Machine learning and algebraic approaches towards complete matter spectra in 4d F-theory

Martin Bies,[a] **Mirjam Cvetič,**[b,c,d] **Ron Donagi,**[c,b] **Ling Lin,**[e] **Muyang Liu**[b] and **Fabian Ruehle**[e,f]

[a]*Mathematical Institute, University of Oxford,*
Woodstock Road, Oxford, OX2 6GG, U.K.

[b]*Department of Physics and Astronomy, University of Pennsylvania,*
209 South 33rd St, Philadelphia, PA 19104-6396, U.S.A.

[c]*Department of Mathematics, University of Pennsylvania,*
209 South 33rd St, Philadelphia, PA 19104-6396, U.S.A.

[d]*Center for Applied Mathematics and Theoretical Physics, University of Maribor,*
Slomškov trg, Center 15, 2000 Maribor, Slovenia

[e]*CERN Theory Department,*
CH-1211 Geneva, Switzerland

[f]*Rudolf Peierls Centre for Theoretical Physics, University of Oxford, Department of Physics,*
Parks Road, Oxford OX1 3PU, U.K.

E-mail: bies@maths.ox.ac.uk, cvetic@physics.upenn.edu,
donagi@math.upenn.edu, lling@physics.upenn.edu, muyang@sas.upenn.edu,
fabian.ruehle@cern.ch

ABSTRACT: Motivated by engineering vector-like (Higgs) pairs in the spectrum of 4d F-theory compactifications, we combine machine learning and algebraic geometry techniques to analyze line bundle cohomologies on families of holomorphic curves. To quantify jumps of these cohomologies, we first generate 1.8 million pairs of line bundles and curves embedded in dP_3, for which we compute the cohomologies. A white-box machine learning approach trained on this data provides intuition for jumps due to curve splittings, which we use to construct additional vector-like Higgs-pairs in an F-Theory toy model. We also find that, in order to explain quantitatively the full dataset, further tools from algebraic geometry, in particular Brill-Noether theory, are required. Using these ingredients, we introduce a diagrammatic way to express cohomology jumps across the parameter space of each family of matter curves, which reflects a stratification of the F-theory complex structure moduli space in terms of the vector-like spectrum. Furthermore, these insights provide an algorithmically efficient way to estimate the possible cohomology dimensions across the entire parameter space.

Contents

1 Introduction

The spectrum of light chiral particles is a defining feature of any four dimensional quantum field theory. Their precise number affects aspects such as the moduli space of vacua, or the behavior of the theory under RG flow. Moreover, they are also of paramount importance to phenomenology, in particular when it comes to models of beyond-the-Standard-Model physics. Therefore, to be able to draw formal and phenomenological lessons from string theory about 4d field theories, one needs efficient methods to compute the spectrum in compactification scenarios.

From an effective field theory perspective, the chiral excess $\chi(\mathbf{R})$ — the difference between chiral and anti-chiral modes of the same matter representation $\mathbf{R}$ — is a discrete parameter, whereas the individual number of light (anti-)chiral modes depend on continuous mass parameters. In string theory, this is reflected by the fact that $\chi(\mathbf{R})$ is typically

a topologically protected quantity, whereas the (perturbative) mass parameters[1] are captured by continuous deformations, or *moduli*, which for certain values can lead to a pair of chiral and anti-chiral modes — a *vector-like pair* — to become massless.

In many string compactification scenarios, we know in principle what the relevant computations are: massless fields are zero modes of some differential operators on the internal space, and therefore counted by appropriate sheaf cohomologies. However, oftentimes these computations are so complicated that in practice, they can only be carried out explicitly for toy models, or for specialized values of the deformation parameters. On the other hand, an exact understanding of how the cohomologies depend on these parameters is necessary for a complete description of the physical interpretation. The moduli dependence and the possibility of jumps in the massless spectrum have been first discussed in the context of heterotic string theory in [1–6]. More recently, the complex structure moduli dependence of the cohomology dimensions has been studied in [7, 8] and [9] in the context of instanton and perturbative superpotential terms, respectively.

In comparison, an analogous analysis in the context of F-theory compactifications [10] is largely missing and has only been discussed in part in [11]. The main reason is because, unlike the chiral spectrum which is accessible via intersection theory [12–25], the vector-like spectrum in F-theory depends on a gauge background, which is encoded in mathematically rather intricate objects such as the intermediate Jacobian and Deligne cohomology [26–29]. Recent progress [30, 31] has made the spectrum computationally more accessible. Namely, for a four-dimensional $\mathcal{N} = 1$ F-theory compactifications on an elliptically fibered Calabi-Yau fourfold $\pi : Y_4 \to B_3$ with a given gauge background, the massless spectrum of chiral particles in representation $\mathbf{R}$ can be counted by certain line bundle cohomologies $h^i(C_{\mathbf{R}}, \mathcal{L}_{\mathbf{R}}), i = 0, 1$ on complex curves $C_{\mathbf{R}} \subset B_3$ — the matter curves — in the base. Given a compact model with a fixed gauge background, $C_{\mathbf{R}}$ and $\mathcal{L}_{\mathbf{R}}$ are specified by global data in terms of polynomials on B_3 , whose coefficients are (parts of) the complex structure parameters of Y_4. In this case, one can model the line bundle as a coherent sheaf on B_3, whose cohomology computation can be systematized in a computer algebra system [32]. While this algorithm can be applied to a broad class of global F-theory models, the calculations for almost all phenomenologically interesting examples overburden even super-computers specifically designed for such tasks. The reason is that here, and in fact in many cohomology computations using commutative algebra or computational algebraic geometry, we need to compute Groebner bases, whose computational complexity scales extremely poorly.

