physics-inspired data science where machine learning sometimes involves prohibitively large computational resources. In those cases, t-SNE could be used to quickly iterate through hypotheses about correlations in the dataset without running a full learning algorithm.

2.E Other experiments

In this appendix, we state the results of a number of mathematical experiments. In applying artificial intelligence to problems in theoretical physics and mathematics, there is no *a priori* understanding of what features are machine learnable. The

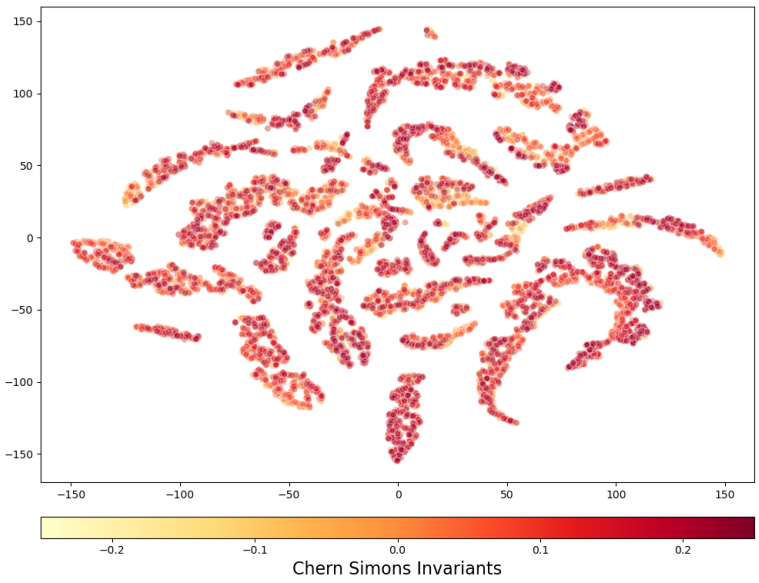


Figure 2.D.2: t-SNE plot of 2000 Jones polynomials, colored by Chern–Simons invariant (default perplexity of 30).

problems for which machine learning provides a useful prediction are often discovered by trial and error. Having isolated such an instance in connecting the Jones polynomial with the hyperbolic volume, we believe it is useful to publicize variations on this central theme.

We as well note some of the less successful experiments among our investigations. When reverse engineering a neural network to obtain analytic results, we expect failed experiments to also guide us. As reported in [74] and in Appendix 2.D.2, the fact that the volume of the knot complement and the Chern–Simons invariant appear on an equal footing in the generalized volume conjecture (2.C.11) but only the former is machine learnable from $J_2(K;q)$ using a simple neural network architecture is particularly striking and perhaps worth highlighting again. It should also be noted that using evaluations of $J_2(K;q)$ at complex phases, we were unable to predict the Chern–Simons invariant.

2.E.1 The HOMFLY-PT and Khovanov polynomials

The HOMFLY-PT polynomial $P(K;a,z)$ [116; 117] is a two variable knot invariant that generalizes both the Alexander polynomial $\Delta(K;q)$ [118] and the Jones polynomial $J_2(K;q)$. In particular,

$$\Delta(K;q) = P(K;1,q^{1/2}-q^{-1/2}) \ , \qquad J_2(K;q) = P(K;q^{-1},q^{1/2}-q^{-1/2}) \ . \quad (2.E.1)$$

We have seen in Appendix 2.B.1 that the Khovanov polynomial $Kh(K;t,q)$ [105; 106] is another such two variable polynomial invariant with the property that

$$J_2(K;q^2) = \frac{Kh(K;-1,q)}{q+q^{-1}} \ . \tag{2.E.2}$$

Using a similar architecture to what is quoted in Appendix 2.A, we machine learn the volume from the HOMFLY-PT polynomial to an accuracy of 93.9% from the full dataset of knots up to 15 crossings. The Khovanov polynomial predicts the volume of the knot complement to an accuracy of 97.2% from a partial dataset of 196,002 knots up to 15 crossings. In both experiments, we use 10% of the dataset for training. The input is a flattened matrix of coefficients appearing in the polynomials where the dimensions of the matrix are determined by the difference between the maximum and minimum degrees of the polynomials in each of the two variables. We notice that though the HOMFLY-PT polynomial contains more information about the knot than the Jones polynomial, it performs significantly worse when used as the input in a two hidden layer neural network.

2.E.2 Representations of $J_2(K;q)$

In [74], the Jones polynomial was faithfully represented as a vector. The first two terms of the input vector are the minimum and maximum degree of the polynomial, and the subsequent entries are the coefficients in the polynomial, padded to the right with zeros to form an 18-vector for knots up to 15 crossings and a 19-vector for knots up to 16 crossings. If we represent the Jones polynomial, again faithfully, as a sparse vector in a long representation whose length is determined by the difference between the maximum and minimum degrees of all polynomials in the dataset, the performance is unchanged.

Surprisingly, unfaithful representations of the Jones polynomial perform almost as well. If we eliminate the degrees from the vector and simply train on the coefficients (that is to say, dimensions of particular Khovonov homology groups that differ for the various inputs), the mean absolute error increases only slightly, from 2.65% to 2.75%. If the input vector sorts the coefficients from smallest to largest, the results remain largely unchanged, but if the ordering of the elements in each vector is randomized, then the mean absolute error is 6.44%, a significant loss of accuracy. While the ordering of the coefficients does matter (or else the randomized vectors would perform just as well), the sorted vectors provide reasonable accuracy, so some rearrangements are allowed. Linear combinations of the coefficients also give an error less than 3%.

Let us drop the degrees but preserve the ordering of coefficients. This yields 16-vector inputs for knots up to 15 crossings. In Appendix 2.D, t-SNE was used

to reduce the data to two dimensions for visualization purposes: it allowed us to see the clustering of the inputs according to volumes. Dimensional reduction may also be useful in determining how the neural network learns the volume. Principal component analysis (PCA) is used as a method to reduce the dimension of the data (length 16 to length 4). With this reduced input, the error grows to 2.80%.

Extending Dunfield's initial observation [72] that for alternating knots $\log|J_2(K;-1)| \propto \mathrm{Vol}(S^3 \setminus K)$, we consider the Taylor expansion of $J_2(K; e^{i\pi(1-x)})$. For instance, for the figure-eight knot 4_1, we have

$$J_2(4_1;q) = q^{-2} - q^{-1} + 1 - q + q^2 \,, \tag{2.E.3}$$

$$J_2(4_1;e^{i\pi(1-x)}) = 5 - \frac{10\pi^2}{2!}x^2 + \frac{34\pi^4}{4!}x^4 + \dots \,. \tag{2.E.4}$$

We adopt the convention of dropping factors of i and π and write the first five coefficients as a vector; this is $(5, 0, -5, 0, 1.417)$ in the example 4_1. Using a dataset of $10,000$ alternating knots between 11 and 15 crossings, the error in the neural network prediction of the volume was 6.5%, and the performance was particularly poor at low volumes. Interesting, training on the first coefficient alone, *i.e.*, training on just $J_2(K;-1)$ yields superior performance.

2.E.3 Small networks

In this experiment, we trained a neural network with one hidden layer, five neurons wide, using the relevant roots from the $r = 8$ row of Table 2.1 as input. Since $J_2(K;-1)$ is always real, the input vector is also of length 5. There are 36 parameters to be optimized: 25 from the 5×5 weight matrix after the input layer, 5 from the bias vector added to the input layer, 5 from the weight vector between the hidden layer and the output layer, and 1 for the bias added before the output layer. Averaged over 100 training runs on the same dataset, the network achieved an error of 5.68%. The large standard deviation on the bias vectors (Figure 2.E.1, right) suggests that the bias vectors may be irrelevant in the function that the neural network is learning. Indeed, setting all five elements of the bias vector to zero does not significantly impact the performance of the network. Although three of the five eigenvalues are zero in Figure 2.E.1 (left), the performance is significantly worsened when fewer than five neurons are used.

The fact that such a simple neural network was able to perform so well is evidence that a simple analytic function exists between the Jones polynomial and the hyperbolic volume. The initial intention of this experiment was to write out the function learnt by the neural network and use it to develop a simple analytic function. However, the number of parameters (which is after zeroing out the bias vector

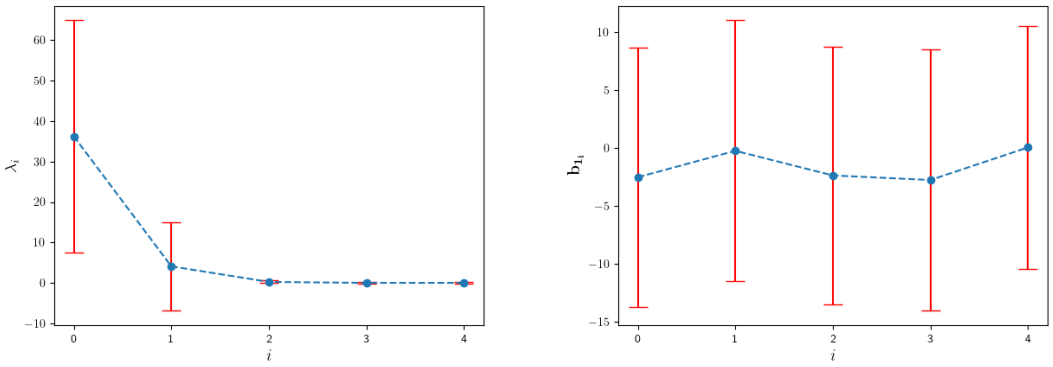


Figure 2.E.1: (Left) Eigenvalues of the matrix $(W_1)^T W_1$, with W_1 being the weight matrix of the first layer, averaged over 100 runs. (Right) The bias vector $\mathbf{b}_1$ of the first layer, averaged over 100 runs. Error bars mark standard deviations.

in the first layer) as well as the application of the activation function made the neural network difficult to interpret, despite its relatively small size.

2.E.4 Symbolic regression

In [119], a combination of deep learning and symbolic regression is used to discover analytic expressions. A deep learning model, such as a Graph Neural Network, is trained on available data. Symbolic expressions are then fitted to the functions learned by the model. The software `PySR` [120] automates this process. We used `PySR` to look for functions of the form $V(K) = f(|J_2(K; e^{3\pi i/4})|)$, allowing the search to use all functions provided in the package. The best fit predicts the volumes to 96.56% accuracy. While the precise formula is unenlightening, it does include a logarithm of $J_2(K; e^{3\pi i/4})$, which reassures us that the formulas proposed in Section 2.4 are reasonable ansätze.

Chapter 3

Learning knot invariants across dimensions

We use deep neural networks to machine learn correlations between knot invariants in various dimensions. The three-dimensional invariant of interest is the Jones polynomial $J(q)$, and the four-dimensional invariants are the Khovanov polynomial $\mathrm{Kh}(q,t)$, smooth slice genus g, and Rasmussen's s-invariant. We find that a two-layer feed-forward neural network can predict s from $\mathrm{Kh}(q,-q^{-4})$ with greater than 99% accuracy. A theoretical explanation for this performance exists in knot theory via the now disproven knight move conjecture, which is obeyed by all knots in our dataset. More surprisingly, we find similar performance for the prediction of s from $\mathrm{Kh}(q,-q^{-2})$, which suggests a novel relationship between the Khovanov and Lee homology theories of a knot. The network predicts g from $\mathrm{Kh}(q,t)$ with similarly high accuracy, and we discuss the extent to which the machine is learning s as opposed to g, since there is a general inequality $|s| \leq 2g$. The Jones polynomial, as a three-dimensional invariant, is not obviously related to s or g, but the network achieves greater than 95% accuracy in predicting either from $J(q)$. Moreover, similar accuracy can be achieved by evaluating $J(q)$ at roots of unity. This suggests a relationship with $SU(2)$ Chern–Simons theory, and we review the gauge theory construction of Khovanov homology which may be relevant for explaining the network's performance.

3.1 Introduction

Knots and links[1] are defined most naturally as embeddings of disjoint circles S^1 in a three-dimensional ambient space, often taken to be S^3 or $\mathbb{R}^3$. However, the numerous isotopy invariants associated with a knot may be more appropriately defined in

[1] In this work, we will focus mostly on knots, though much of the discussion to follow will apply to links as well.

two, three, or four dimensions.[2] While a completely dimension independent theory of knots and their invariants is currently out of reach, progress has been made in relating invariants which are naïvely defined in different dimensions.

Perhaps the most famous example of such a relationship involves the Jones polynomial $J(q)$ of a knot, which is a Laurent polynomial in q with integer coefficients. The original definition by Jones [40; 67] is intrinsically two-dimensional, as $J(q)$ is computed by the trace in a Hecke algebra of a braid representative, which is a two-dimensional projection of the knot (a knot diagram) that has been cut open. However, soon after, Witten discovered [51] that the evaluation of $J(q)$ at a specific root of unity, $q = e^{2\pi i/(k+2)}$, is equal to the appropriately normalized expectation value of a fundamental representation Wilson loop operator along the knot in three-dimensional Chern–Simons gauge theory on S^3 with gauge group $SU(2)$ and integer coupling constant k. This intrinsically three-dimensional perspective on $J(q)$ led to a deeper understanding of various objects in both knot theory and quantum field theory, and in some sense the problem of finding this three-dimensional perspective in the first place was one of the greatest mysteries about the Jones polynomial.[3]

In a later development, Khovanov showed [105] that $J(q)$ may be categorified. What this means is that there exists a $\mathbb{Z}\times\mathbb{Z}$-graded homology theory $\mathcal{H}(L)$, known as Khovanov homology, associated to a link L, for which the graded Euler characteristic is $J(q)$. More precisely, there is a finite decomposition

$$\mathcal{H}(L) = \bigoplus_{m,n} \mathcal{H}^{m,n}(L)\,, \tag{3.1.1}$$

where $\mathcal{H}^{m,n}(L)$ are vector spaces over $\mathbb{Q}$,[4] with $m \in \mathbb{Z}$ and $n \in 2\mathbb{Z} + |L|$, where $|L|$ denotes the number of components in L. Then $J(q)$ is given by

$$(q+q^{-1})J(q^2) = \sum_{\substack{m,n \\ \mathcal{H}^{m,n}(L)\neq 0}} \dim \mathcal{H}^{m,n}(L)(-1)^m q^n\,. \tag{3.1.2}$$

The original definition of $\mathcal{H}(L)$, as explained succinctly in [106], is again intrinsically two-dimensional. Just as with $J(q)$, $\mathcal{H}(L)$ is computed using a two-dimensional projection of the link, and the main technical challenge in [105] is to show that it is actually an invariant (*i.e.*, independent of the chosen projection).

[2]Examples of invariants for each of these dimensions are crossing number, hyperbolic volume, and smooth slice genus, respectively. There are also physics inspired invariants which are naturally defined in five or six dimensions, some of which will be important for our discussion.

[3]Indeed, Witten wrote [51] that "the challenge of the knot polynomials has been to... learn what it is that is three dimensional about two dimensional conformal field theory."

[4]Khovanov homology may be defined more generally over other rings and fields. Indeed, Khovanov's original definition was over the ring $\mathbb{Z}[c]$, where c has degree two.

Despite its two-dimensional definition in [105], there are intimations that $\mathcal{H}(L)$ has a higher-dimensional interpretation, just as $J(q)$ does. One of the most direct hints appears in the work of Rasmussen [121], who showed that for a knot K the homology $\mathcal{H}(K)$ contains enough information to extract an even integer s (the aptly named "s-invariant" of K) whose magnitude supplies a lower bound for the smooth slice genus g:

$$|s(K)| \leq 2g(K)\,. \tag{3.1.3}$$

The smooth slice genus of a knot K is the least integer g such that there exists a smoothly embedded orientable surface $\Sigma \subset D^4$ with $K = \partial\Sigma \subset \partial D^4 = S^3$. In this case Σ is called a slice surface for the knot K, and due to the fact that slice surfaces are embedded in the disk D^4, the invariant g is intrinsically four-dimensional. More generally, linear transformations between $\mathcal{H}(L)$ and $\mathcal{H}(L')$ can be associated with link cobordisms: properly embedded orientable surfaces $\Sigma \subset S^3 \times I$ with $\partial\Sigma = L \sqcup L' \subset S^3 \times \partial I$. Upon concatenating cobordisms, the associated linear transformations simply compose. (Rasmussen's arguments make use of the fact that a punctured slice surface for L can be viewed as a cobordism between L and the unknot.) This strongly suggests that $\mathcal{H}(L)$ has another definition which is natural in four or higher dimensions. Such a definition was proposed by Witten [122], who constructed a candidate for $\mathcal{H}(L)$ as the Hilbert space of a certain five-dimensional gauge theory.[5]

While the Jones polynomial is a success story of the general program of freeing invariants from their dimension dependent definitions, the situation with Khovanov homology, the s-invariant, and the slice genus remains unclear. In particular, the connection between $\mathcal{H}(L)$ and s described by Rasmussen is not entirely geometric, which makes it difficult to assign a particular dimension to the definition of s even if we assume $\mathcal{H}(L)$ is a natural four- or five-dimensional invariant. It is therefore worthwhile to understand what relationships, if any, exist beyond the obvious ones between $J(q)$, $\mathcal{H}(L)$, s, and g. In addition to uncovering new structure in physics, finding different perspectives on these objects could also be helpful in addressing important questions in pure mathematics. One such outstanding question is the smooth four-dimensional Poincaré conjecture (SPC4), which asserts that any manifold that is homotopy equivalent to S^4 is also diffeomorphic to S^4. Because (3.1.3) informs us that g must be positive whenever $s \neq 0$, [125] develops a strategy for constructing exotic four-spheres. Techniques for producing potential counterexamples to SPC4 are also proposed in [48], which rely on showing that certain topologically

[5] Actually, the five-dimensional theory in question is not ultraviolet complete. This fact is discussed at length in [122], and a more powerful formulation in six dimensions is proposed. There are also other physics-based proposals for Khovanov homology [71; 123; 124], but we will focus on the approach described in [122].

slice knots are smoothly slice.[6]

A powerful method for establishing relationships between knot invariants has emerged recently in the form of deep neural networks [81; 74]. The rough idea is to use a dataset to train a universal function approximator to predict one knot invariant from another. Such techniques have been applied fruitfully in knot theory due to the enormous amount of data available on knot invariants, and more general machine learning explorations have yielded new perspectives on the unknotting problem [82] and the volume conjecture [2].[7,8]

The purpose of this chapter is to employ these deep neural network techniques to extract subtle correlations between $J(q)$, $\mathcal{H}(L)$, s, and g, with the broader goal of understanding the properties of each invariant that are surprising from a dimensional perspective. We find strong correlation between properties of $\mathcal{H}(L)$ and s, partially explained by the so-called knight move conjecture in knot theory.[9] Another correlation points toward a novel relationship between a specialization of Khovanov homology and the s-invariant, similar to (but distinct from) the standard passage through Lee homology [121]. Perhaps more surprisingly, we discover large correlation between $J(q)$ or evaluations at roots of unity of $J(q)$ and the invariants s and g. We do not know of a conjecture in knot theory which would explain this, and the physical picture is also unclear. We may phrase this issue as a question, following [51]: what is it that is four-dimensional about the Jones polynomial?[10]

Four sections follow. In Section 3.2, we review the two-dimensional construction of Khovanov homology and Rasmussen's construction of the s-invariant through Lee homology. This will aid our analysis of the mysterious correlation between an unusual specialization of $\mathcal{H}(L)$ and s. In Section 3.3, we briefly review the gauge theoretic ideas of [51; 122] which are relevant for our speculations concerning what

[6] If Σ is a locally flat disk properly embedded in D^4 it is called a topologically slice disk, and $K = \partial\Sigma \subset \partial D^4$ is a topologically slice knot. A knot is smoothly slice when its slice genus is zero.

[7] One of the main surprises in [74; 2] is that a form of the volume conjecture holds approximately for many knots even for the fundamental representation Jones polynomial. The results of [74; 2] could have been partially anticipated from the analytically continued Chern–Simons theory by physicists familiar with the main ideas of [76]. By contrast, we do not know of a physical explanation for the results we find here, even with the apparently most relevant work [122] in mind. We will comment further on this issue later.

[8] Concurrent with this chapter, machine learning methods were applied to motivate a theorem relating the signature, slope, injectivity radius, and volume of hyperbolic knots [126; 127].

[9] Although a counterexample to the knight move conjecture is presented in [128], the statement holds for knots quite broadly and can in fact be proven for certain classes of knots.

[10] Anticipated in [76; 129], discussed in [122], and explained in greater detail in [130], there is a clear sense in which $J(q)$ is the result of a four-dimensional gauge theory calculation. But, as we will discuss, this statement alone does not explain our findings.

higher dimensional information is contained in the Jones polynomial. In Section 3.4, we recall some standard facts about neural networks and give an overview of our experiments and results in extracting correlations among knot invariants. We conclude in Section 3.5 with a discussion and speculate on the possible knot and gauge theoretic explanations of our results.

3.2 Khovanov homology and the s-invariant

3.2.1 Khovanov homology

In what follows, let D be a diagram of the link L. Then each crossing of D can be smoothed in one of two different ways, which we call the 0– and 1–smoothings, respectively (see Figure 3.2.1). Suppose that D has n_+ positive crossings and n_- negative crossings and that we fix an ordering on the combined set of $n = n_+ + n_-$ crossings. Then for for each vertex $v \in \{0,1\}^n$ of the n–dimensional cube $[0,1]^n$ we define a total smoothing D_v by applying the smoothing specified by the k-th coordinate of v to the k-th crossing of D. After applying these smoothings the resulting diagram D_v will be a family of disjoint embedded circles in the plane of projection. Define $s(v)$ to be the number of connected components of D_v, and let $r(v)$ denote the sum of the components of v, *i.e.*, the number of 1–smoothings performed on D to obtain D_v. Now let V be a graded $\mathbb{Q}$–vector space with basis $\{\mathbf{v}_+, \mathbf{v}_-\}$, where $\deg \mathbf{v}_+ = 1$ and $\deg \mathbf{v}_- = -1$. For each total smoothing D_v define the vector space $\mathbb{V}_v = V^{\otimes s(v)}\{r(v)+n_+-2n_-\}$. Here, we let $W\{k\}$ denote the graded vector space obtained by shifting the gradings of elements of W by k. Moreover, to each component of D_v we associate one of the tensor product factors of $\mathbb{V}_v$ via some one-to-one correspondence.

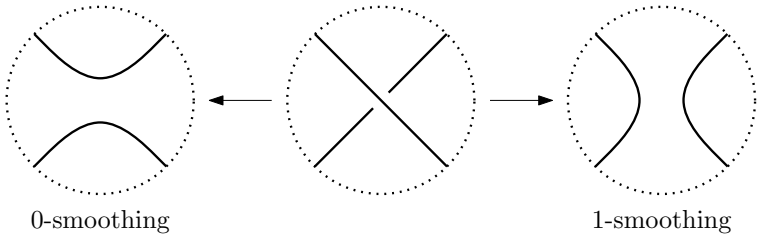


Figure 3.2.1: Two smoothings of a crossing.

Now, suppose that $v = \{v_1, \ldots, v_n\}$ and $v' = \{v'_1, \ldots, v'_n\}$ are vertices in $\{0,1\}^n$ that agree in all but their j-th coordinates, where $v_j = 0$, $v'_j = 1$, and $v_i = v'_i$ for $i \neq j$. It follows then that the sequences of smoothings used to obtain D_v and $D_{v'}$ differ only at the j-th crossing. If e is the edge of the cube $[0,1]^n$ connecting v and v', then we associate an edge map $\varphi_e : \mathbb{V}_v \to \mathbb{V}_{v'}$ to e as follows. First, if changing the smoothing performed at the j-th crossing from a 0–smoothing to a

1–smoothing results in a pair of circles of D_v merging into a single circle, then on the corresponding tensor product factors of $\mathbb{V}_v$ we define the map φ_e as

$$\begin{aligned} \mathbf{v}_- \otimes \mathbf{v}_- &\mapsto 0\,, & \mathbf{v}_+ \otimes \mathbf{v}_+ &\mapsto \mathbf{v}_+\,, \\ \mathbf{v}_- \otimes \mathbf{v}_+ &\mapsto \mathbf{v}_-\,, & \mathbf{v}_+ \otimes \mathbf{v}_- &\mapsto \mathbf{v}_-\,. \end{aligned} \tag{3.2.1}$$

On the other hand, if changing the smoothing performed at the j-th crossing from a 0–smoothing to a 1–smoothing corresponds to a single circle of D_v splitting into two circles, then we define φ_e on the associated tensor product factors of $\mathbb{V}_v$ as

$$\mathbf{v}_- \mapsto \mathbf{v}_- \otimes \mathbf{v}_-\,, \qquad \mathbf{v}_+ \mapsto \mathbf{v}_+ \otimes \mathbf{v}_- + \mathbf{v}_- \otimes \mathbf{v}_+\,. \tag{3.2.2}$$

The map φ_e is then defined to be the identity map on the unaffected tensor product factors of $\mathbb{V}_v$. Defined in this way, it is not difficult to show that the maps φ_e commute around any 2–dimensional face of the cube $[0,1]^n$.

We can now define the Khovanov chain complex $\mathcal{C}(D) = (C^k(D), d^k)$ associated to the diagram D, with chain groups

$$C^k(D) = \bigoplus_{r(v)=k+n_-} \mathbb{V}_v\,, \tag{3.2.3}$$

and differential $d^k : C^k(D) \to C^{k+1}(D)$ given by

$$d^k = \sum \epsilon_e \varphi_e\,. \tag{3.2.4}$$

The sum in (3.2.4) is taken over all edges joining pairs of vertices v and v', with $r(v) = k + n_-$ and $r(v') = r(v) + 1$, and the signs $\epsilon_e \in \{-1, 1\}$ are chosen so that the squares of the resulting differentials are zero (such a choice of the ϵ_e is always possible). Elements of $C^k(D)$ inherit a grading from the vertex vector spaces $\mathbb{V}_v$, which we refer to as the quantum grading, and we define the homological grading of an element of $C^k(D)$ to be k. It can then be checked that the differentials d^k preserve the quantum grading of elements of $C^k(D)$, but increase their homological grading by 1. The homology $\mathcal{H}(L)$ of the resulting chain complex $\mathcal{C}(D) = (C^k(D), d^k)$ is therefore bigraded, with gradings induced by the homological and quantum gradings on $\mathcal{C}(D)$. This homology $\mathcal{H}(L)$, called the Khovanov homology of L, was first defined in [105], where Khovanov showed that up to bigrading-preserving isomorphism it only depends on the underlying isotopy class of the link L. For a knot K the homological grading of $\mathcal{H}(K)$ takes values in $\mathbb{Z}$, while the quantum grading values are all odd integers. It is often convenient to express the Khovanov homology of a knot or link in terms of its Poincaré series, which is a Laurent polynomial defined by

$$\mathrm{Kh}(L; q, t) = \sum_{i,j} \dim(\mathcal{H}^{i,j}(L)) t^i q^j\,, \tag{3.2.5}$$

where $\mathcal{H}^{i,j}(L)$ denotes the subspace of $\mathcal{H}(L)$ consisting of all elements with homological grading equal to i and quantum grading equal to j. We refer to $\mathrm{Kh}(L; q, t)$ as the Khovanov polynomial of L.

3.2.2 Lee homology and Rasmussen's s-invariant

In [131], Lee defined a modification of this homology theory, by constructing a chain complex with the same chain groups $C^k(D)$ as Khovanov's original theory, but with modified edge maps $\varphi'_e : \mathbb{V}_v \to \mathbb{V}_{v'}$. When the edge e corresponds to the merging of two circles the map φ'_e is defined on the affected tensor product factors as

$$\begin{aligned} \mathbf{v}_- \otimes \mathbf{v}_- &\mapsto \mathbf{v}_+ \,, & \mathbf{v}_+ \otimes \mathbf{v}_+ &\mapsto \mathbf{v}_+ \,, \\ \mathbf{v}_- \otimes \mathbf{v}_+ &\mapsto \mathbf{v}_- \,, & \mathbf{v}_+ \otimes \mathbf{v}_- &\mapsto \mathbf{v}_- \,. \end{aligned} \tag{3.2.6}$$

When e corresponds to a circle being split, the edge map φ'_e is defined by

$$\mathbf{v}_- \mapsto \mathbf{v}_- \otimes \mathbf{v}_- + \mathbf{v}_+ \otimes \mathbf{v}_+ \,, \qquad \mathbf{v}_+ \mapsto \mathbf{v}_+ \otimes \mathbf{v}_- + \mathbf{v}_- \otimes \mathbf{v}_+ \,. \tag{3.2.7}$$

Notice now that the resulting differential $d^k_{\mathrm{Lee}} : C^k(D) \to C^{k+1}(D)$ is no longer grading-preserving, but is still non-decreasing in the quantum grading. As a result of this, we can define a filtration on $C^k(D)$, which is a nested sequence of subspaces

$$\{0\} = F^n C^k(D) \subseteq F^{n-1} C^k(D) \subseteq \ldots \subseteq F^m C^k(D) = C^k(D) \,, \tag{3.2.8}$$

where $F^j C^k(D)$ is defined to be the space of all elements of $C^k(D)$ with quantum grading greater than or equal to j. Note then that $d^k_{\mathrm{Lee}}(F^j C^k(D)) \subseteq F^j C^{k+1}(D)$ for all j and k, and hence we obtain a filtered chain complex $\mathcal{C}_{\mathrm{Lee}}(D) = (C^k(D), d^k_{\mathrm{Lee}})$.

The filtration on $\mathcal{C}_{\mathrm{Lee}}(D)$ gives rise to a spectral sequence, whose E^1 page is the ordinary Khovanov homology $\mathcal{H}(L)$, and whose E^∞ page gives the homology $\mathcal{H}_{\mathrm{Lee}}(L)$ of the chain complex $\mathcal{C}_{\mathrm{Lee}}(D)$. Lee [131] showed that for a knot K, the homology $\mathcal{H}_{\mathrm{Lee}}(K)$ is isomorphic to $\mathbb{Q} \oplus \mathbb{Q}$. Rasmussen [121] then proved that these $\mathbb{Q}$–summands both have homological grading zero, and have quantum gradings $s \pm 1$ for some $s \in 2\mathbb{Z}$. This s is an invariant of the knot K, called the Rasmussen s-invariant.

The Lee spectral sequence gives rise to the following important corollary: for any knot K, the Khovanov polynomial can be factored as

$$\mathrm{Kh}(K; q, t) = q^s(q + q^{-1}) + \sum_{\ell \geq 1} f_{2\ell}(q, t)(1 + tq^{4\ell}) \,, \tag{3.2.9}$$

for some Laurent polynomials $f_{2\ell} \in \mathbb{N}[q^{\pm 1}, t^{\pm 1}]$ with positive coefficients, where s is the Rasmussen s-invariant of the knot. The so-called knight move conjecture posits that $f_{2\ell}(q, t) = 0$ for all $\ell > 1$. Although the conjecture is known to be false — Manolescu and Marengon [128] present a counterexample by way of a diagram with 38 crossings — it is known to hold for alternating [132] and quasi-alternating [133]

knots, knots with unknotting number less than or equal to two [134], and homologically thin knots [121].[11]

In the case of homologically thin knots (3.2.9) implies that the Khovanov homology is entirely determined by the Jones polynomial and the s-invariant. Lee proved that alternating knots are homologically thin, and Rasmussen showed that for alternating knots the s-invariant is equal to the classical signature σ. Hence the Khovanov homology of alternating knots is entirely determined by the signature σ and Jones polynomial $J(q)$.

3.2.3 Rasmussen's s-invariant and the slice genus

In addition to proving that s is a knot invariant, Rasmussen also showed that it provides the lower bound on the slice genus of a knot given in (3.1.3). To see why this is, let $\Sigma \subset S^3 \times I$ be a smoothly embedded surface with $\partial\Sigma = L \sqcup L'$, where $L \subset S^3 \times \{0\}$ and $L' \subset S^3 \times \{1\}$ are both links. After a small perturbation of Σ, we may assume that the projection map $h : S^3 \times I \to I$ restricts to a Morse function $h|_\Sigma : \Sigma \to I$. Let $L_t := h|_\Sigma^{-1}(t) = \Sigma \cap (S^3 \times \{t\})$ for all $t \in I$. Then L_t will be a non-singular link embedded in $S^3 \times \{t\}$ for all but finitely many $t \in I$, which correspond to the critical points of the Morse function $h|_\Sigma$. Furthermore, if we select a projection map $\pi : S^3 \to S^2$, then π induces a projection map $\pi_t : S^3 \times \{t\} \to S^2$ for all $t \in I$. After another small perturbation we may assume that $D_t := \pi_t(L_t)$ is a regular link diagram of L_t in S^2 for all but finitely many points $0 < t_1 < t_2 < \cdots < t_m < 1$.

Starting at $t = 0$ and moving up, we see that between any two values t_i and t_{i+1} the diagram D_t changes only by isotopy in S^2. When passing any t_i value, the diagram D_t changes in one of the following ways:

1. By applying a single Reidemeister move to the diagram D_t.
2. By adding a small circle that is disjoint from the rest of the diagram (corresponding to a local minimum of the Morse function $h|_\Sigma$).
3. By removing a small circle that is disjoint from the rest of the diagram (corresponding to a local maximum of the Morse function $h|_\Sigma$).
4. By performing a saddle surgery to to the diagram (corresponding to a saddle point of the Morse function $h|_\Sigma$).

[11] Homologically thin knots are knots K for which all nontrivial Khovanov groups $\mathcal{H}^{i,j}(K)$ satisfy $j - 2i \in \{n-1, n+1\}$ for some integer n. In fact, for such knots $j - 2i = s \pm 1$, where s is the Rasmussen s-invariant.

Then by setting $s_0 = 0$, $s_m = 1$, and selecting points $s_1, \ldots, s_{m-1}$ which satisfy $t_k < s_k < t_{k+1}$, we may cut Σ at each of the heights s_k, thereby splitting Σ into a union of elementary cobordisms $\Sigma = \Sigma_1 \cup \Sigma_2 \cup \cdots \cup \Sigma_m$, where each Σ_i contains precisely one of the singular points t_i. Note that for each k the boundary of Σ_k splits as $\partial\Sigma_k = L_{s_{k-1}} \sqcup L_{s_k}$, and to each Σ_k Rasmussen associates a map $\phi_k : \mathcal{H}_{\mathrm{Lee}}(L_{s_{k-1}}) \to \mathcal{H}_{\mathrm{Lee}}(L_{s_k})$ on the Lee homology of its boundary components. Each of these maps is an isomorphism, and to the full cobordism Σ Rasmussen defines an isomorphism $\phi_\Sigma : \mathcal{H}_{\mathrm{Lee}}(L) \to \mathcal{H}_{\mathrm{Lee}}(L')$ given by the composition $\phi_\Sigma = \phi_m \circ \cdots \circ \phi_2 \circ \phi_1$.

Although each ϕ_k is an isomorphism, they behave differently with respect to the induced filtration on their respective Lee homology. When the singular level t_k corresponds to performing a single Reidemeister move on the diagram $D_{s_{k-1}}$, the map ϕ_k is a filtered map of degree zero. On the other hand, when t_k corresponds to adding or removing a small circle to the diagram $D_{s_{k-1}}$, the map ϕ_k is a filtered map of degree one. Finally, ϕ_k will be filtered of degree -1 whenever t_k corresponds to performing a saddle surgery on the diagram. It follows then that the map ϕ_Σ will be filtered of degree $v - e = \chi(\Sigma)$, where v is the number of local maximum and minimum points in Σ, e is the number of saddle points, and $\chi(\Sigma)$ is the Euler characteristic of Σ.

Suppose now that $K \subset S^3$ is a knot, and that F is a slice surface for K of genus g_F. In other words, F is a compact oriented surface, smoothly embedded in D^4, with $\partial F = K$. By removing a disk from F, we obtain a cobordism Σ between K and the unknot, which induces a filtered map ϕ_Σ of degree $\chi(\Sigma) = \chi(F) - 1 = -2g_F$. However, as $\mathcal{H}_{\mathrm{Lee}}(K) \cong \mathbb{Q} \oplus \mathbb{Q}$ is supported at quantum gradings $s \pm 1$, while $\mathcal{H}_{\mathrm{Lee}}(\bigcirc) \cong \mathbb{Q} \oplus \mathbb{Q}$ is supported at quantum gradings ± 1, it follows that $\mathrm{degree}(\phi_\Sigma) = -2g_F \leq -s$. By repeating the same calculation with the mirror image of K (and observing that taking them mirror image of a knot changes the s-invariant by a sign), we obtain the bound in (3.1.3).

3.3 Gauge theory and knot invariants

There are well-known gauge theory constructions for some of the knot invariants under consideration, and it will be useful for us to review them. To begin, we discuss the three-dimensional Chern–Simons [51] and four-dimensional super-Yang–Mills [129; 130; 122] pictures of $J(q)$. We also briefly review the five- and six-dimensional pictures of $\mathcal{H}(L)$ from [122].