The introduction of ideas from *Big Data* and *machine learning* to string phenomenology [33–36] provides new perspectives; see [37] for an introduction and comprehensive overview. One advantage that a trained algorithm provides is that it recognizes more subtle patterns without the need of a complete, "microscopic" understanding of the task. In particular, recent studies suggest that supervised learning can be used to predict line bundle cohomologies in string compactifications [35, 38, 39]. One may be tempted to apply these techniques, which are mostly motivated by heterotic compactifications, directly to

---

[1]In this work we will neglect moduli stabilization, flux-induced superpotentials and non-perturbative effects.

the F-theory. However, there is a significant difference in the way the line bundle data are specified in global heterotic vs. F-theory models. In heterotic examples, the line bundles are typically given in a "canonical" way, namely as an element of the Picard group $\text{Pic}(X)$ of the underlying manifold X. This was used, e.g., in [40, 41] to derive formulae for line bundle cohomologies in terms of topological indices.

However, in the F-theory setting, there is no straightforward fashion to extract even the structure of the Picard group of $C_\mathbf{R}$, given its polynomial description. Likewise, because the same data specifies $\mathcal{L}_\mathbf{R}$ essentially as a sum of points p_i on B_3 that also lie on $C_\mathbf{R}$, it is by no means obvious if, say, $p_1 - p_2$ is trivial or not on $C_\mathbf{R}$. What makes the situation particularly challenging is that, by varying the complex structure parameters, the structure of $\text{Pic}(C_\mathbf{R})$ as well as the points specifying $\mathcal{L}_\mathbf{R}$ will change. Together with the fact that we simply do not have a large data set of non-trivial F-theory examples, it is a priori unclear whether we could train an algorithm that reliably predicts the cohomologies for realistic models with arbitrary complex parameters.

Instead, we will use machine learning techniques on less complex examples to gain some intuition for circumstances under which line bundle cohomologies *jump*. Physically, this is already interesting as such a jump can engineer one or possibly more massless vector-like pairs in situations where one generically expects none. Even if the trained algorithm does not perform perfectly, understanding its strategy can provide a guiding principle for the behavior of the vector-like spectrum in non-trivial examples. For this reason, we focus on white-box machine learning techniques, in particular on decision trees.

To fully understand the results of the machine learning, we further employ "formal" techniques from algebraic geometry, in the form of *Brill-Noether theory*. This allows to identify "microscopically" the sources for jumps in cohomology, either from the curve $C_\mathbf{R}$ or the line bundle $\mathcal{L}_\mathbf{R}$ becoming non-generic. With these insights, we provide an algorithmic way to estimate the admissible numbers of vector-like pairs over the entire parameter space of a matter curve in a global F-theory model with given gauge background. Furthermore, our analysis also reveals a convenient diagrammatic way to encode the *stratification* on the parameter space induced by the number of vector-like pairs. We believe that this is progress towards understanding the full complex structure dependence of the vector-like spectrum in global F-theory models.

The paper is organized as follows. In section 2 we discuss our machine learning approach. Using the exact methods implemented in [42], we generate a database [43] of cohomologies of pullback line bundles on hypersurface curves in dP_3. Interpreting these results with decision trees, we find that curve splittings typically lead to jumps in the vector-like spectrum. In section 3, we demonstrate that such curve splittings provide a practial way to engineer jumps in a global F-theory GUT-model. To investigate the origin of these jumps, we turn in section 4 to algebraic and analytic techniques. We find a unified perspective on jumps due to curve splittings and non-generic line bundles described by Brill-Noether theory, and introduce a diagrammatic way to illustrate the natural stratification of the complex structure parameter space in terms of the vector-like spectrum. In section 5, we present a refined analysis of jumps due to curve splittings. This rests on a procedure to count the global sections by gluing "local contributions" along intersections

of curve components, which leads to two interesting results: first, we are able to formulate sufficient conditions for jumps of vector-like spectra. Second, we can propose an algorithmic h^0 estimate, which relies mostly on topological data, and hence provides a quick, approximative scan of the vector-like spectrum over the entire parameter space of a matter curve. In contrast to currently existing exact methods, such as [42], our implementation [44] has a much lower demand of computational resources and run times.

2 Machine learning

2.1 Introduction to decision trees

We are interested in tuning complex structure moduli to engineer jumps in the dimensions of sheaf cohomologies over complex curves. It is a priori not clear how to efficiently identify these subloci in complex structure moduli space. In order to state (at least) necessary conditions for jumps to occur, we address the problem using machine learning. Since we are interested in interpreting the results of the machine learning algorithm, we resort to white-box models, in particular to binary decision trees.

In more detail, we use binary decision trees as classifiers in supervised machine learning, following the notation and conventions of [37]. Supervised learning means that we have a set of inputs x_i^μ (called features) together with associated labels[2] y_i, where $i = 1, \ldots, N$ counts the feature-label-pairs, and $\mu = 1, \ldots, F$ counts the F features of each input. This set of feature-label combinations is now divided into a train set and a test set (typically around 90 percent of the pairs are assigned to the train set and 10 percent to the test set). Using the train set, an algorithm is trained to learn a map from the features to the labels. The training consists of adjusting parameters of the algorithm to optimize the map. This is typically done by minimizing the loss, which is a measure for how well the algorithm reproduces the labels. Once training ends, the algorithm is tested on the test set. This is necessary in order to see how well it performs on (hitherto unseen) data. If the test set have been chosen generically enough, performance on the test set will serve as an indicator for how well the trained algorithm will perform.

After this general discussion, let us describe these steps in the context of binary decision trees. Trees are data structures that appear abundantly in computer science. They can be thought of as acyclic, directed, connected graphs with a unique root vertex (in trees, vertices are called nodes). In binary trees, each node has either zero or exactly two vertices, each of which is connected to a unique node. These two subnodes are called child nodes, and the original node is called parent node. A node with no children is called a leaf node.

A decision tree expects numerical features $x_i^{(0)}$. It then introduces boolean splitting criteria of the type $x_i^{(0)} \leq \kappa_i$ for some constant $\kappa_i \in \mathbb{R}$. All data that satisfy this criterion are assigned to one child node, while data that does not satisfy the criterion is assigned to the other child node. The tree is now built recursively by splitting each child node according to some other feature $x_j^{(0)} \leq \kappa_j$, etc. This procedure segments feature space

---

[2]In general, there could be more than one label for each feature vector; however, for the cases studied in this paper, the label corresponds to a class the input belongs to, labeled by an integer.