3.3.1 3d Chern–Simons theory

Let us recall how the Jones polynomial arises in Chern–Simons theory [51]. The Chern–Simons action is

$$S_{\rm CS}(A) = \frac{k}{4\pi}\int_M \mathrm{tr}\big(A\wedge dA + \frac{2}{3}A\wedge A\wedge A\big)\,, \tag{3.3.1}$$

where M is a three manifold and the one-form A is a $\mathfrak{g}$ valued connection, with $\mathfrak{g}$ the Lie algebra of the gauge group G. The Chern–Simons level k is integer quantized to ensure single valuedness of $e^{iS_{\rm CS}}$ under large gauge transformations. Chern–Simons theory is a topological quantum field theory, meaning that correlation functions are independent of the metric on M. The operators of interest are Wilson loops:

$$U_R(\gamma) = \mathrm{tr}_R\ \mathcal{P}\exp\left(\oint_\gamma A\right)\,, \tag{3.3.2}$$

where R identifies a representation of the gauge group G and $\mathcal{P}$ denotes path ordering. Taking M to be S^3, γ to be a knot K embedded as $S^1\subset S^3$, and $G=SU(2)$, the colored Jones polynomial is the unknot normalized vacuum expectation value of the Wilson loop operator:

$$J_n(K;q=e^{2\pi i/(k+2)}) = \frac{\langle U_n(K)\rangle}{\langle U_n(\bigcirc)\rangle}\,, \tag{3.3.3}$$

where

$$\langle U_n(K)\rangle = \frac{1}{Z}\int_{\mathcal{U}}[DA]\ U_n(K)\,e^{iS_{\rm CS}(A)}\,,\qquad Z=\int_{\mathcal{U}}[DA]\ e^{iS_{\rm CS}(A)}\,, \tag{3.3.4}$$

and n labels the n-dimensional irreducible representation of $SU(2)$. The path integrals are taken over the space $\mathcal{U}$ of $\mathfrak{su}(2)$ connections modulo gauge transformations. The shift $k\mapsto k+2$ in the definition of q in (3.3.3) accounts for one loop corrections to the path integral. (The 2 is the dual Coxeter number of $SU(2)$.) The Jones polynomial $J(q)=J_2(K;q)$ corresponds to working in the fundamental representation of $SU(2)$.

The Chern–Simons path integral gives a manifestly three-dimensional definition for evaluations of $J(q)$. However, it does not explain at least two very basic facts about $J(q)$ which are obvious from the original two-dimensional definition. The path integral does not explain why the Jones polynomial should be a polynomial in the first place, and also why it should have integer coefficients. The second of these is perhaps the deeper issue, and it is addressed by Khovanov homology. In Khovanov homology, the coefficients of $J(q)$ are interpreted as (graded) dimensions of homology groups, as in (3.1.1). However, to understand the appearance of Khovanov homology in gauge theory, following [122], we must venture beyond Chern–Simons theory.[12]

[12]It is not actually proven in [122] that the construction described there coincides with Khovanov homology. This point will not concern us in this work. We will continue to describe the homology

3.3.2 4d $\mathcal{N} = 4$ super-Yang–Mills

The first step in constructing Khovanov homology from gauge theory is rewriting the Chern–Simons path integral on M as a four-dimensional path integral in (a twisted version of) $\mathcal{N} = 4$ super-Yang–Mills theory on a manifold with boundary $V = M \times \mathbb{R}_+$. This construction originated in a study of integration cycles in quantum mechanics and Chern–Simons theory [76; 129], as the equivalence can sometimes necessitate considering Chern–Simons theory with a rather exotic integration cycle. As the two path integrals are equal, the $\mathcal{N} = 4$ path integral will produce the Jones polynomial $J(q)$ [130].

The details of the construction will not be so important for us, but it will be useful to understand both what the computation entails as well as just what the variable q represents in the four-dimensional language. In a certain sense, this will tell us something about the Jones polynomial which is intrinsically four-dimensional. In the $\mathcal{N} = 4$ $SO(3)$ gauge theory, there is a theta angle[13] multiplying the following term in the action:

$$S_{\mathcal{N}=4} \supset \frac{\theta^\vee}{64h^\vee\pi^2} \int_V \epsilon^{\mu\nu\alpha\beta} \operatorname{Tr} F_{\mu\nu} F_{\alpha\beta} \,, \tag{3.3.5}$$

where the trace is taken in the adjoint representation and $h^\vee = 1$ is the dual Coxeter number of $SO(3)$. This term is topological in nature, and counts the "instanton number" of a classical solution to the $\mathcal{N} = 4$ field equations. If the instanton number of a solution is n, the saddle point contribution to the $\mathcal{N} = 4$ path integral from (3.3.5) is $e^{-in\theta^\vee}$.[14] The Jones polynomial is supposed to be computed by summing over certain supersymmetric field configurations in the twisted $\mathcal{N} = 4$ theory, and this accounts for the full path integral (up to a ratio of one-loop determinants) due to a supersymmetric localization. The ratio of determinants has unit magnitude by supersymmetry, and its sign is determined by the fermion component. As these sums of exponentials are supposed to reproduce the Jones polynomial, the variable

theory constructed in [122] as Khovanov homology, and it is enough for us that [122] constructs some kind of enhanced Floer-like homology theory which has the Jones polynomial as its graded Euler characteristic.

[13]Following [122], we denote this angle by $\theta^\vee$, as it appears in a gauge theory with gauge group $G^\vee$ that is Langlands or GNO dual to the original Chern–Simons gauge group G. For the Jones polynomial, which arises for Chern–Simons gauge group $SU(2)$, the dual group is $G^\vee = SO(3)$, though in this case the Lie algebras coincide.

[14]We are pretending that the result of the instanton number integral is an integer, but this may not be strictly true. It depends on the boundary conditions on $M \times \mathbb{R}_+$, among other subtle points such as a framing of the spin bundle; see Section 3.5 of [122] for a detailed discussion. The gauge group itself can also play a role, and the fact that $SO(3)$ is not simply connected means that we at least have $n \in \mathbb{Z}/4 + \delta$ where δ depends on the boundary conditions.

q must be simply the weight of the topological term in the path integral:

$$q = \exp(-i\theta^\vee)\,. \tag{3.3.6}$$

This gives us some geometric intuition for how evaluations of the Jones polynomial can carry four-dimensional information. The quantity $J(e^{-i\theta^\vee})$ is the result of the $\mathcal{N} = 4$ path integral. This still has not told us how to produce Khovanov homology, however. For that, we need yet another dimension.

3.3.3 5d maximal super-Yang–Mills

To get Khovanov homology, we must consider a five-dimensional theory. This is motivated by an idea which is suggested by the form of (3.1.1): the Jones polynomial should be produced by the graded trace in some vector space. Such graded trace interpretations of path integrals arise when there is an S^1 factor of the manifold and the Hilbert space of the Cauchy surface is the vector space in question. As the Jones polynomial was obtained previously as a sum over solutions of four-dimensional supersymmetric equations which are essentially the time-independent versions of the lifted five-dimensional equations, we interpret these solutions as approximate supersymmetric ground states of the five-dimensional theory.

The theory in question ought to reduce at low energies to the twisted $\mathcal{N} = 4$ theory we had previously, which means we are considering 5d maximally supersymmetric Yang–Mills theory on $S^1 \times M \times \mathbb{R}_+$. Due to various issues discussed in [122], taking $M = \mathbb{R}^3$ is the appropriate choice to produce something like Khovanov homology. The topological supercharge Q from the $\mathcal{N} = 4$ theory remains a symmetry, and has an action on the Hilbert space of the five-dimensional theory. The candidate for $\mathcal{H}(K)$ in this construction is the cohomology of this supercharge Q acting on the Hilbert space of the five-dimensional theory on $M \times \mathbb{R}_+$.

The resulting $\mathcal{H}(K)$ is $\mathbb{Z} \times \mathbb{Z}$-graded, as expected of Khovanov homology. The "quantum grading" arises from the instanton number which we discussed in the $\mathcal{N} = 4$ context, and the "homological grading" is due to a fermion number symmetry.[15] Of course, in the five-dimensional formulation, the object we call instanton number is really an operator acting on the Hilbert space of the theory, whereas in the previous $\mathcal{N} = 4$ description it was a term in the action.

We mentioned that the Jones polynomial is produced by a graded trace in the space of approximate supersymmetric ground states. It is also given by the graded

[15] Technically, the homological grading is associated with a certain R-symmetry generator. See [122] for a discussion of why this object is essentially a fermion number operator.

trace in the space of exact ground states, as we have just said this is to be identified with $\mathcal{H}(L)$. These computations are equal for the following reason. The space of exact states is constructed in the usual manner of Morse homology. The unbroken supercharge Q acts as a differential on the chain complex of critical points (approximate ground states), and the cohomology of Q yields the true space of ground states. In physics language, this incorporates the effects of instanton corrections to the energy levels of the approximate ground states. However, due to supersymmetry, states can only be lifted from zero energy in boson-fermion pairs with the same instanton number, which leaves the graded trace (a supersymmetric equivariant index) unchanged.

3.3.4 6d (0,2) theory

Though it is not crucial for our discussion, we note that an ultraviolet complete setup is provided in [122] in the form of dimensional reduction from the (appropriately twisted) $(0, 2)$ superconformal field theory in six dimensions.[16] This theory has no Lagrangian description, and is known largely as the worldvolume theory of M5-branes in M-theory. Compactifying the theory on a surface leads to the celebrated AGT correspondence between Liouville theory on the surface and four-dimensional $\mathcal{N} = 2$ superconformal field theories [137].

To get Khovanov homology in this theory, we consider a six-manifold $\mathbb{R}_t \times M \times \mathbb{R}^2$. The factor $\mathbb{R}^2$ admits a $U(1)$ action which has a fixed point at the origin, which we call $p \in \mathbb{R}^2$. Then the proposal made in [122] is that $\mathcal{H}(K)$ is the cohomology of an unbroken supercharge Q acting on the Hilbert space of this theory when the Cauchy surface is $M \times \mathbb{R}^2$, and to obtain the result for a nontrivial knot we need a time-independent surface operator inserted on $\mathbb{R}_t \times K \subset \mathbb{R}_t \times M \times p$. The $U(1)$ action is generated by an object P which obeys $[P, Q] = 0$ and reduces to the instanton number in the previously discussed five-dimensional theory.

3.3.5 Rasmussen's s-invariant and gauge theory

We have discussed the Jones polynomial and Khovanov homology from the perspective of gauge theory following [122] in the interest of understanding what four-dimensional information the Jones polynomial may contain. Previously, we have seen that the s-invariant can often be extracted from Khovanov homology along with the Jones polynomial, so there is at least some naïve relationship between them. We now briefly describe two attempts to construct the s-invariant in gauge theory.

[16] See also [135; 136] for the connection between the five-dimensional super-Yang–Mills theory and the six-dimensional $(0, 2)$ theory compactified on S^1.

The first attempt to construct s in gauge theory uses instanton Floer homology. The resulting knot invariant, $s^\sharp$, was defined by Kronheimer and Mrowka [138] and turns out not to be equal to s, but the two do share some qualitative properties. Instanton Floer homology is related to Khovanov homology, but the two are not exactly the same. Roughly speaking, instanton Floer homology is constructed as the Morse homology of the space of gauge fields on a three-manifold M with the Chern–Simons action playing the role of the Morse function.[17] Critical points are flat connections, and the gradient flow equations are the self-dual instanton equations $F^+ = 0$ of four-dimensional Yang–Mills theory on $M \times \mathbb{R}$. The invariant $s^\sharp$ is then computed from instanton Floer homology in a manner very reminiscent of how s arises from Khovanov homology, as we reviewed in Section 3.2. One constructs a cobordism between K and the unknot, and subsequently studies the properties of the map between instanton Floer homologies induced by this cobordism. It has been shown that $s^\sharp \neq s$ even for the trefoil [138], but the precise relationship between the two quantities is unclear as their magnitudes are both lower bounds on twice the slice genus. One fact which seems worth mentioning is that the instanton Floer homology used in [138] is naturally $\mathbb{Z}/4\mathbb{Z}$-graded,[18] while Khovanov homology is fully $\mathbb{Z}$-graded (arising from the fermion number in 5d super-Yang–Mills). Perhaps more insight could be gained in this direction by considering instanton Floer homology with gauge group $SU(N)$ and using level-rank duality [139].

The second attempt at constructing an invariant like s from gauge theory makes use of the techniques in [122] which we reviewed above. There was a conjecture in knot theory which we discussed around (3.2.9), now known to be false in general [128], involving the Khovanov polynomial $\mathrm{Kh}(K;q,t)$ of a knot and Rasmussen's s-invariant [125].[19] It can be proven for homologically thin knots and states

$$q^s = \frac{\mathrm{Kh}(q,-q^{-4})}{q+q^{-1}} = \frac{\mathrm{Kh}(K;q,-q^{-4})}{\mathrm{Kh}(\bigcirc;q,-q^{-4})}\,. \tag{3.3.7}$$

Indeed, the Khovanov polynomial for the unknot is independent of the variable t. The expression (3.3.7) is structurally similar to (3.3.3). In light of the observation that Khovanov homology may have a formulation as the ground state Hilbert space (on $\mathbb{R}^3 \times \mathbb{R}_+$ with suitable operator insertions) of a five-dimensional gauge theory, this conjecture immediately suggests a trace formula for s, at least when K is thin:

[17] This is certainly related to the constructions in [122], and the Morse theory perspective on Khovanov homology is explained in Section 5.3.2 of [122].

[18] As reviewed in Section 5.3.2 of [122], due to the fact that the Chern–Simons action is only well-defined on the space of gauge fields modulo gauge transformations up to multiples of $2\pi k$, standard Floer theory is actually $\mathbb{Z}/4h\mathbb{Z}$-graded, where h is the dual Coxeter number of the gauge group. In [138], the gauge group is $SO(3)$, which has $h = 1$.

[19] To streamline notation, where the meaning is unambiguous, we drop the label K in writing knot polynomial invariants.

$$\lim_{k\to\infty} k \log\left(\frac{Z_5[K;k]}{Z_5[\bigcirc;k]}\right) = 2\pi i s\,, \tag{3.3.8}$$

where $Z_5[K;k]$ is shorthand for the doubly-graded trace in the cohomology of Q in the Hilbert space on $\mathbb{R}^3 \times \mathbb{R}_+$ with appropriate operator insertions to produce the Khovanov polynomial of the knot K evaluated at quantum grading $q = e^{2\pi i/(k+2)}$ and homological grading $t = -e^{-8\pi i/(k+2)}$. An alternate expression for (3.3.8), which physicists would call a "saddle point approximation," is[20]

$$Z_5[k] \underset{k\to\infty}{\sim} e^{2\pi i s/k + \log 2}\,. \tag{3.3.9}$$

This formula is reminiscent of the volume conjecture [69; 70; 113; 71] which expresses the volume of the knot complement of a hyperbolic knot as an evaluation of the colored Jones polynomial $J_n(q)$ in the large-n limit. However, unlike the volume conjecture, (3.3.8) involves only the $n = 2$ fundamental representation of $SU(2)$ in the associated Chern–Simons theory.

The ordinary Jones polynomial is obtained from evaluating the Khovanov polynomial at the specific value $t = -1$, as in (3.1.1), and this is the point reached above as $k \to \infty$. Moreover, we have $q \to 1$ when $k \to \infty$. We see from this that, at least for certain knots, the analytic structure of the Khovanov polynomial $\mathrm{Kh}(q,t)$ contains information about s near a point where it coincides with the Jones polynomial evaluation, namely $J(1)$. Of course, the function $J(q)$ only appears from $\mathrm{Kh}(q,t)$ on the hypersurface $t = -1$, so we apparently must move off the $t = -1$ surface at a particular angle set by $t = -q^{-4}$ to learn about s.

In general, the knight move conjecture can fail, and counterexamples are known in the literature. These counterexamples have higher order terms ($f_{2\ell}(q,t) \neq 0$ for some $\ell > 1$) in the factorization (3.2.9), which spoils the simple extraction of s from gauge theory that we outlined above. From the physics perspective, the failure of the knight move conjecture corresponds to pairs of additional supersymmetric solutions in the path integral which have quantum numbers that differ by more than one in fermion number or more than four in instanton number.

[20] We have written this formula as if the graded trace in the Hilbert space of ground states can be also written as a path integral. For reasons described in [122], this is not exactly true. It is true for the Jones polynomial because that is a type of index of an operator, but the most general path integral on S^1 will receive contributions from states of nonzero energy. Nevertheless, due to supersymmetric localization, there is a classical solution of the supersymmetric equations with the appropriate quantum numbers; what we cannot say is whether this solution is unique or if there are others which complicate the limit.

3.4 Experiments

Machine learning, and in particular neural networks, have been employed in knot theory to predict knot invariants like the slice genus and Ozsváth–Szabó τ-invariant [81] and the hyperbolic volume [74; 2], as well as to solve problems like the unknot recognition problem [82].[21] Neural networks are particularly useful when working with large datasets, and millions of knots have been tabulated together with efficient algorithms for computing their invariants. Such networks are therefore appropriate tools to study the relationships between knot invariants.

A neural network is a function, $f_\theta(\mathbf{v})$, which approximates the relationship between input and output features, $\mathcal{A} : \mathbf{v}_{\text{in}} \to \mathbf{v}_{\text{out}}$, by adjusting its parameters θ to recreate the assignment $\mathcal{A}$ as closely as possible. A single layer, finite width neural network can approximate any well-behaved function on a compact subset of $\mathbb{R}^N$ [83; 84]. What is noteworthy about the neural networks we employ is that they have relatively small architectures and achieve high accuracy when trained on small fractions of the dataset. This suggests the underlying mathematical structures allow a simple analytic description. In this way, machine learning can function as a discovery tool in mathematics and theoretical physics.[22]

The dataset is divided into training and testing sets. The network uses the training set to minimize a chosen loss function in the space of parameters. The neural network is built out of n hidden layers which perform matrix multiplication by a weight matrix, W_θ^m, addition of a bias vector, $\mathbf{b}_\theta^m$, and elementwise application of an activation function σ. The network can thus be written as the function

$$f_\theta(\mathbf{v}_{\text{in}}) = L_\theta^n\Big(\sigma\big(\dots L_\theta^2\big(\sigma\big(L_\theta^1(\mathbf{v}_{\text{in}})\big)\big)\big)\Big)\,, \qquad L_\theta^m(\mathbf{v}) = W_\theta^m \cdot \mathbf{v} + \mathbf{b}_\theta^m\,. \tag{3.4.1}$$

The values of the neurons in the hidden layers, after application of the activation function, are called the activations $\mathbf{a}^m = \sigma(W_\theta^m \cdot \mathbf{a}^{m-1} + \mathbf{b}_\theta^m)$, where $\mathbf{a}^0 = \mathbf{v}_{\text{in}}$. The role of the activation function is to introduce non-linearity into the network. In this work, we use the Rectified Linear Unit (ReLU) activation function, $\sigma(x) = x\Theta(x)$, where $\Theta(x)$ is the Heaviside step function. The loss function, which in this work is a cross entropy loss function appropriate to classification tasks, is minimized on the training data using the backpropagation algorithm. The gradients of the loss function are computed for each training datum and the parameters are adjusted layer by layer in the network. The batch size of a network defines how many data points are used in one iteration of the algorithm, while the number of epochs determines

[21] As well, [140; 141] study many of the same invariants we do from dimensionality reduction and topological data analysis perspectives.