(which is in our case $\mathbb{R}^N$) along hyperplanes $x_i = \kappa_i$ with the goal to find regions such that all inputs in that region belong to the same class.

At each node, it is checked how many of the data carry which label. For single membership classification problems, which is what we will be using, the labels are just the different classes which the input feature vector belongs to. A typical loss function is the Gini impurity of a node, which measures how "impure" the data at that node actually is, i.e., how many features with different classes are in the region in feature space corresponding to this node. Denoting the set of features in the region of node a by N_a, we find for K classes the fraction of elements that belong to a class $y_k \in K$ via

$$p_{a,k} = \frac{1}{|N_a|} \sum_{i \in N_a} \delta_{i,k}\,. \tag{2.1}$$

The Gini impurity G_a at node a can then be written as

$$G_a = \sum_{k=1}^{K} p_{a,k}(1 - p_{a,k})\,. \tag{2.2}$$

In particular, if all elements of N_a belong to the same class, $G_a = 0$. In such a case, the node is turned into a leaf, since no further splits are necessary.

The decision tree is now trained by starting from the root node and trying to split by any of the F features. For κ_i, one tries all[3] intermediate values between consecutive values of feature i. The solution that leads to the lowest Gini impurity at the child nodes is accepted, and the procedure is repeated for the two child nodes and the remaining features, etc.

In cases where the map from the input to the labels is not one-to-many, one can eventually reach a perfect classification, if need be with a single element in each region. Typically, this is undesired and hence one stops splitting a node if there are less than some fixed number of elements in its corresponding region. Turning this around, if the minimal number at which a node is split is set to 2, and if the tree does not find a solution where all leaves have Gini impurity zero, this means that the map defined by the input-label-pairs is many-to-one, i.e., even all features combined are not sufficient to distinguish between the class labels.

2.2 Divisors and line bundles on dP_3

While in the general F-theoretic setup, matters curves $C_{\mathbf{R}}$ are a priori defined on a threefold $\mathcal{B}_3$, in most models there is a distinguished surface $S \subset \mathcal{B}_3$ that is wrapped by the 7-branes supporting a non-abelian gauge theory, in which the matter curve sits. A part of the complex structure moduli then parametrizes deformations of the curve inside S, which will in general affect the vector-like spectrum. These deformations can be described by pulling back all defining polynomials on $\mathcal{B}_3$ onto S, and then simply consider the coefficients of these in terms of the homogeneous coordinates on S.

---

[3]In case of many different values for a feature, this might be unfeasible, in which case a number of equally spaced values are tried for κ_i.

For our data collection, we will mimic such a "pulled back" description by focusing on curves embedded inside the del Pezzo surface dP_3. One advantage of this choice is that dP_3 has a toric description in terms of a reflexive polygon, which simplifies many computations. Another one is that it fits the setup for section 3, where we consider an F-theory toy model with non-abelian gauge degrees of freedom localized precisely on a dP_3 surface.

To set the notation, we denote the toric coordinates of dP_3 by $x_i, i = 1, \ldots, 6$. They are graded by homogeneous scalings with associated divisor classes, which are summarized in the following table:

$$
\begin{array}{c|cccccc}
 & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\
\hline
H & 1 & & & 1 & 1 \\
E_1 & -1 & 1 & & -1 & \\
E_2 & -1 & & 1 & & -1 \\
E_3 & & & & -1 & -1 & 1
\end{array}
\tag{2.3}
$$

The columns give the divisor classes of the coordinate's vanishing loci. E.g., $[\{x_1\}] = H - E_1 - E_2$. The Stanley-Reisner ideal is

$$
I_{\mathrm{SR}} = \langle x_3 x_6, x_2 x_6, x_1 x_6, x_4 x_5, x_2 x_5, x_1 x_5, x_3 x_4, x_1 x_4, x_2 x_3 \rangle \,,
\tag{2.4}
$$

and the anti-canonical class is $-K_{dP_3} = \sum_i [\{x_i\}] = 3H - E_1 - E_2 - E_3$. The independent intersection numbers are

$$
H^2 = 1 \,, \quad E_i \cdot E_j = -\delta_{ij} \,, \quad H \cdot E_i = 0 \,.
\tag{2.5}
$$

In order to simplify the notation, we introduce the short-hand notation $(a; b, c, d)$ with $a, b, c, d \in \mathbb{Z}$ for a divisor $D = aH + bE_1 + cE_2 + dE_3$.

We then define curves C inside dP_3 via $C = \{P = 0\} \equiv V(P)$ with

$$
P = \sum_i c_i m_i(x_1, \ldots, x_6) \,,
\tag{2.6}
$$

where $m_i(x_1, \ldots, x_6)$ are monomials of appropriate multi-degree under the grading in eq. (2.3). Importantly, the coefficients c_i parametrize the shape of the curve and thus model (parts of) the complex structure parameters of a global F-theory compactification. The (arithmetic) genus of the curve depends only on the divisor class $[C]$ of the curve (equivalently, the multi-degree of the monomials in P) and is given via adjunction formula as

$$
g = 1 + \frac{1}{2}[C] \cdot ([C] + K) \,.
\tag{2.7}
$$

Next, we also need to specify a line bundle $\mathcal{L}$ on C. Again, instead of focusing on the most general setup, where $\mathcal{L}$ is directly specified by a set of points on C, we consider the slightly simpler cases where $\mathcal{L}$ is a pullback of a line bundle $L = \mathcal{O}_{dP_3}(D)$ on dP_3:

$$
\mathcal{L} = \mathcal{O}_{dP_3}(D)|_{dP_3} \,.
\tag{2.8}
$$

One can think of the points then as the (weighted) intersections $\{a_i \, p_i\}$ between C and a generic representative in the class D. Note that in this case, another representative of D,

intersecting C at $\{b_j\,p'_j\}$, necessarily must give the same divisor on C, i.e., $\{a_i\,p_i\} \sim \{b_j\,p'_j\}$ are linearly equivalent on C. However, in general we cannot say anything about linear equivalences among any two of the points. Therefore, we expect, and also will find, that even for pullback line bundles, there can be special divisor alignments, i.e., p_1 and p_2, say, move into special positions, when we deform C, thus leading to jumps in the cohomology.

2.3 Generating the data set

We generate training data by picking 6 different curve classes $[C]$ with genus $1 \le g \le 6$. For each class we consider several line bundles L on dP_3 and compute (using techniques from [32]) the cohomologies $h^i(C(\mathbf{c}), L|_{C(\mathbf{c})})$, where we vary the curve $C(\mathbf{c})$ by considering all possible combinations of $c_i \in \{0, 1\}, i = 1, \ldots, d$ for the coefficients.[4] This way, we calculate cohomologies of L pulled back to $2^d - 1$ genus g curves in the class $[C]$. While this seems to be a very limited choice, it nevertheless reveals enough structures to correlate jumps in cohomology with degenerations of the geometry. On the other hand, it also introduces some bias in the data. For example, a common way the curve degenerates is if all monomials in the defining polynomial share a common variable; this happens frequently if many c_i are set to 0. However, for certain polynomials, restricting $c_i \in \{0, 1\}$ misses out possible factorizations, where factors are not just a single variable. We will see later that we can easily generalize the interpretation based on our data with algebraic methods to these cases as well.

For this data set, we then compute/collect the following features for each choice of line bundle L on each curve C with coefficients c_i:

F1) The coefficients c_i that define the curve.

F2) The genus of the curve.

F3) The number of global sections of the line bundle on the curve.[5]

F4) Are the curves smooth?

F5) The number of components the curve splits into.

F6) Are the splits smooth?

F7) Are the splits reduced?

F8) The genera of the split components.

F9) The intersection numbers among the split components.

Note that all of this data is numerical (the true/false features are encoded as 1/0). We aggregate the features F4–F9 into a single feature called the *split type*. We want to consider two curves as identical if their features F4–F9 are identical (up to relabeling the individual components). In order to check this, we would in principle have to check all permutations of all split components and see whether any of them have the same data. Since this becomes prohibitively expensive, we perform the following necessary checks:

---

[4]We exclude the case where all $c_i = 0$.

[5]Note that $h^1(C, \mathcal{L})$ can then be computed from the index which is topological and independent of c_i.

- Are the data F4 and F5 identical for the two curves?

- Are the data F6–F8 identical as sets for the two curves? This can be checked by ordering the tuples and comparing them, which is much faster than checking actual permutations.

- Is the determinant of the intersection matrix in F9 identical for the two curves? Note that the determinant is permutation invariant. However, at that point we do not check whether the permutations that make all sets match are actually the same.

Curves which are identical under these checks are assigned the same integer that encodes the split type.

Equipped with this data, we generate four different data sets which we use to train the decision trees and compare the results. In the first, we use the coefficients c_i as features and assign a label of 0 if the cohomology dimension of $H^0(C(\mathbf{c}), \mathcal{L})$ has the generic (i.e., the lowest) value and a label of 1 if there is a jump. Note that at this point, we only classify the curve according to whether a jump occurs, but not according to how large the jump is. For the second data set, we use the same labels, while the features are taken to be the topological intersection numbers between the curve components and the line bundle divisors. For the third data set we use the split type as explained above. Finally, for the fourth data set, we use both the split type and the topological intersection numbers between the curve components and the line bundle divisor as features. In addition, we perform a train:test split of 90:10 for all four data sets.

2.4 Decision trees to learn cohomology jumps

Training the decision trees only takes a few seconds on a modern desktop computer. We train a separate decision tree for each line bundle and each of the four data sets. It is instructive to compare the performance of all four training sets on both the train and the test set.

The results for the accuracy of the trained trees on the test set are summarized in figure 1. One notices that the accuracy of all data sets improves with the genus of the curve. This is due to the fact that the size of the data set grows with the genus: the genus 0 curve, that we are considering, admits only 7 coefficients c_i and hence only $2^7 - 1 = 128$ data points per line bundle. In contrast, our genus 6 curve has $2^{18} - 1 = 262143$ data points.