[22] For two success stories in employing machine learning techniques to discover analytic results in mathematical physics, see [104; 2].

how many passes the network makes through the entire training set.

In this work, we typically employ a neural network with two hidden layers, each consisting of 100 nodes.[23] The networks are trained for between 50 and 100 epochs. The training fraction ranges from 10% up to 50%. The machine learned results were stable under adjustments to the neural network architecture. This includes adding more layers, including more neurons per layer, including dropout layers, or using different activation functions.

3.4.1 Data generation

Data for our experiments were generated via two independent methods: random braid words (with special care given to tracking the slice genus) and random knots via the petaluma model [142]. Invariants for these knots were then computed via the `KnotTheory` package [86; 143] in `Mathematica` and and Schütz's `KnotJob` software [144].

By a theorem of Alexander, any knot can be represented as the closure of a braid. Braids in turn may be represented as finite words in the letters $\sigma_1^{\pm 1}, \sigma_2^{\pm 1}, \ldots, \sigma_{n-1}^{\pm 1}$, where n is the number of strands in the braid. Here, σ_i denotes the simple braid formed by taking n parallel strands, and adding a single positive crossing between the i-th and $(i+1)$-th strands (to obtain σ_i^{-1} we add a single negative crossing instead). By forming a random sequence in the letters $\sigma_1^{\pm 1}, \sigma_2^{\pm 1}, \ldots, \sigma_{n-1}^{\pm 1}$ and taking the closure, we obtain a random link.

Our algorithm for generating random braid words required more care, however, to ensure that we could compute the slice genus of a suitable number of the resulting knots. To do this each random braid began with a randomly selected seed braid. These seed braids were of two types: braids from the KnotInfo database [145] for which the slice genus of the closure was known, and randomly generated quasipositive and quasinegative braids.[24]

Given such a seed braid word β whose closure has known slice genus, we then randomly inserted braid words of the form $\alpha\sigma_i^{\pm 1}\alpha^{-1}$ into β. Each such word we inserted into β either changed the slice Euler characteristic of the closure by ± 1 or

[23] These were constructed with `python` using `TensorFlow` with a `Keras` wrapper. We use `Adam` for optimization. The essential code is quoted in Appendix 3.A.

[24] An n–strand braid is quasipositive if it can be represented as a product of braid words of the form $\alpha\sigma_i\alpha^{-1}$, for some n–strand braid α (quasinegative braids are defined similarly). The slice genus of the closure of any quasipositive or quasinegative braid can be easily computed via the slice-Bennequin inequality [146].

0, depending on the strands involved. Together with the slice genus of the seed braid β, the number of braid words inserted into β allowed us to obtain upper and lower bounds on the slice genus of the resulting knot. These bounds were then combined with slice genus bounds obtained from invariants computed using the `KnotTheory` and `KnotJob` software packages, allowing us to exactly pinpoint the slice genus of many of the knots we produced. Finally, we then applied a sequence of random Markov moves to the braids we obtain to further randomize the dataset.

A smaller, independent dataset was also generated using random petal permutations. Indeed, Adams et al. [147] proved that every knot can be represented via a petal diagram, and that such diagrams can be represented by a permutation of the integers $\{1, 2, \ldots, 2n+1\}$ for some n. By sampling random permutations we thus obtained an independent set of random knots, which we again computed invariants for using the `KnotTheory` and `KnotJob` software packages. In the case of knots obtained via random petal permutations, we have less information regarding the slice genus, however, and hence these knots were excluded from experiments involving the slice genus.

3.4.2 What does the data look like?

Our dataset includes the Khovanov polynomials, Jones polynomials, Rasmussen s-invariants, and signatures of 535239 knots. We also know the slice genus of 82% of these knots (438295 knots). The relationship in (3.1.3) is known to be saturated for 414615 of the knots. The relationship

$$s(K) = \sigma(K)\,, \tag{3.4.2}$$

where σ is the signature of the knot, is true for 96.49% (516450) of the knots. These relationships are summarized in Figure 3.4.1.

As discussed in Section 3.2.2, the Khovanov homology of alternating knots is completely determined by classical invariants (the Jones polynomial $J(q)$ and signature σ).[25] While Howie [149] and Greene [150] both gave topological characterizations of alternating knots, the resulting algorithms for identifying alternating knots are not practical to implement on our dataset. We thus content ourselves with using the Jones and Alexander polynomials, as well as the signature and s-invariant to obstruct knots in our dataset from being alternating. By identifying knots whose Jones or Alexander polynomials are not alternating,[26] along with knots for which

[25] It is natural to ask to what extent the Jones polynomial determines the signature of a knot. While there are knots with the same Jones polynomial and different signature, it is a folklore result that the Jones polynomial determines the signature mod 4. See, for instance, [148].

[26] Alternating knots are known to have Jones and Alexander polynomials which are alternating, meaning that coefficients of terms with consecutive powers alternate sign.

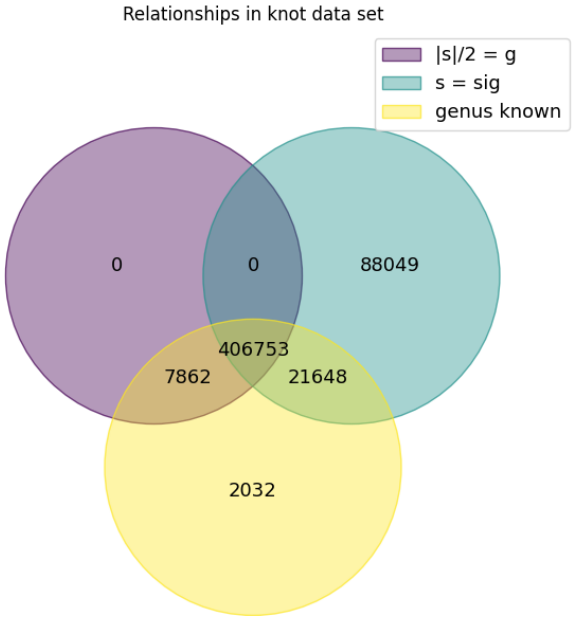


Figure 3.4.1: Relationships between knot invariants in the dataset. There are a total of 535239 knots, 438295 knots with known slice genus, 414615 knots where (3.1.3) is known to be saturated, and 497099 knots satisfying (3.4.2). As an example, the number 2032 represents the number of knots where we know the slice genus, but where (3.1.3) and (3.4.2) are not satisfied.

$\sigma \neq s$, we identified 117510 knots which are not alternating (approximately 21.1% of the total dataset). As none of these conditions are sufficient to imply that a knot is alternating, the remaining knots in our dataset may or may not be alternating.

3.4.3 Khovanov polynomials

In this section, we use the neural network architecture described above to learn the s-invariant and the slice genus from the Khovanov polynomials of the knots. The polynomials are encoded as vectors $\mathrm{Kh}(q,t) = ((e_{q_1}, e_{t_1}, c_1), (e_{q_2}, e_{t_2}, c_2), \ldots)$, where e_{q_i} is the exponent of q, e_{t_i} is the exponent of t, and c_i is the coefficient of that term. For instance, the Khovanov polynomial for the trefoil knot,

$$\mathrm{Kh}(3_1; q, t) = q^9 t^3 + q^5 t^2 + q^3 + q\,, \tag{3.4.3}$$

is written as $(9, 3, 1, 5, 2, 1, 3, 0, 1, 1, 0, 1)$, where we have flattened the vector. The vectors are padded with zeros from the right to ensure consistently sized input. Training the neural network to predict the s-invariant and slice genus based on the Khovanov polynomial, we predict these invariants with 98.30% and 98.6%, respectively (averaged over five runs). The results are robust under a decrease in training size, as shown in Figure 3.4.2. When the prediction is incorrect, we can also look at how far off the predictions are. For s, we look at the value $|s_{\text{true}} - s_{\text{pred}}|$. When the network incorrectly predicts s, it is off by two 91.42% of the time, off by four

5.16% of the time, and off by six 1.9% of the time. For g, when the predictions are incorrect, they are off by one 98.56% of the time and off by two 1.33% of the time.

For the purpose of finding slice knots, we need only learn whether or not g is zero. The network is able to successfully predict this correctly 99.26% of the time, averaged over 5 runs.

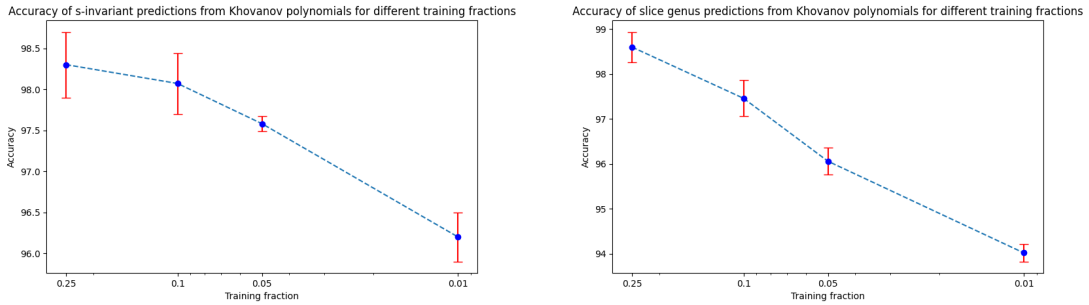


Figure 3.4.2: Left: Accuracy of neural network predictions of the s-invariant for decreasing training fractions. Right: Accuracy of slice genus predictions for decreasing training fractions. In both cases, the neural network inputs are the Khovanov polynomials, as encoded above.

3.4.3.1 Knight move experiments

The inputs here are the Khovanov polynomials $\mathrm{Kh}(q, -q^{-n})$ evaluated at $t = -q^n$. The polynomials are encoded as vectors, $\mathrm{Kh}(q, -q^{-n}) = (e_1, c_1, e_2, c_2, \ldots)$. The $n = -4$ case corresponds to (3.3.7). Indeed, upon making the substitution $t = -q^{-4}$, (3.4.3) becomes

$$\mathrm{Kh}(3_1; q, -q^{-4}) = q^3 + q = q^2(q + q^{-1})\,, \tag{3.4.4}$$

from which we read off $s = 2$. Table 3.4.1 shows the accuracies of the neural network in predicting the s-invariant and the slice genus for $|n| \leq 5$. The mean and variance are computed over five runs.

The excellent results for $n = -4$ are expected because they agree with the conjecture in (3.3.7), which all of the knots in the dataset satisfy. A potential source of misidentifications may stem from ambiguities in how the Khovanov polynomial is factored. We could in principle cast the Laurent expansion another way so that in addition to (3.2.9), we have

$$\mathrm{Kh}(q,t) = q^{s'}(q + q^{-1}) + \sum_{\ell \geq 1} \widetilde{f}_{2\ell}(q,t)(1 + tq^{4\ell})\,, \tag{3.4.5}$$

n	s-invariant accuracy	slice genus accuracy	n	s-invariant accuracy	slice genus accuracy
-5	0.9567 ± 0.0024	0.9700 ± 0.0021	5	0.9377 ± 0.0114	0.9633 ± 0.0040
-4	0.9977 ± 0.0010	0.9452 ± 0.0007	4	0.9457 ± 0.0028	0.9652 ± 0.0037
-3	0.9791 ± 0.0043	0.9716 ± 0.0068	3	0.9554 ± 0.0019	0.9656 ± 0.0066
-2	0.9988 ± 0.0005	0.9456 ± 0.0002	2	0.9612 ± 0.0013	0.9663 ± 0.0040
-1	0.9771 ± 0.0054	0.9751 ± 0.0051	1	0.9577 ± 0.0033	0.9765 ± 0.0011
0	0.9480 ± 0.0021	0.9720 ± 0.0016	–	–	–

Table 3.4.1: Accuracy of neural network predictions for the Rasmussen s-invariant and the slice genus from the Khovanov polynomial evaluated at $t = -q^n$, $n \in [-5, 5]$; $\mathrm{Kh}(q, -q^n)$. The $n = -4$ case corresponds to (3.3.7) and $n = 0$ reduces the Khovanov polynomial to the Jones polynomial. High accuracies are achieved for $n = -4$ and $n = -2$.

with $s \neq s'$.[27] If this is what happens, the effect must be subtle. Picking out the first and second summands only, we have computed the $t = -q^{-4}$ and $t = -q^{-8}$ specializations to test whether these can give rise to different s values. This does occur, but only for the unknot. Our specializations are not able to detect s values that originate from knight moves of mixed sizes, however.

The $n = -2$ results, which are comparable to, and sometimes better than, the $n = -4$ results are more surprising. Table 3.4.2 demonstrates why the accuracy may be so high: we see that a majority of knots at a certain s-invariant have the same polynomial $\mathrm{Kh}(q, -q^{-2})$.

For all knots in the table, the polynomials evaluate to ± 2 at $q = \pm 1$. The Khovanov polynomial admits an expansion (3.2.9). Putting $t = -q^{-2}$, this becomes

$$\mathrm{Kh}(q, -q^{-2}) = q^s(q + q^{-1}) + \sum_{\ell \geq 1} f_{2\ell}(q, -q^{-2})(1 - q^{4\ell-2})\,. \tag{3.4.6}$$

The summands vanish for $q = \pm 1$, and indeed the $q^s = 1$ term is multiplied by ± 2. According to the knight move conjecture, which is false, $f_{2\ell} = 0$ for all $\ell \geq 2$. We can express the most common polynomial for $s = -6, -4, -2, 0$ (and the second most common for $s = -8$ and $s = -10$) as

$$\mathrm{Kh}(q, -q^{-2}) = q^s\left(-\left(\tfrac{s}{2} - 1\right)q + \left(\tfrac{s}{2} + 1\right)q^{-1}\right)\,. \tag{3.4.7}$$

Note that there are some cases where making this substitution produces the second most common polynomial for s, rather than the most common. The second most common polynomial, for $s = 0, -6$, as well as the first most common polynomial for

[27] Though no example is stated, this possibility was already noted in [125].

s	**total count**	**most common** $\mathrm{Kh}(q,-q^{-2})$	**tally**	**second most common** $\mathrm{Kh}(q,-q^{-2})$	**tally**
-18	12	$-q^{-19}+q^{-17}-2q^{-15}+3q^{-13}+q^{-11}$	4	$-q^{-19}+q^{-17}-3q^{-15}+5q^{-13}$	4
-16	28	$-q^{-17}+q^{-15}-2q^{-13}+4q^{-11}$	19	$-q^{-17}+3q^{-11}$	6
-14	66	$-q^{-15}+q^{-11}+2q^{-9}$	37	$-q^{-15}+q^{-13}-q^{-11}+3q^{-9}$	17
-12	264	$-q^{-13}-q^{-11}+4q^{-9}$	151	$-q^{-13}+q^{-7}+2q^{-9}$	64
-10	1173	$-q^{-11}+3q^{-7}$	989	$-4q^{-11}+6q^{-9}$	154
-8	4549	$-q^{-9}+q^{-7}+2q^{-5}$	2304	$-3q^{-9}+5q^{-7}$	2223
-6	15075	$-2q^{-7}+4q^{-5}$	11259	$-q^{-7}+2q^{-5}+q^{-3}$	3775
-4	32621	$-q^{-5}+3q^{-3}$	32582	$q^{-3}+q^{-1}$	16
-2	57591	$2q^{-1}$	57302	$-q^{-3}+4q^{-1}-q$	2261
0	339140	$q+q^{-1}$	339081	$-q^{-1}+5q-2q^{3}$	15
2	52440	$2q$	52175	$-q^{-1}+4q-q^{3}$	235
4	23009	$3q^{3}-q^{5}$	22973	$q+q^{3}$	17
6	7464	$4q^{5}-2q^{7}$	5361	$q^{3}+2q^{5}-q^{7}$	2057
8	1214	$5q^{7}-3q^{9}$	604	$2q^{5}+q^{7}-q^{9}$	593
10	352	$3q^{7}-q^{11}$	311	$6q^{9}-4q^{11}$	22
12	140	$-q^{13}-q^{11}+4q^{9}$	63	$-q^{13}+2q^{9}+q^{7}$	40
14	62	$-q^{15}-q^{11}+q^{13}+3q^{9}$	30	$-q^{15}+q^{11}+2q^{9}$	18
16	30	$-q^{17}+q^{15}-2q^{13}+4q^{11}$	20	$-q^{17}+3q^{11}$	7
18	7	$2q^{11}+q^{13}-q^{15}+q^{17}-q^{19}$	3	$3q^{13}-2q^{15}+q^{17}+q^{11}-q^{19}$	2
20	2	$3q^{13}-q^{17}+q^{19}-q^{21}$	2	–	–
–	535239	–	525270	–	11526

Table 3.4.2: Most common Khovanov polynomials, $\mathrm{Kh}(q,-q^{-2})$, for each s-invariant. The neural network achieves over 99% accuracy when predicting the s-invariant from $\mathrm{Kh}(q,-q^{-2})$.

$s=-4,-8,-10,-12$ can be expressed as

$$\mathrm{Kh}(q,-q^{-2}) = -q^{s-1} + \left(5+\tfrac{s}{2}\right) q^{s+1} - \left(2+\tfrac{s}{2}\right) q^{s+3}\,. \tag{3.4.8}$$

This expression applies for $s \leq 0$. We obtain an expression for positive values of s by sending $q \to q^{-1}$. There also appear to be related expressions for the cases where $\mathrm{Kh}(q,-q^{-2})$ has four or five terms.