For the blue data points, which uses the coefficients c_i as labels, we find that the decision tree performs best. This is to be expected, since these are the finest feature set, i.e., the one with the most information, out of the four feature sets we studied. Indeed, the trees reach an accuracy of essentially 1 as soon as the training set becomes large enough (there are 3685 points in the training set for genus 3). For the other three data sets, we see that they perform worse, but still reach high accuracies. Using just the split type as a feature, for the larger genus cases where enough data is available, we reach accuracies around 80 to 85 percent. Using the intersection numbers, accuracies around 94 percent are obtained. Lastly, combining the split type and the intersection numbers, improves the results obtained when either is used individually, to an accuracy of around 97 percent. This
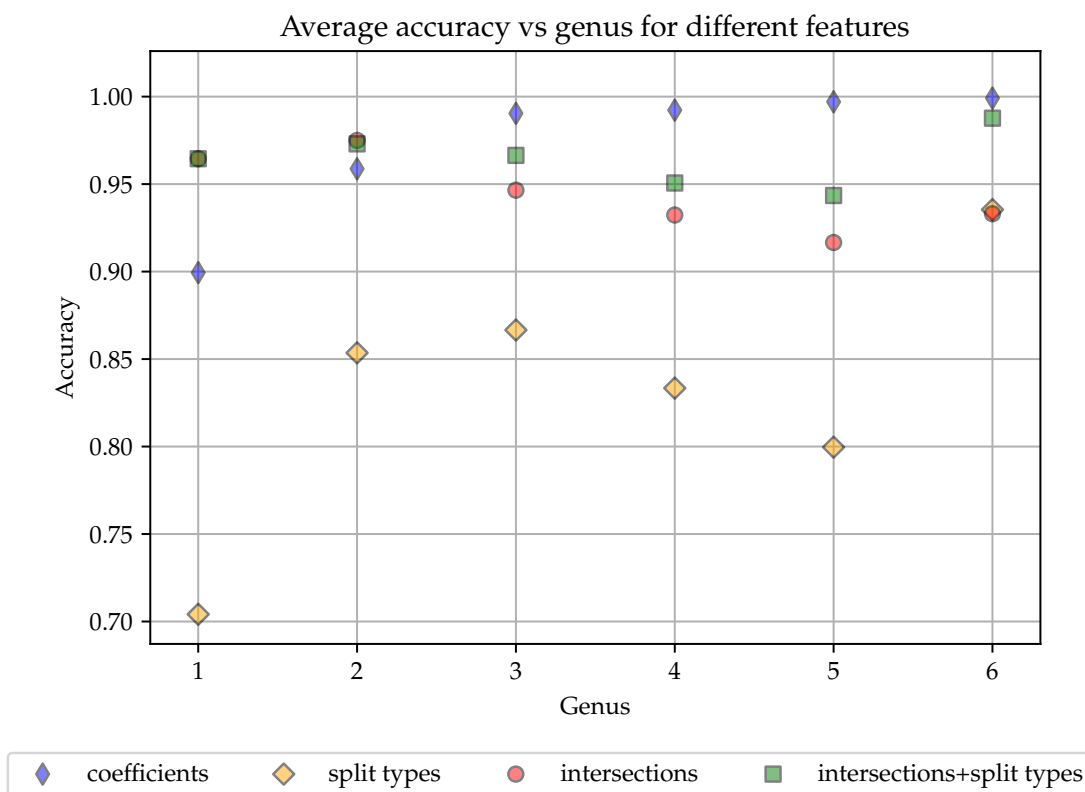
Figure 1. Average accuracy on the test set as a function of the genera of the curves for different features.

means that the two features contain different types of information which the three can use in order to improve its prediction when given access to both.

One can learn more information about the data by also analyzing the performance on the training set, as explained in section 2.1. Indeed, we find that, when not imposing constraints on the tree, the accuracy on the train set when using the coefficients as features is always 100 percent. This is not surprising, since the coefficients uniquely identify each case and hence the tree can learn a sequence of splits that puts each data point in the correct leaf node (if necessary, this leaf might only contain this single data point). For the other data sets, we find that the performance on the test set is already below 100 percent. Hence, the features are not enough to decide whether a jump in cohomology occurs, not even in principle.

Let us illustrate this by looking at the decision tree trained on the full data set for a genus three curve $D_C = (4; -1, -1, -1)$ inside dP_3 with line bundle $D_L = (1, 2, -2, -1)$, cf. section B.1.5. We give the full decision tree in figure 2. Looking at the root node, we see that for this bundle, there are 4095 different data points ("samples"). Out of these, 1791 exhibit a cohomology jump for this line bundle, while 2304 do not. The tree assigns a class label to this (non-leaf) node based on the majority, which is "no jump". However,

there are almost as many data points with a jump as there are data points without, which is why the uncertainty is high. This is encoded in the light blue color: the more certain a node predicts no jump, the darker blue it is colored. Similarly, the more certain there is a jump, the darker orange it is.

Recall that integers labelling the split type (based on the features F4–F9) are by construction small if the number of components the curve splits into is small. Hence, small split types correspond to irreducible curves, or curves with only few split components. We expect such curves being close to generic (in a sense that will be made mathematically more precise in section 4), hence the cohomologies should also take generic values.
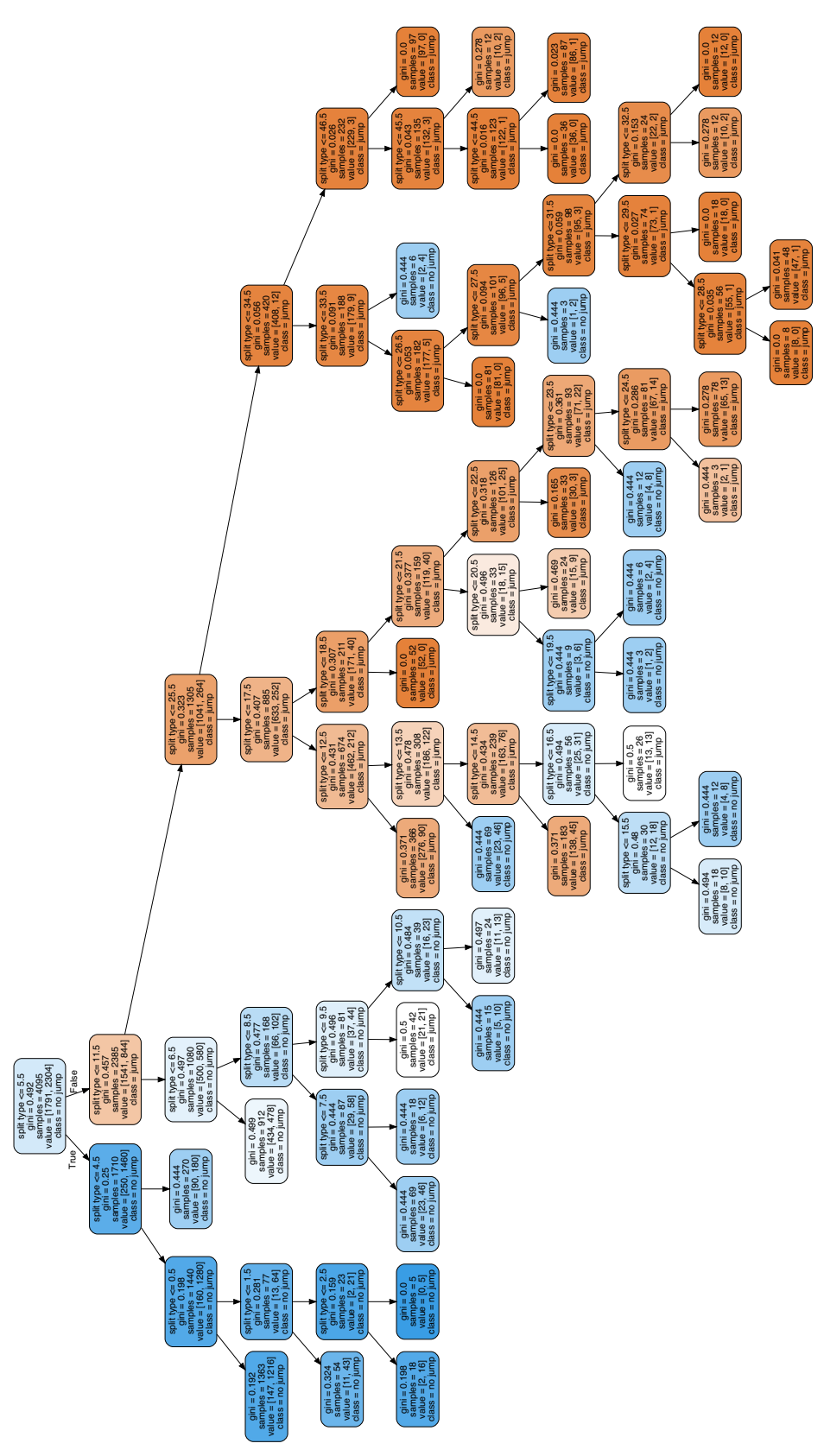
Figure 2. Trained decision tree that classifies the presence of cohomology jumps based on split types for the genus three curve example.

Indeed, we observe that the first split is performed according to whether or not the split type is smaller than 5.5. This first split already gives a good indicator in the sense that out of the 1710 training data points that have a split type of 5 or smaller, 85 percent actually do not have a jump in their cohomologies. This also illustrates that decision trees can be used for feature selection: important features that are good indicators for the classes tend to be used for splitting higher up in the tree, while more unimportant features are used further down the tree (or not at all, if they do not have any predictive power for the class membership). Now, in our case, we only have a single feature, but it is a composite feature of several quantities. The fact that the first split does not occur around the median (which would be 27) but at much smaller value indicates that the number of split components is a good criterion to distinguish jumps.

While the split types are integers, the tree always chooses half-integer decision boundaries. The reason is that the tree does not know that the feature only takes integer values. Hence, splitting in the middle between the feature values that appear in the train set will allow the most slack in either direction when the tree is presented with unseen data.

By focusing on the leaf nodes, we can also see that the tree is not classifying the data perfectly, not even the training data. Indeed, many nodes have a non-zero Gini impurity, i.e., both curves with and without jumps share the same split type associated with this leaf node. Looking for example at the bottom right leaf node, we see that 48 curves have the same split type (with value 29). However, one of these does not jump while the others do. This means that the topological data F4–F9 used to construct the split type is not enough to decide whether or not a cohomology jump occurs.

2.5 Interpretation of results

2.5.1 Jumps from curve splittings

We have seen that the decision tree trained on a combination of split types and intersection numbers performs very well. Moreover, the tree trained with just the split types splits on small split types first. This suggests that there is a tight correlation between changes in the topology of the curve and jumps in the line bundle cohomology. In particular, the data set has an abundance of cases with jumps where the curve C splits off one or more rigid components: for 78 (about 95%) of the 82 pairs of geometries D_C and line bundles D_L considered in our database, we find that we can split off a rigid component E, i.e., $C \to \tilde{C} \cup E$, such that

$$h^0_{\min}(\tilde{C} \cup E, L|_{\tilde{C} \cup E}) > h^0_{\min}(C, L|_C) \,. \tag{2.9}$$

Put differently, for almost all pairs (D_C, D_L) in our database, there exists a rigid divisor such that splitting off this rigid divisor from the curve C leads to a jump in the number of global sections on that curve. At the same time, for a given combination (D_C, D_L), we observe a jump of $h^0_{\min}$ only for a subset of all possible splits $C \to \tilde{C} \cup E$, suggesting that E and D_L must have some correlation in order for the cohomology to enhance. We list the details of these splittings and jumps in section B.1.

It is obvious that the jumps stemming from rigid component splittings can be associated with the curve C becoming non-generic. While per se not unexpected, the machine

learning process reveals — without explicitly "knowing" algebraic geometry — these features.

It is important in this context to address the bias in the data coming from considering only values of $\{0, 1\}$ for the coefficients. Namely, within the data, we only observe jumps associated with splittings of rigid components. Naively, one might conclude that rigidity of a split component is a necessary condition. However, as we already stressed in the beginning of subsection 2.3, setting enough coefficients to 0 usually factors out one of the homogeneous coordinates x_i. The corresponding curve splitting then always involves the toric divisor $V(x_i)$ which on a dP_3 is rigid for any $i = 1, \ldots, 6$. Therefore, the strong correlation between a rigid component and a jump is likely due to the bias in the data.

Indeed, we will find in section 4 and section 5 with insights from algebraic geometry, that the main source for cohomology jumps in cases of curve splittings is actually insensitive to components being rigid. We will also supplement a concrete example in subsubsection 4.1.3 where we find a jump from non-rigid curve splittings. Furthermore, we will combine these arguments with the intuition about curve splittings we gained through the data to phrase a sufficient condition for a jump in cohomology to occur in terms of topological data only. We will discuss this idea in section 5.

2.5.2 Unpredicted jumps

The fact that the decision tree cannot predict all jumps hints towards sources for additional sections (and hence cohomology jumps) beyond curve splitting. Within the data set, we observe that in rare occasions, the curve remains smooth despite a deformation which induces a jump.