At least one aspect of the simple polynomials (3.4.7) and (3.4.8) can be explained by considering a simple change of variables in the Khovanov polynomial. The key observation is that sending $t \to tq^{-2}$ leaves the leading term in the decomposition (3.2.9) unchanged, since $q^s(q+q^{-1})$ is independent of t. This transformation acts in a predictable way on a large class of knots with bounded homological width. In essence, it normalizes the powers of q appearing in the Khovanov polynomial around those appearing at t^0. When the replacement $t=-q^{-2}$ is made, the resulting Khovanov polynomial reduces to alternating sums of terms with a fixed set of quantum gradings centered on s. For knots with homological width two, for instance, the only powers which can appear in these alternating sums are q^{s+1} and q^{s-1} as these are the powers appearing at t^0. For knots with homological width three, the relevant powers are q^{s+3}, q^{s+1}, and q^{s-1}. So, this argument explains the

	−7	−6	−5	−4	−3	−2	−1	0	1	2
1										1
−1									1	
−3								(2 + 1)	1	
−5							3	(1 + 1)		
−7						4	2			
−9					3	3				
−11				3	4					
−13			2	3						
−15		1	3							
−17		2								
−19	1									

Table 3.4.3: The Khovanov homology of $K = 9_{20}$, represented in tabular form. Here the vertical axis represents the quantum grading, while the horizontal axis represents the homological grading. We have highlighted the single pawn move corresponding to the $q^{-4}(q + q^{-1})$ term in (3.2.9) at homological grading 0.

number of distinct powers appearing in the polynomial $\mathrm{Kh}(q, -q^{-2})$ as well as their relation to s.

To illustrate this, consider a homologically thin knot, for example $K = 9_{20}$, with Khovanov polynomial

$$\begin{aligned}\mathrm{Kh}(q,t) =& \frac{1}{q^{19}t^7} + \frac{2}{q^{17}t^6} + \frac{1}{q^{15}t^6} + \frac{3}{q^{15}t^5} + \frac{2}{q^{13}t^5} + \frac{3}{q^{13}t^4} + \frac{3}{q^{11}t^4} + \\ &+ \frac{4}{q^{11}t^3} + \frac{3}{q^9t^3} + \frac{3}{q^9t^2} + \frac{4}{q^7t^2} + \frac{2}{q^7t} + \frac{3}{q^5t} + \frac{2}{q^5} + \frac{3}{q^3} + \frac{t}{q^3} + \frac{t}{q} + qt^2\,. \end{aligned} \tag{3.4.9}$$

The polynomial $\mathrm{Kh}(q,t)$ can be represented in table form as shown in Table 3.4.3. Notice that $\mathrm{Kh}(q,t)$ can be decomposed as in (3.2.9) as

$$\mathrm{Kh}(q,t) = q^{-4}(q + q^{-1}) + f_2(q,t)(1 + tq^4)\,, \tag{3.4.10}$$

where

$$f_2(q,t) = \frac{1}{q^{19}t^7} + \frac{2}{q^{17}t^6} + \frac{3}{q^{15}t^5} + \frac{3}{q^{13}t^4} + \frac{4}{q^{11}t^3} + \frac{3}{q^9t^2} + \frac{2}{q^7t} + \frac{1}{q^5} + \frac{t}{q^3}\,. \tag{3.4.11}$$

Expanding out the product $f_2(q,t)(1 + tq^4)$ above, we see that $\mathrm{Kh}(q,t)$ contains terms of the form $\alpha q^m t^n(1 + tq^4)$, for integers m, n, and α, with $\alpha \geq 1$, each of which can be represented as a knight move as shown in Table 3.4.4 (left). The only remaining term in $\mathrm{Kh}(q,t)$ is $q^{-4}(q + q^{-1})$, which represents a lone pawn move in

	n	$n+1$
$m+4$		α
$m+2$		
m	α	

	0
$s+1$	1
$s-1$	1

Table 3.4.4: Left: Knight move corresponding to the term $\alpha q^m t^n(1+tq^4)$. Right: Pawn move corresponding to the term $q^s(q+q^{-1})$.

	-7	-6	-5	-4	-3	-2	-1	0	1	2
-3		1	2	3	3	4	3	$(2+1)$	1	1
-5	1	2	3	3	4	3	2	$(1+1)$	1	

Table 3.4.5: A tabular representation of $\mathrm{Kh}(q, tq^{-2})$, where each diagonal line in $\mathrm{Kh}(q,t)$ has been flattened to a horizontal line. Notice that the terms at homological degree zero remain fixed.

Table 3.4.4 (right). Notice that the location of the pawn move for 9_{20} at quantum gradings -3 and -5 indicates that $s=-4$ for this knot.

We may make the substitution $t \to -q^{-2}$ in two stages. First we make the substitution $t \to tq^{-2}$. This shifts the quantum grading of every term down by -2 times its homological grading, which has the effect of flattening each diagonal line in Table 3.4.3 into a horizontal line as in Table 3.4.5. Note that the terms at homological grading 0 remain fixed when making this substitution. Furthermore, since (aside from lone pawn move pair) the terms of $\mathrm{Kh}(q,t)$ are arranged in knight move pairs, after substituting $t \to tq^{-2}$ we obtain two identical lines (again ignoring the pawn move pair) that are offset by one.

In the new polynomial $\mathrm{Kh}(q, tq^{-2})$, we may now set $t=-1$ to implement our desired substitution $t=-q^{-2}$ in the original polynomial $\mathrm{Kh}(q,t)$. This step may be interpreted as taking the alternating sum of the rows of Table 3.4.5, and multiplying by q^{-3} and q^{-5} respectively. Since the rows are offset by one, the alternating sums will differ by a sign until the pawn move pair is factored in (2 and -2 respectively), which then shifts both sums up by one. In this case we obtain

$$\mathrm{Kh}(q, -q^{-2}) = -q^{-5} + 3q^{-3}\,, \tag{3.4.12}$$

which matches (3.4.7) for $s=-4$. A similar argument may be used to partially explain (3.4.8).

Notice that although this explains the general form that the polynomials $\mathrm{Kh}(q, -q^{-2})$ often take, it does not explain why the coefficients of these terms seem to also be completely fixed by the s-invariant. The path through Lee homology taken by Rasmussen in order to define the s-invariant makes reference only to the remaining quantum grading since Lee homology is isomorphic to $\mathbb{Q} \oplus \mathbb{Q}$ at homological degree 0. Our observation suggests that perhaps s is encoded in some additional way within Khovanov homology, one which does not make reference to Lee homology or the quantum grading.

3.4.3.2 Learning from polynomial evaluations

We can also learn the s-invariant from the polynomials $\mathrm{Kh}(q, -q^{-4})$ and $\mathrm{Kh}(q, -q^{-2})$, evaluated at roots of unity $e^{\frac{\pi i n}{(k+2)}}$ for $n \in [0, k+2)$. The input data for a knot K is represented as a vector $\vec{v}(K; k)$, where the entries at position $2p$ and $2p{+}1$ correspond to the real and complex parts of the p-th evaluation. For instance, taking $k = 3$, the Khovanov polynomial for the trefoil knot at $t = -q^{-4}$, which we write in (3.4.4), is encoded as

$$\vec{v}(3_1; 3) = (2, 0, 0.5, 1.54, -0.5, 0.36, 0.5, 0.36, -0.5, 1.54) \ . \tag{3.4.13}$$

When training on either evaluations of $\mathrm{Kh}(q, -q^{-4})$ or of $\mathrm{Kh}(q, -q^{-2})$, the accuracy of predictions is over 99% for the s-invariant and over 94% for the slice genus, even for $k = 3$, where we train on only five evaluations of the polynomial.

3.4.4 Jones polynomials

The Jones polynomials are encoded, like the Khovanov polynomials, as vectors, $J(q) = (e_1, c_1, e_2, c_2, \ldots)$. Using this input and training on 25% of the data, the network predicts the Rasmussen s-invariant with $\sim 95\%$ accuracy, and the slice genus with $\sim 96.6\%$ accuracy. The results for the s-invariant are fairly robust under a decrease in training size, Figure 3.4.3. When the prediction is incorrect, we can also look at how far off the predictions are. For s, we look at the value $|s_{\text{true}} - s_{\text{pred}}|$. When the network incorrectly predicts s, it is off by two 67.83% of the time, off by four 18.28% of the time. For g, when the predictions are incorrect, they are off by one 81.71% of the time and off by two 17.52% of the time. The accuracy of these networks is surprising, as there are knots with the same Jones polynomial but different slice genus and s-invariant.[28] This means that the neural network is not learning a function, but merely an association.

For the purpose of finding slice knots, we need only learn whether or not g is zero. The network is able to successfully predict this correctly 98.61% of the time,

[28]The knots 5_1 and 10_{132} are a pair of such knots.

k	Rasmussen accuracy	Slice genus accuracy
3	0.9314 ± 0.0013	0.9672 ± 0.0017
4	0.9186 ± 0.0009	0.9601 ± 0.0010
5	0.9650 ± 0.0006	0.9826 ± 0.0009
6	0.9674 ± 0.0007	0.9828 ± 0.0008
7	0.9676 ± 0.0007	0.9825 ± 0.0008
8	0.9673 ± 0.0005	0.9825 ± 0.0006
9	0.9669 ± 0.0011	0.9814 ± 0.0014
10	0.9680 ± 0.0006	0.9826 ± 0.0008

Table 3.4.6: Learning the Rasmussen s-invariant and the slice genus from evaluations of the Jones polynomial at roots of unity, $e^{\pi i n/(k+2)}$ for $n \in [0, k+2)$. This experiment is repeated for $k \in [3, 10]$.

averaged over 5 runs. We can also learn the invariants from evaluations of the Jones

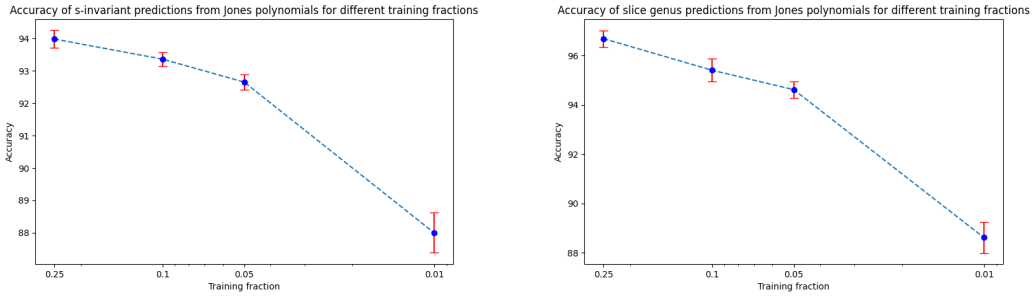


Figure 3.4.3: Left: Accuracy of neural network predictions of the s-invariant for decreasing training fractions. Right: Accuracy of slice genus predictions for decreasing training fractions. In both cases, the neural network inputs are the Jones polynomials, as encoded above.

polynomial at roots of unity. For a fixed k, we generate a list of roots $\{e^{\pi i n/(k+2)}\}$, $n \in [0, k+2)$ at which to evaluate the Jones polynomials. The results for training the network on these inputs, for $k \in [3, 10]$, are given in Table 3.4.6.

3.4.5 Interdependence of results

As discussed previously, the Rasmussen s-invariant provides a lower bound for the slice genus. Our neural networks are able to successfully predict both s and g, but we do not know if the two invariants are learnt independently or if one is learnt via the other. Additionally, since the signature and the s-invariant coincide for most of the knots, it could be that the network is learning s indirectly via the signature. In this section, we aim to diagnose which invariants the neural networks are learning

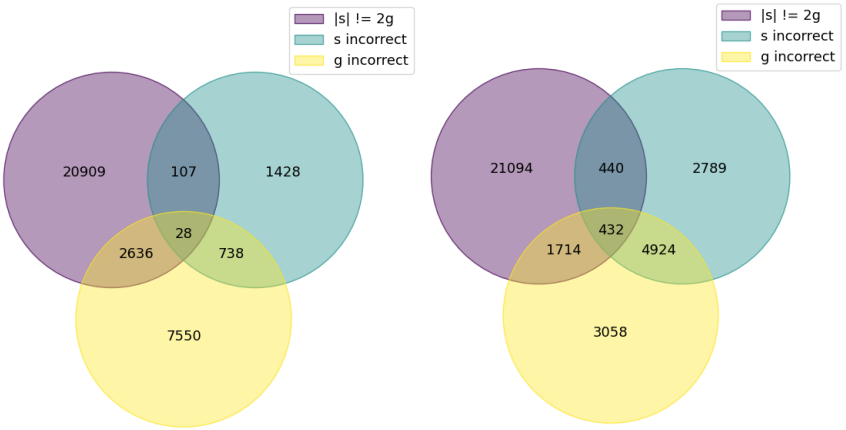


Figure 3.4.4: Comparing the number of s-invariants and slice genuses predicted incorrectly by neural networks with the number of cases where the slice genus saturates the lower bound provided by the s-invariant. Left figure is trained on the Khovanov polynomial and right figure is trained on the Jones polynomial.

directly.

First, we train two neural networks to learn the s-invariant and the slice genus, using the Khovanov (or Jones) polynomials. Then, we count how many times s and g were predicted incorrectly when $|s| = 2g$, versus when $|s| \neq 2g$. These counts are summarized in Figure 3.4.4. When training on the Khovanov polynomial (Figure 3.4.4 left), a total of 2301 s-invariants and 10952 slice genuses are predicted incorrectly. The majority of the s-invariants which are predicted incorrectly lie in the region where $|s| = 2g$. This does not support the claim that the network is learning s via g. A similar trend is apparant when training on the Jones polynomials (Figure 3.4.4 right).

We can perform a similar experiment with the signature and the s-invariant. The results for this experiment are given in Figure 3.4.5. When training on the Khovanov polynomial (Figure 3.4.5 left) a total of 7438 s-invariants and 4280 signatures were predicted incorrectly. When s and σ coincide, s is predicted incorrectly 1648 times and σ is predicted incorrectly 6693 times. When s and σ do not coincide, s is predicted incorrectly 3510 times and σ is predicted incorrectly 2939 times. This experiment does not offer any conclusive evidence as to which invariant the network is learning.

3.4.5.1 Multi-task learning

We can further train a neural network to learn the s-invariant and the slice genus simultaneously. This may provide insight into which invariant the network is learning. In a multi-task problem, we can weight the various tasks. This weight determines

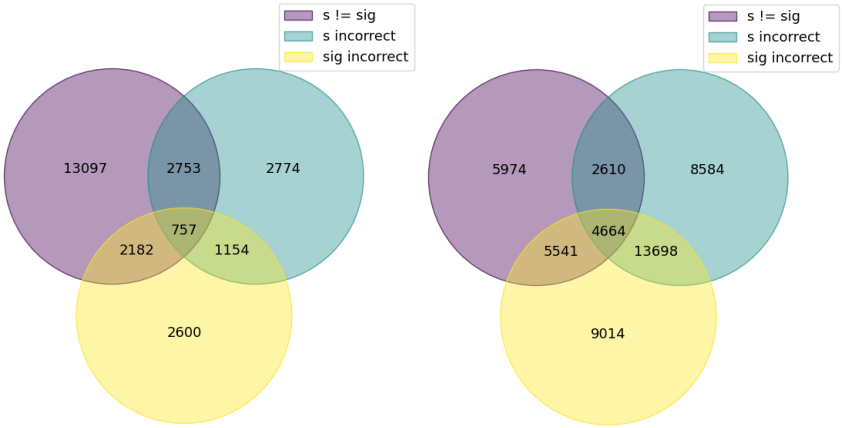


Figure 3.4.5: Comparing the number of s-invariants and signatures predicted incorrectly by a neural network trained on Khovanov polynomials with the number of cases where $s = \sigma$. Left figure is trained on the Khovanov polynomials and right figure is trained on the Jones polynomials.

how strongly that task will influence the change of parameters as the network is trained. We can ask how changing the weights on the two outputs affects the overall learning of the network. Denote the weight of the s-invariants to be $\mathcal{W}_s$ and the weights of the slice genus to be $\mathcal{W}_g$. We find that weighting either the s-invariant or the slice genus more heavily did not significantly influence the performance of the network on either of the invariants.

First, we try this experiment using the Khovanov polynomials as input. For $\mathcal{W}_s/\mathcal{W}_g = 1$, we get accuracies of 0.9952 ± 0.0002 and 0.9866 ± 0.0026 on the s-invariant and the slice genus, respectively. When $\mathcal{W}_s/\mathcal{W}_g = 0.1$, we get accuracies of 0.9924 ± 0.0010 and 0.9859 ± 0.0044. For $\mathcal{W}_s/\mathcal{W}_g = 10$, we get accuracies of 0.9954 ± 0.0004 and 0.9797 ± 0.0025. Repeating this experiment only on the dataset where $|s| \neq 2g$, we get the following results. For $\mathcal{W}_s/\mathcal{W}_g = 1$, we get accuracies of 0.9959 ± 0.0002 and 0.9943 ± 0.0008. When $\mathcal{W}_s/\mathcal{W}_g = 0.1$, we get accuracies of 0.9954 ± 0.0022 and 0.9935 ± 0.0027. For $\mathcal{W}_s/\mathcal{W}_g = 10$, we get accuracies of 0.9960 ± 0.0003 and 0.9948 ± 0.0008.

We can also do this experiment using the Jones polynomials as input. For $\mathcal{W}_s/\mathcal{W}_g = 1$, we get accuracies of 0.9775 ± 0.0007 and 0.9723 ± 0.0043 on the s-invariant and the slice genus, respectively. When $\mathcal{W}_s/\mathcal{W}_g = 0.1$, we get accuracies of 0.9766 ± 0.0006 and 0.9731 ± 0.0032. For $\mathcal{W}_s/\mathcal{W}_g = 10$, we get accuracies of 0.9774 ± 0.0021 and 0.9640 ± 0.0084. Repeating this experiment only on the dataset where $|s| \neq 2g$, we get the following results. For $\mathcal{W}_s/\mathcal{W}_g = 1$, we get accuracies of 0.9803 ± 0.0006 and 0.9786 ± 0.0021. When $\mathcal{W}_s/\mathcal{W}_g = 0.1$, we get accuracies

of 0.9806 ± 0.0008 and 0.9788 ± 0.0010. For $\mathcal{W}_s/\mathcal{W}_g = 10$, we get accuracies of 0.9811 ± 0.0006 and 0.9780 ± 0.0016.

3.5 Discussion

In this work, we have demonstrated that the Jones polynomial of a knot is strongly correlated with the knot's Rasmussen s-invariant and slice genus g. We also found a specialization for the Khovanov polynomial which makes the s-invariant manifest in our dataset, though this specialization is not the one suggested by the knight move conjecture. These correlations were extracted by training deep neural networks to predict s or g with the Jones or specialized Khovanov polynomials as input. We reviewed both the combinatorial construction of Khovanov homology and Rasmussen's invariant as well as the gauge-theoretic constructions in four, five, and six dimensions.

Though we were unable to determine whether the neural network learned s, g, some combination of the two, or something else altogether,[29] there is at least one naïve reason to believe that the Jones polynomial is correlated with s specifically. This is because both arise from Khovanov homology, whereas the slice genus is only known through Khovanov homology relative to the s-invariant. The gauge theoretic construction of Khovanov homology in [122] combined with the knight move conjecture makes it clear that (in our dataset) s may be extracted from the Hilbert space of a five- or six-dimensional theory with surface operators, whereas the construction of $s^\sharp$ in [138] shows that invariants like s can arise in the instanton homology of simpler gauge theories.

Despite these relationships, there is no obvious route by which the Jones polynomial may be manipulated into revealing the s-invariant. A potential avenue, motivated by the index formula for $J(q)$ in the five-dimensional theory, is that the analytic structure of $J(q)$ around $q = 1$ is sensitive to s because s can be extracted from the graded trace with $t = -q^{-4}$ at large k. However, this requires $J(q)$ to somehow be sensitive to analytic structure in the t variable of the Khovanov polynomial. It is not clear why this would be so, as $J(q)$ is obtained from $\mathrm{Kh}(q,t)$ by immediately taking $t = -1$. These results could imply that Chern–Simons gauge theory knows a little bit more about the five-dimensional gauge theory Hilbert space than is naïvely expected. In a previous situation where Chern–Simons theory appeared to contain more information than it should [74], the explanation was found

[29] A major obstruction to answering this question was that the vast majority of our dataset obeyed $|s| = 2g$. A useful step forward would be to find a knot generation procedure which yields many knots for which $|s| < 2g$.

in an analytic continuation of the path integral [2]. Perhaps the same effect is at work here. Another approach which may be useful begins with the observation that there are spectral sequences relating Khovanov homology to *e.g.*, instanton Floer homology [151; 152]. The gauge theory interpretation of these spectral sequences is unclear, but finding such an interpretation may help in relating $J(q)$ to s, since the s-like invariant $s^\sharp$ is easily extracted from instanton Floer homology. The question concerning the four-dimensional information in $J(q)$ that we posed in the introduction still stands, and a complementary question is: what is three-dimensional about the Rasmussen s-invariant?

The specialization of the Khovanov polynomial at $t = -q^{-2}$, on the other hand, had at least a partial explanation via the knight move conjecture for knots in our dataset. This explanation was given in terms of a rearrangement of the Khovanov homology groups into a number of rows determined by the homological width of the knot. This width determined the number of nonzero terms (in the most general case with no cancellations), and the powers of these terms had an offset which depended on s. However, we were not able to explain why the coefficients of the resulting polynomial also seem to be determined completely by s. This suggests that the s-invariant may be extracted from Khovanov homology in a way that differs from the route through Lee homology taken by Rasmussen.

The fact that machine learning is successful at identifying relationships between knot invariants is suggestive, both with respect to low dimensional topology and to quantum field theory. Because knot invariants have been extensively tabulated [86; 145], one could easily imagine using neural networks to scan through all pairs of invariants and determine which can be predicted from knowledge of the other. As we do not have a minimal list of topological invariants that identifies a knot uniquely, such an investigation may provide some guidance in this endeavor. We have also seen that knot invariants have a physical interpretation in diverse dimensions. The Jones polynomial, for example, can be thought of as a mathematical object with a two-dimensional definition or as a physical quantity associated to quantum field theories in three, four, five, or six dimensions. Translating the relationships between knot invariants to a physical language may expose non-obvious connections between quantum field theories.

For instance, in the five-dimensional theory discussed in Section 3.3, the formula we presented for s holds for a certain subset of knots. What this formula means is that there exists a very special pair of approximate supersymmetric ground states in the five-dimensional theory which have instanton numbers equal to $s \pm 1$. These

states survive the five-dimensional instanton corrections to the ground state energies. Our results suggest that the Chern–Simons path integral has privileged access to these states in particular, through some yet undiscovered gauge theory mechanism. Furthermore, the information contained in Lee homology and the specialization $\mathrm{Kh}(q,-q^{-2})$ about the s-invariant may be interpreted as a statement about the existence of deformations of the topological supercharge Q. The nontrivial prediction from Lee's theorem in this context would be that the cohomology of a deformed supercharge is two-dimensional for any knot, and our results on the unusual specialization $\mathrm{Kh}(q,-q^{-2})$ suggest that correlating the fermion number grading with the instanton number grading in this way produces, *e.g.*, a number of ground states with instanton numbers that depend on s in the manner we described in Section 3.4. Perhaps the perspective on deformations of supercharges via spectral sequences presented in [153] would be useful here.[30]

Finally, as poetically and playfully contemplated by the title of [125], machine learning may fill an important niche in constructing potential counterexamples to SPC4. A counterexample establishes the existence of exotic smooth structures on manifolds that are homeomorphic to S^4. If we can conclude whether a knot is slice from a description of the knot (*e.g.*, by using a picture or the braid representation or a collection of other topological invariants), this may highlight interesting knots to examine using analytical methods. Since we are only interested in distinguishing $g=0$ from $g\neq 0$, this is a conceptually simpler task than learning the slice genus exactly as we have aimed to do from the Khovanov and Jones polynomials. Applying a neural network trained to identify sliceness on these knots is a promising attack for the future.

[30]However, it is difficult to relate the two-dimensional perspective of [153] with the five-dimensional picture of [122]. In particular, while the two-dimensional construction seems to suggest the explicit symmetry breaking pattern $SU(2)\to U(1)\times U(1)$ (implemented by modifying the superpotential) is important to obtain Lee homology, this may not be the relevant mechanism in five dimensions to implement Lee's deformation. We thank Edward Witten for pointing this out to us.

Appendix

3.A Implementation

In this appendix, we supply the key part of our implementation. Listing 3.1 outlines the building and training of the neural network.

Listing 3.1: Keras implementation

```
import numpy as np
import keras
from sklearn.model_selection import train_test_split

train_size = 0.25  # training fraction
epochs = 50  # number of epochs to train for
density = 100  # density of hidden layers in network
activation = 'relu'  # activation function for input/hidden layers
loss = 'sparse_categorical_crossentropy'  # loss function for multi-label classification
optim = 'adam'  # optimizer


#  parameters: full set of inputs and outputs
#  returns: trained model and testing data (input, output)
def learn(inputs, outputs):
    # fix input dimensions
    if np.isscalar(inputs[0]):
        dims = 1
    else:
        dims = len(train_in[0])

    # calculate number of output classes
    num_classes = max(outputs) - min(outputs) + 1

    # split into training/testing data
    train_in, test_in, train_out, test_out = train_test_split(inputs, outputs,
                                                              train_size=train_size)

    # build the model
```

```
30      model = keras.models.Sequential([
31          keras.layers.Dense(density, activation=activation,
    input_dims=dims),
32          keras.layers.Dense(density, activation=activation),
33          keras.layers.Dense(num_classes, activation='softmax'
34          )]
35
36      # compile and train the model
37      model.compile(optimizer=optim, loss=loss, metrics=['accuracy']
38      model.fit(train_in, train_out, epochs=epochs, verbose=0.5)
39
40      return model, test_in, test_out
```

Chapter 4

Learning holographic quantum chaos

We explore an artificial neural network representation of many-body wave functions to study out of time ordered correlators (OTOCs) in the SYK model. We provide a simplest working example, computing an OTOC for the $N = 4$ SYK model, to demonstrate the effectiveness of this technique. We also provide a discussion of the numeric implementation, including how to construct and train the network, as well as how to act with the SYK Hamiltonian on the neural network states. The techniques presented have applications in studying quantum chaos.

4.1 Introduction

As discussed in Section 1.4, the SYK model is a 0+1 dimensional model of Majorana fermions with all-to-all couplings, given by the Hamiltonian

$$H = \sum_{1 \leq i_1 < i_2 \ldots < i_q \leq N} J_{i_1 \ldots i_q} \psi_{i_1} \ldots \psi_{i_q} \,. \tag{4.1.1}$$

The SYK model saturates the universal chaos bound given by $\lambda_L \leq 2\pi T$, where λ_L is the Lyapunov exponent. The Lyapunov exponent characterizes the early time behavior of the out of time ordered correlator (OTOC), $F(t) = \langle W^\dagger(t) V^\dagger W(t) V \rangle$: $F(t) \sim e^{\lambda_L t}$. Studying the chaotic behavior of the SYK model can provide insight into the physics of black holes, as there is a correspondence between the SYK model and the theory of near extremal black holes.

Numerical studies of OTOCs in large-N systems are challenging as the task of diagonalizing $2N \times 2N$ Hamiltonian matrices quickly becomes impossible. Recent studies [154] have made great progress in this area and the authors are able to compute OTOCs for system sizes up to $N = 60$ in under three days. Neural network quantum states (NNQS) [10] present another possible approach to these problems.

Instead of the large Hamiltonian matrix, we use a RBM representation for many-body wavefunctions to compute quantities of interest.

In this chapter, we present an approach to computing OTOCs for the SYK model using the NNQS framework. The layout of this chapter is as follows. In Section 4.2, we discuss motivations for studying the SYK model and recent interesting developments in the area. In Section 4.3, we introduce the process of computing OTOCs using the RBM representations. In Section 4.4, we present some preliminary numerical results. The two appendices, Appendix 4.A and Appendix 4.B, give further details on the numerical implementation and include some key algorithms for implementing the SYK model for the RBM representations.

4.2 Motivation

A primary motivation for this work is to analyze the effectiveness of the RBM representation in studying quantum chaos. As a first step, we aim to compute OTOCs and extract their early-time decay parameters λ_{RBM} by fitting the OTOCs to exponential functions. Comparing λ_{RBM} to the chaos bound, $2\pi T$, will determine whether the RBM approach is useful for studying these chaotic quantum systems. This is explored in Section 4.4. In addition to confirming that the RBM method is effective for these computations, there are a number of interesting recent developments involving the SYK model which may be amenable to the RBM approach.

It was recently proposed in [155] that sparse versions of the SYK model still demonstrate the chaotic behavior exhibited by the full model. The would be interesting to study using the RBM approach. In [156], the author proposes a reformulation of theories of fast scrambling and of the holographic complexity correspondence so that they apply to de Sitter, rather than anti-de Sitter, space. It is claimed that the SYK model satisfies some criteria that are needed to describe de Sitter space, particularly that the complexity growth be "hyperfast". This claim would be interesting to test using the RBM approach.

4.3 Numeric background

As introduced in [10], the wavefunction of a many-body quantum system can be represented by a Restricted Boltzmann Machine (RBM). Suppose we have a system with N spins; $\mathcal{S} = (S_1, S_2, \ldots, S_N)$. The RBM representation consists of an artificial neural network with two layers: the visible spins $\mathcal{S}$ and the hidden spins $\mathcal{H} =$

$(h_1, \ldots, h_M)$. There are two bias vectors, $\mathbf{a} = (a_1, \ldots, a_N)$ and $\mathbf{b} = (b_1, \ldots, b_M)$, as well as an $N \times M$ weight matrix, W. The state can be expressed as

$$\Phi(S) = \sum_{\{h_i\}} \exp\left(\sum_j a_j s_j + \sum_i b_i h_i + \sum_{i,j} W_{ij} h_i s_j \right) . \tag{4.3.1}$$

Tracing over the hidden spins, we get

$$\Phi(S) = \exp\Big(\sum_j a_j s_j \Big) \prod_i 2\cosh\Big(b_i + \sum_j W_{ij} s_j \Big) . \tag{4.3.2}$$

The RBM is trained using the Stochastic Reconfiguration (SR) algorithm, which involves a Monte Carlo sampling procedure (see Appendix 4.A for details). The loss function is the expectation value of the energy; $E(\mathcal{W}) = \langle \Phi | \mathcal{H} | \Phi \rangle / \langle \Phi | \Phi \rangle$. Training the RBM with the SR algorithm finds the set of parameters that best represents the ground state of the system.

Consider an out of time ordered correlator (OTOC),

$$F(t) = \langle W^\dagger(t) V^\dagger W(t) V \rangle. \tag{4.3.3}$$

We will now describe a general procedure to compute the OTOC in the RBM representation, as developed in [157]. First, prepare a random state $|\Psi_0\rangle$ (this involves generating $N + M + N \times M$ random complex numbers. In this work, we choose these numbers according to the normal distribution $\mathcal{N}(0, \sigma^2)$, where $\sigma \ll 1$.) The OTOC can be computed as an overlap between two states $|\psi_1\rangle = W(t)V |\Psi_0\rangle = U^\dagger W U V |\Psi_0\rangle$, and $|\Psi_2\rangle = VW(t) |\Psi_0\rangle = VU^\dagger W U |\Psi_0\rangle$. The time evolution can be achieved via the SR algorithm, which can be thought of as an imaginary time evolution. In order to achieve real time evolution, we apply an imaginary learning rate; for instance $\mathrm{d}t = 0.001i$. In this work, the operators V and W are the SYK fields ψ_i (Appendix 4.B details how to apply these operators to the RBM states.)

If we further want to compute OTOCs at inverse temperature, β, then we should rather compute

$$F_\beta(t) = \langle W^\dagger(t) V^\dagger(\beta) W(t) V(\beta) \rangle , \tag{4.3.4}$$

where $V(\beta) = e^{-\beta/4} V e^{-\beta/4}$. The effect of the $e^{-\beta/4}$ is an imaginary time evolution, which can also be achieved via the SR algorithm.

The last idea we need in order to compute OTOCs is how to calculate the overlap between two RBM states. This can also be achieved via a Monte Carlo sampling process (see Appendix 4.A for details on how this is achieved). First, we sample spins according to $|\Phi_1\rangle$ and compute the quantity $v_1 = \sum_S \frac{\langle \Phi_1 | S \rangle \langle S | \Phi_2 \rangle}{\langle \Phi_1 | \Phi_1 \rangle}$. Assuming we sample enough spin configurations, this is equivalent to $\frac{\langle \Phi_1 | \Phi_2 \rangle}{\langle \Phi_1 | \Phi_1 \rangle}$. Similarly, by

sampling according to Φ_2, we compute $v_2 = \frac{\langle \Phi_2 | \Phi_1 \rangle}{\langle \Phi_2 | \Phi_2 \rangle}$. Then, the overlap between the states is given simply by $\sqrt{v_1^* v_2}$.

4.4 Experiments and results

The simplest possible SYK Hamiltonian is given by the case $N = 4$, $q = 4$; $H = J_{1234}\psi_1\psi_2\psi_3\psi_4$. We compute the OTOC using the RBM method and compare to the exact diagonalization results. We also fix $J = 1$, $\beta = 1$. The results for this experiment are shown in Figure 4.4.1.

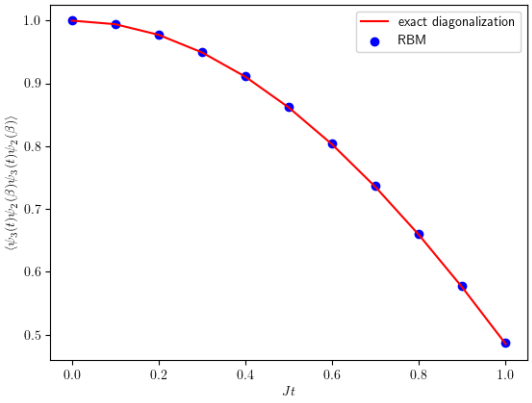


Figure 4.4.1: Simplest working example of computing OTOCs using the RBM representation. The OTOC is computed for $N = 4$ and $q = 4$, meaning the Hamiltonian has a single term. The results agree nearly identically with the exact diagonalization.

We see that in the $N = 4$ case, the RBM method precisely reproduces the exact diagonalization results. This serves as a simplest working example that demonstrates the effectiveness of the RBM method. We leave careful analysis, including studying larger N and the computation of Lyapunov exponents, to later work. The eventual goal of this research is to consider different Hamiltonians and numerically map out the phase space of chaos.

Appendix

4.A Stochastic Reconfiguration

In [10], the authors found a representation of quantum many-body wave functions in terms of RBMs (and more general artificial neural networks). To find the best representation of the ground state of the system, the Stochastic Reconfiguration (SR) algorithm is utilized. The first step is to generate a Markov chain of spin configurations; $\mathcal{S}^{(1)} \to \mathcal{S}^{(2)} \to \dots \mathcal{S}^{(p)}$. The Markov chain is generated through the Metropolis-Hastings algorithm [158], with acceptance probability

$$A(\mathcal{S}^{(k)} \to \mathcal{S}^{(k+1)}) = \min\left(1, \left|\frac{\Phi(\mathcal{S}^{(k+1)})}{\Phi(\mathcal{S}^{(k)})}\right|^2\right) \tag{4.A.1}$$

We then compute the local energy,

$$E_{\text{loc}}(\mathcal{S}) = \frac{\langle \mathcal{S}|\mathcal{H}|\phi\rangle}{\phi(\mathcal{S})} \tag{4.A.2}$$

and the variational derivatives

$$\mathcal{O}_k(\mathcal{S}) = \frac{1}{\Phi(\mathcal{S})}\partial_{\mathcal{W}_k}\Phi(\mathcal{S}). \tag{4.A.3}$$

Recalling the definition of Φ in (4.3.2), we see that

$$\frac{1}{\Phi(\mathcal{S})}\partial_{a_i}\Phi(\mathcal{S}) = \sigma_i^z\,, \tag{4.A.4a}$$

$$\frac{1}{\Phi(\mathcal{S})}\partial_{b_j}\Phi(\mathcal{S}) = \tanh(\theta_j(\mathcal{S}))\,, \tag{4.A.4b}$$

$$\frac{1}{\Phi(\mathcal{S})}\partial_{W_{ij}}\Phi(\mathcal{S}) = \sigma_i^z \tanh(\theta_j(\mathcal{S}))\,. \tag{4.A.4c}$$

Using (4.A.2) and (4.A.3), we compute

$$S_{kk'} = \langle \mathcal{O}_k^* \mathcal{O}_{k'}\rangle - \langle \mathcal{O}_k^*\rangle\langle \mathcal{O}_{k'}\rangle \tag{4.A.5}$$

and

$$F_k(p) = \langle E_{\text{loc}}\mathcal{O}_k^*\rangle - \langle E_{\text{loc}}\rangle\langle \mathcal{O}_k^*\rangle. \tag{4.A.6}$$

The RBM parameters are then updated according to

$$\mathcal{W}(p+1) = \mathcal{W}(p) - \mathrm{d}tS^{-1}(p)F(p). \tag{4.A.7}$$

4.B SYK Hamiltonian

The SYK Hamiltonian is given by (4.1.1). The basic operators implemented in the code are the Pauli spin operators, σ_i^α, with $i = 1, 2, 3$ and $\alpha = 1, \ldots, N$. The ψ_is in (4.1.1) are built up from these matrices:

$$\psi_i^{(N)} = \psi_i^{(N-2)} \otimes \sigma_3, \tag{4.B.1a}$$

$$\psi_{N-1}^{(N)} = \mathbb{I} \otimes \sigma_2, \tag{4.B.1b}$$

$$\psi_N^{(N)} = \mathbb{I} \otimes \sigma_1. \tag{4.B.1c}$$

The action of the Pauli matrices on the RBM state can be understood as follows (see [157]). σ_1^i flips the sign of element i in the visible bias vector and of row i in the weight matrix: $a_i \to -a_i$, $W_{ij} \to -W_{ij}$ for all $j \in [1, M]$. σ_3^i adds a phase to the visible bias vector, $a_i \to a_i + \frac{i\pi}{2}$. Finally, σ_2^i can be implemented as $\sigma_3^i \sigma_1^i$.