For illustration purposes, consider again the genus three curve with the line bundle discussed above. Generically, this genus 3 curve is cut out by the polynomial

$$
\begin{aligned}
P(\mathbf{c}) = {} & c_1 x_1^3 x_2^3 x_3^2 x_4 + c_2 x_1^2 x_2^3 x_3 x_4^2 x_6 + c_3 x_1 x_2^3 x_4^3 x_6^2 + c_4 x_1^3 x_2^2 x_3^3 x_5 + c_5 x_1^2 x_2^2 x_3^2 x_4 x_5 x_6 \\
& + c_6 x_1 x_2^2 x_3 x_4^2 x_5 x_6^2 + c_7 x_2^2 x_4^3 x_5 x_6^3 + c_8 x_1^2 x_2 x_3^3 x_5^2 x_6 + c_9 x_1 x_2 x_3^2 x_4 x_5^2 x_6^2 \\
& + c_{10} x_2 x_3 x_4^2 x_5^2 x_6^3 + c_{11} x_1 x_3^3 x_5^3 x_6^2 + c_{12} x_3^2 x_4 x_5^3 x_6^3 \, .
\end{aligned}
\tag{2.10}
$$

The pullback of $\mathcal{O}_{dP_3}(D_L)$ onto C defines a line bundle $\mathcal{L}$ of degree $d = 3$. By Riemann-Roch we have $\chi(\mathcal{L}) = h^0 - h^1 = 1$.

In our database, we have computed the number of global sections for this line bundle for coefficient choices $\mathbf{c} \in \{0, 1\}^{12} - \mathbf{0}$. For these 4095 curves, we find

- $h^0 = 1$: 2304 (56.3%) ,

- $h^0 = 2$: 1664 (40.6%) ,

- $h^0 = 3$: 127 (3.1%) .

Our database indicates that a jump to $h^0 = 3$ occurs whenever $c_1 = c_2 = c_3 = c_{11} = c_{12} = 0$. This corresponds to a splitting

$$
C = V(x_2) \cup V(x_5) \cup V(P|_{c_1 = c_2 = c_3 = c_{11} = c_{12} = 0}) \, .
\tag{2.11}
$$

The majority of the cases with $h^0 = 2$ are where either $V(x_2)$ or $V(x_5)$ splits off, each being a rigid $\mathbb{P}^1$. This is in line with the above observation. However, we also have instances (about 9% of all curves with $h^0 = 2$) where the curve remains smooth and irreducible. Despite having $h^0 = 2$, the split type features cannot distinguish these cases from the generic setup with $h^0 = 1$, thus leading to an imperfect performance of the decision tree.

While we will come back to a detailed discussion of this phenomenon and the associated algebraic description in terms of Brill-Noether theory in subsection 4.2, it is evident that these cases of jumps are associated to the line bundle $\mathcal{L}$ on C becoming non-generic. Moreover, we also observe that such Brill-Noether-type jumps can sometimes produce values of h^0 that cannot be obtained by splittings off rigid curve components. This becomes particularly important in F-theory models, as we will discuss now.

3   Application: F-theory model building

In the previous section, we have used machine learning techniques to gain some intuition on how line bundle cohomologies jump under complex structure deformations. While we will discuss the underlying "precise" description of these various sources of jumps in the next section, we would like to show that these "rules of thumb" inferred from the white-box machine learning results can be applied directly in string phenomenology. To this end, we consider an F-theory toy model and exemplify how curve splittings help "controlling" the number of vector-like pairs.[6]

Let us first summarize the relevant features of the model, whose explicit construction is detailed in [32]. The model has an SU(5) gauge symmetry localized on a dP_3 surface inside the compact base threefold $\mathcal{B}_3$, which itself is a smooth hypersurface inside a toric variety. There are matter states in the representations $\mathbf{10}_1$, $\mathbf{5}_3$ and $\mathbf{5}_{-2}$, where the subscript denote the charges under an additional U(1) gauge symmetry. Each representation $\mathbf{R}$ resides on a curve $C_{\mathbf{R}}$ inside the dP_3 surface. One can find a globally consistent vertical G_4-flux configuration that induces the chiral spectrum

$$\chi(\mathbf{10}_1) = 3\,, \quad \chi(\mathbf{5}_3) = 15\,, \quad \chi(\mathbf{5}_{-2}) = -18\,. \tag{3.1}$$

In the following, we will analyze in detail the vector-like spectrum in this setup.

Geometry of curves.   In the global geometry, the matter curves $C_{\mathbf{R}}$ are complete intersections involving the dP_3 surface and another divisor on the base $\mathcal{B}_3$. As discussed in [32], a generic choice of the complex structure parameters for the elliptic fourfold also induces a generic curve $C_{\mathbf{R}}$ on dP_3. In other words, we can parametrize them in terms of global sections of $\mathcal{O}_{dP_3}([C_{\mathbf{R}}])$, where $[C_{\mathbf{R}}]$ denotes the divisor class of the curve inside dP_3.

Furthermore, the data defining the zero mode spectrum in a global F-theory model can be extracted from the G_4-configuration and packaged into a line bundle (or, more generally, a coherent sheaf) for each curve $C_{\mathbf{R}}$ [30, 31]. For the case at hand, the flux inducing the

---

chiral spectrum in eq. (3.1) induces line bundles which are pullbacks of various bundles on dP_3 to the curves [32].

Using the same notation as in the previous section,[7] the curves with their genus and their corresponding zero-modes counting bundles are:

curve	class	genus	bundle	h^i
$C_{\mathbf{10}_1}$	$(4; -1, -1, -2)$	2	$\mathcal{O}_{dP_3}(1; -1, -1, 1)$	$(3, 0)$
$C_{\mathbf{5}_3}$	$(10; -3, -3, -4)$	24	$\mathcal{O}_{dP_3}(5; -4, -4, 3)$	$(15 + n, n)$
$C_{\mathbf{5}_{-2}}$	$(17; -5, -5, -7)$	79	$\mathcal{O}_{dP_3}(6; 0, 0, -6)$	$(7, 25)$

$$(3.2)$$

Note that the cohomologies on $C_{\mathbf{10}_1}$ and $C_{\mathbf{5}_{-2}}$ are fixed by the exactness of the corresponding Koszul resolutions, and hence there are no complex-structure-dependent jumps possible.[8] For the representation $\mathbf{5}_3$, no such arguments apply, and thus we expect the number n of light vector-like pairs to vary.