Now we need to determine the action of a ψ_i on the RBM state. This is given by Algorithm 1. As an example, consider the action of $\psi_4^1 0$ on the RBM state. p is given by $-(\lfloor \frac{10-4}{2} \rfloor + 1) = -4$. So, $\mathbf{a}[6] = -\mathbf{a}[6]$. Also, we apply a sign flip to the seventh row of W. Since $i = 4$ is even, we do not need to apply line 5. Finally, we add $\frac{i\pi}{2}$ to all elements of a that come after position 4. The result of these actions is that we have applied σ_x^4 and σ_z^j, where $j = 8, 9, 10$. Let's compare this to (4.B.1a):

$$\begin{aligned}
\psi_4^{10} &= \psi_4^8 \otimes \sigma_3 \\
&= \psi_4^6 \otimes \sigma_3 \otimes \sigma_3 \\
&= \psi_4^4 \otimes \sigma_3 \otimes \sigma_3 \otimes \sigma_3 \\
&= \mathbb{I} \otimes \sigma_1 \otimes \sigma_3 \otimes \sigma_3 \otimes \sigma_3.
\end{aligned}$$

This is equivalent to the result of Algorithm 1.

We use a similar technique to implement the Hamiltonian, 4.1.1. Algorithm 2 details how to compute a single term of the Hamilton. First, we apply each of the q operators to the RBM state. Then, we compute the local energy: $E_{\text{loc}}(\mathcal{S}) = \frac{\langle \mathcal{S}|\mathcal{H}|\phi\rangle}{\phi(\mathcal{S})}$.

Algorithm 1 $\psi(i)$; acts with ψ_i on the RBM

$p \leftarrow -(\lfloor \frac{N-i}{2} \rfloor + 1)$
$a[p] \leftarrow -a[p]$
$W[p] \leftarrow -W[p]$
if $i \mod 2 \neq 0$ **then**
    $a[p] \mathrel{+}= \frac{i\pi}{2}$
end if
$p \leftarrow p + 1$
$a[p:] \mathrel{+}= \frac{i\pi}{2}$

Algorithm 2 $H_t(J_t, a, b, w, s, t)$ Computes a term of the SYK Hamiltonian

$a_n \leftarrow a$
$w_n \leftarrow w$
$E \leftarrow 0$
if $J_t \neq 0$ **then**
    $i \leftarrow 0$
    while $i \leq q + 1$ **do**
        $n \leftarrow t[q - i]$
        $p \leftarrow -\lfloor (N - n)/2 + 1 \rfloor$
        $a_n[p] \leftarrow -a_n[p]$
        $w_n[p, :] \leftarrow -w_n[p, :]$
        if $\mod n, 2 \neq 0$ **then**
            $a_n[p] \leftarrow a_n[p] + 1j\pi/2$
        end if
        $p \leftarrow p + 1$
        $a_n[p:] \leftarrow a_n[p:] + 1j * \pi/2$
        $i \leftarrow i + 1$
    end while
    logdiff $\leftarrow$ matmul$(a_n - a, s)$
    $\theta_o \leftarrow b +$ matmul(w^T, s)
    $\theta_n \leftarrow b +$ matmul(w_n^T, s)
    $i \leftarrow 0$
    while $i < M$ **do**
        logdiff $\leftarrow$ logdiff $+ \log(\cosh(\theta_n[i])) - \log(\cosh(\theta_o[j]))$
        $i \leftarrow i + 1$
    end while
    $E \leftarrow J_t * \exp($logdiff$)$
end if

Chapter 5

Conclusion

In this work, we have demonstrated some applications of machine learning techniques to theoretical physics. In Chapter 2, we used a technique called Layer-wise Relevance Propagation to reverse engineer a neural network and derive a phenomenological formula that explained the machine learning results. In Chapter 3, we used neural networks to identify new relationships between knot invariants and to explore the nature of these relationships. In Chapter 4, we use artificial neural network structures to study chaos many-body quantum systems.

In both Chapters 2 (preceded by [74]) and 3, artificial neural networks are used to uncover new relationships between knot invariants. These works demonstrate an application for machine learning that humans will hardly ever be able to replicate: the ability to pick out relationships in large datasets. For instance, there was no mathematical or physical reason to expect that the Jones polynomial can predict the Rasmussen s-invariant with over 95% accuracy (Section 3.4.4). The available datasets of knot invariants contain dozens of invariants and millions of knots far more raw data than any scientist would successfully be able to parse. However, by automating a neural network to search for correlations between the knot invariants, we may be able to fast-track discovery by quickly highlighting previously unknown relationships.

While discovering new relationships is certainly interesting, it is only the start of the story. These new relationships open the door to new lines of investigation, which may be pursued with or without machine learning. For instance, the application of interpretable machine learning techniques in Chapter 2 led to a physical interpretation of machine learning results in analytically continued Chern–Simons theory. These findings may open the door to further analytic discoveries, as the work in Chapter 3 suggests the possibility of a monotonic version of the volume conjecture [69; 70]. To explore this direction further, a tabulation of a significant number of colored Jones polynomials with $n \geq 3$ is required. Once this data is available, it

would be simple to repeat Chapter 2's experiments with the higher color polynomials. The findings in Chapter 3 (the relationships between $\mathrm{Kh}(q, -q^{-2})/J(q)$ and s/g) so far lack a full analytic explanation. While LRP techniques were ineffective in these experiments, perhaps successful application of other interpretable machine learning techniques could clear the way to a gauge-theoretic explanation of the relationship between $J(q)$ and s, just as they did for the relationship between $J(q)$ and V.

Machine learning architectures can also be used as tools to perform computationally difficult tasks. Exactly diagonalizing many-body Hamiltonians for system sizes larger than $N = 50$ is still a hopeless task. However, the RBM representation of quantum states, [10], has provided a novel way to approach these sorts of problems. In Chapter 4, we demonstrate a possible application of the RBM quantum states to the study of the SYK model and consequently to quantum chaos. The obvious future applications of this work include exploring larger system sizes and studying novel systems, including sparse SYK models [155] and SYK models which demonstrate hyperfast chaos [156].

The field of data science in string theory is expanding rapidly, so it is difficult to make exact predictions on the future of the field. However, there are a few interesting directions that are currently capturing the focus of the string theory/ML community. The first is further study of the largest and most well-known dataset in string theory; the Kreuzer–Skarke collection of Calabi–Yau threefolds. As discussed in Section 1.1.3, CYs have been explored using a variety of machine learning methods. Some recent works ([159; 160; 161], among others) have demonstrated continued interest in this area. Another interesting area is the use of machine learning structures to represent physical systems/structures. One example of this, neural network quantum states, is introduced in Chapter 4. Other instances include training neural networks to represent bulk geometries in the AdS/CFT correspondence [35; 36; 37; 162] and to define quantum field theories [163].

It is certainly attractive to try to address as many problems as possible with a simple feedforward neural network, given past successes and the relative simplicity of setting up the neural network. However, neural networks are not always the best tool to tackle a problem. There exists a vast landscape of machine learning architectures, designed for a large variety of tasks, which may have many useful physics applications. For instance, a class of machine learning methods focus on finding efficient representations of data. These include autoencoders (which are themselves a type of neural network), which learn encodings of unlabeled data. Dimensional reduction techniques, such as Principal Component Analysis (PCA) and t-SNE (which

we applied in Chapter 2) are used to find a lower-dimensional representation of high-dimensional inputs. These techniques are extremely useful for identifying clusters and patterns in data. Finding efficient, lower-dimensional representations of data makes it easier for people to study the data themselves, without further aid of machines. Generative models, such as generative adversarial networks, can be used to generate new data that follows the same statistics as the old data. These architectures could be useful in expanding our datasets, giving us more data to conduct further experiments on.

A major criticism of the use of data science in physics is that machine learning does not provide acceptable proofs or evidence of its results. To ensure that machine learning is a respected tool in physics research, we must address the problems of reliability and determine exactly what role we expect machine learning to play in our research. One way to tackle these concerns is by embracing explainable machine learning techniques. Paying attention to precisely how the machine is learning can help us test whether its results are useful. Of further concern is the correctness and availability of the datasets used in physics/ML research. In order for machine learning results to be trusted, it is necessary that we have universally available and well-maintained datasets. Another problem with the datasets we use is that they may not be representative of the physics/mathematics problem that we are trying to tackle. For example, there are many relationships that exist between knot invariants with low crossing number which generally do not hold true for higher crossing numbers. We must therefore ensure that we either work with representative datasets, or that we limit our conclusions to the scope of the data that we are working with.

The study of string theory via machine learning can be counted as no more than five years old, so there are presumably many more discoveries to be made and many more lessons to be learned. It is unknown how much we may expect of our computers in the future. Turing himself believed the question "can machines think?" to be a trivial one [164] and that all that was required to achieve true intelligence was sufficient memory space. However, it seems unlikely (even with the rapid progression of computing technology) that the work of a physicist will ever be automated. Machines are incredibly powerful tools and new technology allows us to harness them against dense, time-consuming problems; ultimately freeing us from mundane tasks and making room for the uniquely human creativity and ingenuity that are required to expand our understanding of the universe.

Bibliography

[1] D. Bar-Natan, *The symbol font* `dbnsymb`, *Online* (2020) .

[2] J. Craven, V. Jejjala and A. Kar, *Disentangling a Deep Learned Volume Formula*, *JHEP* **06** (2021) 040, [2012.03955].

[3] J. Craven, M. Hughes, V. Jejjala and A. Kar, *Learning knot invariants across dimensions*, *arXiv preprint arXiv:2112.00016* (2021) .

[4] L. McClendon and N. Meghanathan, *Using machine learning algorithms to analyze crime data*, *Machine Learning and Applications: An International Journal (MLAIJ)* **2** (2015) 1–12.

[5] M. Fatima, M. Pasha et al., *Survey of machine learning algorithms for disease diagnostic*, *Journal of Intelligent Learning Systems and Applications* **9** (2017) 1.

[6] D. W. Otter, J. R. Medina and J. K. Kalita, *A survey of the usages of deep learning for natural language processing*, *IEEE Transactions on Neural Networks and Learning Systems* **32** (2020) 604–624.

[7] D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez et al., *Mastering chess and shogi by self-play with a general reinforcement learning algorithm*, *arXiv preprint arXiv:1712.01815* (2017) .

[8] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche et al., *Mastering the game of go with deep neural networks and tree search*, *nature* **529** (2016) 484–489.

[9] S.-M. Udrescu and M. Tegmark, *Ai feynman: A physics-inspired method for symbolic regression*, *Science Advances* **6** (2020) eaay2631.

[10] G. Carleo and M. Troyer, *Solving the quantum many-body problem with artificial neural networks*, *Science* **355** (2017) 602–606.

[11] G. E. Hinton, *A practical guide to training restricted boltzmann machines*, in *Neural networks: Tricks of the trade*, pp. 599–619. Springer, 2012.

[12] F. Ruehle, *Data science applications to string theory*, *Phys. Rept.* **839** (2020) 1–117.

[13] Y.-H. He, *Deep-Learning the Landscape*, 1706.02714.

[14] D. Krefl and R.-K. Seong, *Machine Learning of Calabi-Yau Volumes*, *Phys. Rev. D* **96** (2017) 066014, [1706.03346].

[15] F. Ruehle, *Evolving neural networks with genetic algorithms to study the string landscape*, *Journal of High Energy Physics* **2017** (2017) 1–20.

[16] J. Carifio, J. Halverson, D. Krioukov and B. D. Nelson, *Machine Learning in the String Landscape*, *JHEP* **09** (2017) 157, [1707.00655].

[17] K. Bull, Y.-H. He, V. Jejjala and C. Mishra, *Machine Learning CICY Threefolds*, *Phys. Lett. B* **785** (2018) 65–72, [1806.03121].

[18] K. Bull, Y.-H. He, V. Jejjala and C. Mishra, *Getting CICY High*, *Phys. Lett. B* **795** (2019) 700–706, [1903.03113].

[19] Y.-H. He and S.-J. Lee, *Distinguishing elliptic fibrations with AI*, *Phys. Lett. B* **798** (2019) 134889, [1904.08530].

[20] H. Erbin and R. Finotello, *Inception neural network for complete intersection calabi–yau 3-folds*, *Machine Learning: Science and Technology* **2** (2021) 02LT03.

[21] H. Erbin and R. Finotello, *Machine learning for complete intersection calabi-yau manifolds: a methodological study*, *Physical Review D* **103** (2021) 126014.

[22] Y.-H. He and A. Lukas, *Machine learning calabi-yau four-folds*, *Physics Letters B* **815** (2021) 136139.

[23] H. Erbin, R. Finotello, R. Schneider and M. Tamaazousti, *Deep multi-task mining calabi-yau four-folds*, *Machine Learning: Science and Technology* (2021) .

[24] A. Ashmore, Y.-H. He and B. A. Ovrut, *Machine learning calabi–yau metrics*, *Fortschritte der Physik* **68** (2020) 2000068.

[25] L. B. Anderson, M. Gerdes, J. Gray, S. Krippendorf, N. Raghuram and F. Ruehle, *Moduli-dependent calabi-yau and su (3)-structure metrics from machine learning*, *Journal of High Energy Physics* **2021** (2021) 1–45.

[26] M. R. Douglas, S. Lakshminarasimhan and Y. Qi, *Numerical calabi-yau metrics from holomorphic networks*, *arXiv preprint arXiv:2012.04797* (2020) .

[27] V. Jejjala, D. K. M. Pena and C. Mishra, *Neural network approximations for calabi-yau metrics*, *arXiv preprint arXiv:2012.15821* (2020) .

[28] R. Deen, Y.-H. He, S.-J. Lee and A. Lukas, *Machine learning string standard models*, *arXiv preprint arXiv:2003.13339* (2020) .

[29] R. Altman, J. Carifio, J. Halverson and B. D. Nelson, *Estimating calabi-yau hypersurface and triangulation counts with equation learners*, *Journal of High Energy Physics* **2019** (2019) 1–29.

[30] D. Klaewer and L. Schlechter, *Machine Learning Line Bundle Cohomologies of Hypersurfaces in Toric Varieties*, *Phys. Lett. B* **789** (2019) 438–443, [1809.02547].

[31] Y.-N. Wang and Z. Zhang, *Learning non-higgsable gauge groups in 4d f-theory*, *Journal of High Energy Physics* **2018** (2018) 1–50.

[32] H. Erbin and S. Krippendorf, *Gans for generating eft models*, *Physics Letters B* **810** (2020) 135798.

[33] A. Mütter, E. Parr and P. K. Vaudrevange, *Deep learning in the heterotic orbifold landscape*, *Nuclear Physics B* **940** (2019) 113–129.

[34] S. Abel, D. G. Cerdeno and S. Robles, *The power of genetic algorithms: what remains of the pmssm?*, *arXiv preprint arXiv:1805.03615* (2018) .

[35] K. Hashimoto, S. Sugishita, A. Tanaka and A. Tomiya, *Deep learning and the ads/cft correspondence*, *Physical Review D* **98** (2018) 046019.

[36] K. Hashimoto, S. Sugishita, A. Tanaka and A. Tomiya, *Deep learning and holographic qcd*, *Physical Review D* **98** (2018) 106014.

[37] K. Hashimoto, *Ads/cft correspondence as a deep boltzmann machine*, *Physical Review D* **99** (2019) 106017.

[38] C. Adams, *The Knot Book*. W.H. Freeman, 1994.

[39] R. G. Scharein and K. S. Booth, *Interactive knot theory with knotplot*, in *Multimedia Tools for Communicating Mathematics*, pp. 277–290. Springer, 2002.

[40] R. J. VF and A. Jones, *A polynomial invariant for knots via von neumann algebras*, *Bull. Am. Math. Soc., New Ser* **12** (1985) 103–111.

[41] W. P. Thurston et al., *Three-dimensional geometry and topology*, vol. 1. Princeton university press, 1997.

[42] J. S. Purcell, *Hyperbolic knot theory*, vol. 209. American Mathematical Soc., 2020.

[43] G. Perelman, *The entropy formula for the ricci flow and its geometric applications*, *arXiv preprint math/0211159* (2002) .

[44] G. Perelman, *Finite extinction time for the solutions to the ricci flow on certain three-manifolds*, *arXiv preprint math/0307245* (2003) .

[45] G. Perelman, *Ricci flow with surgery on three-manifolds*, *arXiv preprint math/0303109* (2003) .

[46] M. H. Freedman et al., *The topology of four-dimensional manifolds*, *Journal of Differential Geometry* **17** (1982) 357–453.

[47] S. K. Donaldson et al., *An application of gauge theory to four-dimensional topology*, *Journal of Differential Geometry* **18** (1983) 279–315.

[48] C. Manolescu and L. Piccirillo, *From zero surgeries to candidates for exotic definite four-manifolds*, *arXiv preprint arXiv:2102.04391* (2021) .

[49] D. Tong, *Lectures on the quantum hall effect*, *Online* (2016) .

[50] G. W. Moore, *Introduction to chern–simons theories*, *Online* (2019) .

[51] E. Witten, *Quantum Field Theory and the Jones Polynomial*, *Commun. Math. Phys.* **121** (1989) 351–399.

[52] P. Mifsud, "Knot theory, the jones polynomial, and chern-simons theory." June, 2018.

[53] S. Hawking, *Particle creation by black holes 1975*, *Comm. Math. Phys* **43** 199.

[54] J. Maldacena, *The large-n limit of superconformal field theories and supergravity*, *International journal of theoretical physics* **38** (1999) 1113–1133.

[55] S. S. Gubser, I. R. Klebanov and A. M. Polyakov, *Gauge theory correlators from non-critical string theory*, *Physics Letters B* **428** (1998) 105–114.

[56] E. Witten, *Anti de sitter space and holography*, *arXiv preprint hep-th/9802150* (1998) .

[57] G. Sárosi, *Ads _2 holography and the syk model*, *arXiv preprint arXiv:1711.08482* (2017) .

[58] E. Witten, *An syk-like model without disorder*, *Journal of Physics A: Mathematical and Theoretical* **52** (2019) 474002.

[59] J. Maldacena, S. H. Shenker and D. Stanford, *A bound on chaos*, *Journal of High Energy Physics* **2016** (2016) 1–17.

[60] L. J. Dixon, *Some worldsheet properties of superstring compactifications, on orbifolds and otherwise*, in *Summer Workshop in High-energy Physics and Cosmology*, 10, 1987.

[61] W. Lerche, C. Vafa and N. P. Warner, *Chiral Rings in N=2 Superconformal Theories*, *Nucl. Phys. B* **324** (1989) 427–474.

[62] P. Candelas, M. Lynker and R. Schimmrigk, *Calabi-Yau Manifolds in Weighted P(4)*, *Nucl. Phys. B* **341** (1990) 383–402.

[63] B. R. Greene and M. Plesser, *Duality in {Calabi-Yau} Moduli Space*, *Nucl. Phys. B* **338** (1990) 15–37.

[64] P. Candelas, X. C. De La Ossa, P. S. Green and L. Parkes, *A Pair of Calabi-Yau manifolds as an exactly soluble superconformal theory*, *AMS/IP Stud. Adv. Math.* **9** (1998) 31–95.

[65] A. Givental, *Topological field theory, primitive forms and related topics*, ch. A mirror theorem for toric complete intersections, pp. 141–175. Springer, 1998.

[66] B. Lian, K. Liu and S.-T. Yau, *Mirror principle. 1.*, *Asian J. Math.* **1** (1997) 729–763.

[67] V. F. R. Jones, *Hecke algebra representations of braid groups and link polynomials*, *The Annals of Mathematics* **126** (Sep, 1987) 335.

[68] W. Thurston, *The Geometry and Topology of 3-Manifolds*, *Lecture Notes, Princeton University* (1978) .

[69] R. M. Kashaev, *The hyperbolic volume of knots from quantum dilogarithm*, *Letters in Mathematical Physics* **39** (1997) 269–275, [arXiv:q-alg/9601025].

[70] H. Murakami and J. Murakami, *The colored Jones polynomials and the simplicial volume of a knot*, *Acta Math.* **186** (2001) 85–104, [arXiv:math/9905075].

[71] S. Gukov, *Three-dimensional quantum gravity, Chern-Simons theory, and the A polynomial*, *Commun. Math. Phys.* **255** (2005) 577–627, [hep-th/0306165].

[72] N. Dunfield, *An interesting relationship between the Jones polynomial and hyperbolic volume*, *Online* (2000) .

[73] S. Garoufalidis and Y. Lan, *Experimental evidence for the volume conjecture for the simplest hyperbolic non-2–bridge knot*, *Algebraic & Geometric Topology* **5** (May, 2005) 379–403, [math/0412331].

[74] V. Jejjala, A. Kar and O. Parrikar, *Deep Learning the Hyperbolic Volume of a Knot*, *Phys. Lett. B* **799** (2019) 135033, [1902.05547].

[75] L. G. Valiant, *A theory of the learnable*, *Commun. ACM* **27** (1984) 1134–1142.

[76] E. Witten, *Analytic Continuation Of Chern-Simons Theory*, *AMS/IP Stud. Adv. Math.* **50** (2011) 347–446, [1001.2933].

[77] V. Balasubramanian, J. R. Fliss, R. G. Leigh and O. Parrikar, *Multi-Boundary Entanglement in Chern-Simons Theory and Link Invariants*, *JHEP* **04** (2017) 061, [1611.05460].

[78] V. Balasubramanian, M. DeCross, J. Fliss, A. Kar, R. G. Leigh and O. Parrikar, *Entanglement Entropy and the Colored Jones Polynomial*, *JHEP* **05** (2018) 038, [1801.01131].

[79] T. Dimofte and S. Gukov, *Quantum Field Theory and the Volume Conjecture*, *Contemp. Math.* **541** (2011) 41–67, [1003.4808].

[80] Y. LeCun, Y. Bengio and G. Hinton, *Deep learning*, *Nature* **521** (May, 2015) 436–444.

[81] M. C. Hughes, *A neural network approach to predicting and computing knot invariants*, 1610.05744.

[82] S. Gukov, J. Halverson, F. Ruehle and P. Sułkowski, *Learning to Unknot*, 2010.16263.

[83] G. Cybenko, *Approximation by superpositions of a sigmoidal function*, *Mathematics of Control, Signals and Systems* **2** (Dec, 1989) 303–314.

[84] K. Hornik, *Approximation capabilities of multilayer feedforward networks*, *Neural networks* **4** (1991) 251–257.

[85] J. Hoste, M. Thistlethwaite and J. Weeks, *The first 1,701,936 knots*, *The Mathematical Intelligencer* **20** (1998) 33–48.

[86] "The knot atlas: The take home database." http://katlas.org/wiki/The_Take_Home_Database.

[87] M. Culler, N. M. Dunfield, M. Goerner and J. R. Weeks, "SnapPy, a computer program for studying the geometry and topology of 3-manifolds." http://snappy.computop.org.

[88] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller and W. Samek, *On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation*, *PloS one* **10** (2015) e0130140.

[89] G. Montavon, A. Binder, S. Lapuschkin, W. Samek and K.-R. Müller, *Layer-wise relevance propagation: an overview*, in *Explainable AI: interpreting, explaining and visualizing deep learning*, pp. 193–209. Springer, 2019.

[90] O. Dasbach and X.-S. Lin, *A volumish theorem for the jones polynomial of alternating knots*, *Pacific Journal of Mathematics* **231** (Jun, 2007) 279–291.

[91] S.-M. Udrescu and M. Tegmark, *AI Feynman: a Physics-Inspired Method for Symbolic Regression*, 1905.11481.

[92] S. Blücher, L. Kades, J. M. Pawlowski, N. Strodthoff and J. M. Urban, *Towards novel insights in lattice field theory with explainable machine learning*, *Phys. Rev. D* **101** (2020) 094507, [2003.01504].

[93] J. S. Cotler, G. Gur-Ari, M. Hanada, J. Polchinski, P. Saad, S. H. Shenker et al., *Black Holes and Random Matrices*, *JHEP* **05** (2017) 118, [1611.04650].

[94] M. Khovanov, *Patterns in knot cohomology, i*, *Experiment. Math.* **12** (2003) 365–374.

[95] K. Habiro, *On the colored jones polynomials of some simple links*, *Sūrikaisekikenkyūsho Kōkyūroku* (2000) 34–43.

[96] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro et al., *TensorFlow: Large-scale machine learning on heterogeneous systems*, 2015.

[97] Y.-H. He and S.-T. Yau, *Graph Laplacians, Riemannian Manifolds and their Machine-Learning*, 2006.16619.

[98] J. Halverson, B. Nelson and F. Ruehle, *Branes with Brains: Exploring String Vacua with Deep Reinforcement Learning*, *JHEP* **06** (2019) 003, [1903.11616].

[99] J. Halverson, M. Plesser, F. Ruehle and J. Tian, *Kähler Moduli Stabilization and the Propagation of Decidability*, *Phys. Rev. D* **101** (2020) 046010, [1911.07835].

[100] J. Halverson and C. Long, *Statistical Predictions in String Theory and Deep Generative Models*, *Fortsch. Phys.* **68** (2020) 2000005, [2001.00555].

[101] M. Bies, M. Cvetič, R. Donagi, L. Lin, M. Liu and F. Ruehle, *Machine Learning and Algebraic Approaches towards Complete Matter Spectra in 4d F-theory*, 2007.00009.

[102] C. R. Brodie, A. Constantin, R. Deen and A. Lukas, *Machine Learning Line Bundle Cohomology*, *Fortsch. Phys.* **68** (2020) 1900087, [1906.08730].

[103] C. R. Brodie, A. Constantin, R. Deen and A. Lukas, *Topological Formulae for the Zeroth Cohomology of Line Bundles on Surfaces*, 1906.08363.

[104] C. R. Brodie, A. Constantin, R. Deen and A. Lukas, *Index Formulae for Line Bundle Cohomology on Complex Surfaces*, *Fortsch. Phys.* **68** (2020) 1900086, [1906.08769].

[105] M. Khovanov, *A categorification of the jones polynomial*, *Duke Math. J.* **101** (02, 2000) 359–426.

[106] D. Bar-Natan, *On khovanov's categorification of the jones polynomial*, *Algebraic & Geometric Topology* **2** (May, 2002) 337–370, [math/0201043].

[107] L. H. Kauffman, *State models and the jones polynomial*, *Topology* **26** (1987) 395 – 407.

[108] K. P. Walsh, *Patterns and Stability in the Coefficients of the Colored Jones Polynomial*. PhD thesis, UC, San Diego, 2014.

[109] K. P. Walsh, *Higher order stability in the coefficients of the colored jones polynomial*, *Journal of Knot Theory and Its Ramifications* **27** (2018) 1840010, [1603.06957].

[110] T. T. Lê, "The colored jones polynomial and the aj conjecture." http://people.math.gatech.edu/~letu/Papers/Lectures_Luminy_2014_new.pdf.

[111] V. F. Jones, "The jones polynomial for dummies." https://math.berkeley.edu/~vfr/jonesakl.pdf.

[112] V. F. Jones, "The jones polynomial." https://math.berkeley.edu/~vfr/jones.pdf.

[113] H. Murakami, J. Murakami, M. Okamoto, T. Takata and Y. Yokota, *Kashaev's conjecture and the chern-simons invariants of knots and links*, *Experimental Mathematics* **11** (2002) 427–435, [math/0203119].

[114] L. van der Maaten and G. Hinton, *Visualizing data using t-SNE*, *Journal of Machine Learning Research* **9** (2008) 2579–2605.

[115] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel et al., *Scikit-learn: Machine learning in Python*, *Journal of Machine Learning Research* **12** (2011) 2825–2830.

[116] P. Freyd, D. Yetter, J. Hoste, W. B. R. Lickorish, K. Millett and A. Ocneanu, *A new polynomial invariant of knots and links*, *Bull. Am. Math. Soc.* **12** (1985) 239–246.

[117] J. H. Przytycki and P. Traczyk, *Conway algebras and skein equivalence of links*, *Proceedings of the American Mathematical Society* **100** (1987) 744–748.

[118] J. W. Alexander, *Topological invariants of knots and links*, *Transactions of the American Mathematical Society* **30** (1928) 275–306.

[119] M. Cranmer, A. Sanchez-Gonzalez, P. Battaglia, R. Xu, K. Cranmer, D. Spergel et al., *Discovering symbolic models from deep learning with inductive biases*, *NeurIPS 2020* (2020) , [2006.11287].

[120] M. Cranmer, *Pysr: Fast & parallelized symbolic regression in python/julia*, Sept., 2020. 10.5281/zenodo.4052869.

[121] J. Rasmussen, *Khovanov homology and the slice genus*, *Inventiones mathematicae* **182** (2010) 419–447, [math/0402131].

[122] E. Witten, *Fivebranes and Knots*, 1101.3216.

[123] M. Aganagic, *Knot Categorification from Mirror Symmetry, Part I: Coherent Sheaves*, 2004.14518.

[124] M. Aganagic, *Knot Categorification from Mirror Symmetry, Part II: Lagrangians*, 2105.06039.

[125] M. H. Freedman, R. E. Gompf, S. Morrison and K. Walker, *Man and machine thinking about the smooth 4-dimensional poincaré conjecture*, *Quantum Topology* **1** (2010) 171–208, [0906.5177].

[126] A. Davies, P. Velickovic, L. Buesing, S. Blackwell, D. Zheng, N. Tomasev et al., *Advancing mathematics by guiding human intuition with ai*, *Nature* **600** (2021) 70–74.

[127] A. Davies, A. Juhász, M. Lackenby and N. Tomasev, *The signature and cusp geometry of hyperbolic knots*, 2111.15323.

[128] C. Manolescu and M. Marengon, *The knight move conjecture is false*, *Proceedings of the American Mathematical Society* **148** (2020) 435–439, [1809.09769].

[129] E. Witten, *A New Look At The Path Integral Of Quantum Mechanics*, 1009.6032.

[130] D. Gaiotto and E. Witten, *Knot Invariants from Four-Dimensional Gauge Theory*, *Adv. Theor. Math. Phys.* **16** (2012) 935–1086, [1106.4789].

[131] E. S. Lee, *An endomorphism of the Khovanov invariant*, *Adv. Math.* **197** (2005) 554–586, [math/0210213].

[132] E. S. Lee, *The support of the khovanov's invariants for alternating knots*, math/0201105.

[133] C. Manolescu and P. Ozsváth, *On the Khovanov and knot Floer homologies of quasi-alternating links*, in *Proceedings of Gökova Geometry-Topology Conference 2007*, pp. 60–81, Gökova Geometry/Topology Conference (GGT), Gökova, 2008. 0708.3249.

[134] A. Alishahi and N. Dowlin, *The lee spectral sequence, unknotting number, and the knight move conjecture*, *Topology and its Applications* **254** (2019) 29–38, [1710.07875].

[135] M. R. Douglas, *On D=5 super Yang-Mills theory and (2,0) theory*, *JHEP* **02** (2011) 011, [1012.2880].

[136] N. Lambert, C. Papageorgakis and M. Schmidt-Sommerfeld, *M5-Branes, D4-Branes and Quantum 5D super-Yang-Mills*, *JHEP* **01** (2011) 083, [1012.2882].

[137] L. F. Alday, D. Gaiotto and Y. Tachikawa, *Liouville Correlation Functions from Four-dimensional Gauge Theories*, *Lett. Math. Phys.* **91** (2010) 167–197, [0906.3219].

[138] P. B. Kronheimer and T. S. Mrowka, *Gauge theory and rasmussen's invariant*, *Journal of Topology* **6** (Apr, 2013) 659–674, [1110.1297].

[139] S. G. Naculich, H. A. Riggs and H. J. Schnitzer, *Group Level Duality in WZW Models and Chern-Simons Theory*, *Phys. Lett. B* **246** (1990) 417–422.

[140] J. S. F. Levitt, M. Hajij and R. Sazdanovic, *Big data approaches to knot theory: Understanding the structure of the jones polynomial*, 1912.10086.

[141] D. Paweł, D. Gurnari and R. Sazdanovic, *Knot invariants and their relations: a topological perspective*, 2109.00831.

[142] C. Even-Zohar, J. Hass, N. Linial and T. Nowik, *The distribution of knots in the petaluma model*, *Algebraic & geometric topology* **18** (2018) 3647–3667, [1706.06571].

[143] "The Mathematica package KnotTheory." http://katlas.org/wiki/The_Mathematica_Package_KnotTheory.

[144] "KnotJob." https://www.maths.dur.ac.uk/users/dirk.schuetz/knotjob.html.

[145] C. Livingston and A. H. Moore, "Knotinfo: Table of knot invariants." URL: knotinfo.math.indiana.edu, November, 2021.

[146] L. Rudolph, *Quasipositivity as an obstruction to sliceness*, *Bulletin of the American Mathematical Society* **29** (1993) 51–59, [math/9307233].

[147] C. Adams, T. Crawford, B. DeMeo, M. Landry, A. T. Lin, M. Montee et al., *Knot projections with a single multi-crossing*, *Journal of Knot Theory and Its Ramifications* **24** (2015) 1550011, [1208.5742].

[148] J. H. Przytycki and P. Traczyk, *Invariants of links of conway type*, 1610.06679.

[149] J. Howie, *A characterisation of alternating knot exteriors*, *Geometry & Topology* **21** (2017) 2353–2371, [1511.04945].

[150] J. E. Greene, *Alternating links and definite surfaces*, *Duke Mathematical Journal* **166** (2017) 2133–2151, [1511.06329].

[151] P. B. Kronheimer and T. S. Mrowka, *Khovanov homology is an unknot-detector*, *Publications mathématiques de l'IHÉS* **113** (2011) 97–208, [1005.4346].

[152] P. B. Kronheimer and T. S. Mrowka, *Filtrations on instanton homology*, *Quantum Topology* **5** (2014) 61–97, [1110.1290].

[153] S. Gukov, S. Nawata, I. Saberi, M. Stošić and P. Sułkowski, *Sequencing BPS Spectra*, *JHEP* **03** (2016) 004, [1512.07883].

[154] B. Kobrin, Z. Yang, G. D. Kahanamoku-Meyer, C. T. Olund, J. E. Moore, D. Stanford et al., *Many-body chaos in the sachdev-ye-kitaev model*, *Physical Review Letters* **126** (2021) 030602.

[155] S. Xu, L. Susskind, Y. Su and B. Swingle, *A sparse model of quantum holography*, *arXiv preprint arXiv:2008.02303* (2020) .

[156] L. Susskind, *Entanglement and chaos in de sitter holography: An syk example*, *arXiv preprint arXiv:2109.14104* (2021) .

[157] Y. Wu, L.-M. Duan and D.-L. Deng, *Artificial neural network based computation for out-of-time-ordered correlators*, *Physical Review B* **101** (2020) 214308.

[158] W. K. Hastings, *Monte carlo sampling methods using markov chains and their applications*, .

[159] D. S. Berman, Y.-H. He and E. Hirst, *Machine Learning Calabi-Yau Hypersurfaces*, 2112.06350.

[160] N. Afkhami-Jeddi, A. Ashmore and C. Cordova, *Calabi-yau cfts and random matrices*, *arXiv preprint arXiv:2107.11461* (2021) .

[161] P. Berglund, B. Campbell and V. Jejjala, *Machine learning kreuzer–skarke calabi–yau threefolds*, *arXiv preprint arXiv:2112.09117* (2021) .

[162] K. Hashimoto, K. Ohashi and T. Sumimoto, *Deriving dilaton potential in improved holographic qcd from meson spectrum*, *arXiv preprint arXiv:2108.08091* (2021) .

[163] J. Halverson, *Building quantum field theories out of neurons*, *arXiv preprint arXiv:2112.04527* (2021) .

[164] C. Machinery, *Computing machinery and intelligence-am turing*, *Mind* **59** (1950) 433.