The curve $C_{\mathbf{5}_3} = \{a_{3,2} = 0\}$ is the vanishing locus of a polynomial with class $(10; -3, -3, -4)$, whose explicit expression in the parametrization of the toric dP_3 coordinates x_i are given in section A, cf. eq. (A.67). With the curve having genus 24, it would be almost impossible to perform a scan by varying all the complex structure parameters (eq. (A.67) has 44 coefficients), as we did previously for the low genus cases. However, the intuition we gained from the low genus examples will help us to "control" n — that is, to efficiently find suitable geometries realizing the desired vector-like spectrum.

3.1 Engineering jumps in cohomology

What we have learned from the machine learning results is that the line bundle cohomology is more likely to jump if the curve in question is reducible. Though we have already emphasized that rigidity of the components is not necessary, the abundance of toric coordinates makes it handy to factor out various different curves which in this case happen to be rigid. For the purpose of finding *a* concrete realization of a particular jump in the vector-like spectrum, these rigid factors turn out to be sufficient.

We thus modify the coefficients of the defining polynomial $a_{3,2}$ in eq. (A.67) such that individual toric coordinates x_i of dP_3 factor out. Of course, not every such factorization will lead to a jump: the rigid component must in some way receive a "non-trivial contribution", i.e., intersection, from the divisor D_L defining the line bundle. The intuitions we gained from the previous section is that a *negative* intersection of D_L with $V(x_i)$ will lead to a jump. It is then intuitive to assume that the more rigid components splits off, the higher the jumps tend to be. With this intuition, we now proceed to engineer step-wise jumps of the vector-like spectrum.

Using the linear relations in eq. (2.3) and intersection numbers in eq. (2.5), we easily verify the divisor defining the line bundle, $D_L = 5H - 4E_1 - 4E_2 + 3E_3$, has only negative

---

[7]Divisor classes $aH + bE_1 + cE_2 + dE_3$ are denoted by $(a; b, c, d)$.

[8]This can change if we modify the flux by, e.g., horizontal pieces. However, for the purpose of this work, we focus on jumps induced by geometric changes.

intersections with $[x_1]$ and $[x_6]$. Inspecting eq. (A.67), one finds that if we set

$$c_{40} = c_{41} = c_{42} = c_{43} = c_{44} = 0 \,, \tag{3.3}$$

the polynomial factors as $a_{3,2} = x_6 R_2$, where R_2 is an irreducible polynomial in the class $(10; -3, -3, -5)$. And indeed, a computer-assisted computation with methods from [32] reveals that for this curve $C_2 = \{x_6 R_2 = 0\}$, we have

$$h^i(C_2, \mathcal{O}_{dP_3}(5; -4, -4, 3)|_{C_2}) = (17, 2) \,, \tag{3.4}$$

We can factor out another factor x_6 from R_2 by setting

$$c_{34} = c_{35} = c_{36} = c_{37} = c_{38} = c_{39} = c_{40} = c_{41} = c_{42} = c_{43} = c_{44} = 0 \,, \tag{3.5}$$

yielding $C_{\mathbf{5}_3} \rightarrow C_3 = \{x_6^2 R_3 = 0\}$, with R_3 an irreducible polynomial of class $(10; -3, -3, -6)$. In this case, we find a jump by three,

$$h^i(C_3, \mathcal{O}_{dP_3}((5; -4, -4, 3)|_{C_3}) = (18, 3) \,. \tag{3.6}$$

To achieve a jump by four, we factorize $C_{\mathbf{5}_3} \rightarrow C_4 = \{x_1 x_6 R_4 = 0\}$, with $[R_4] = (9; -2, -2, -5)$, with the following choice of complex structure:

$$c_1 = c_2 = c_3 = c_4 = c_5 = c_{40} = c_{41} = c_{42} = c_{43} = c_{44} = 0 \,. \tag{3.7}$$

Then we find

$$h^i(C_4, \mathcal{O}_{dP_3}((5; -4, -4, 3)|_{C_4}) = (19, 4) \,. \tag{3.8}$$

Also a jump by five is possible. To this end, we set

$$c_1 = c_2 = c_3 = c_4 = c_5 = c_{34} = c_{35} = c_{36} = c_{37} = 0 \,,$$
$$c_{38} = c_{39} = c_{40} = c_{41} = c_{42} = c_{43} = c_{44} = 0 \,. \tag{3.9}$$

For this choice of parameters, the matter curve factorizes as $C_{\mathbf{5}_3} \rightarrow C_5 = \{x_1 x_6^2 R_5 = 0\}$, with $[R_5] = (9; -2, -2, -6)$. In this case we have

$$h^i(C_5, \mathcal{O}_{dP_3}((5; -4, -4, 3)|_{C_5}) = (20, 5) \,. \tag{3.10}$$

Lastly, we also easily construct a model with six vector-like pairs, by setting

$$c_1 = c_2 = c_3 = c_4 = c_5 = c_6 = c_7 = c_8 = c_9 = c_{10} = c_{11} = 0 \,,$$
$$c_{34} = c_{35} = c_{36} = c_{37} = c_{38} = c_{39} = c_{40} = c_{41} = c_{42} = c_{43} = c_{44} = 0 \,. \tag{3.11}$$

Then $C_{\mathbf{5}_3} \rightarrow C_5 = \{x_1^2 x_6^2 R_6 = 0\}$, with $[R_6] = (8; -3, -3, -6)$. This gives

$$h^i(C_5, \mathcal{O}_{dP_3}((5; -4, -4, 3)|_{C_5}) = (21, 6) \,. \tag{3.12}$$

3.2 Single vector-like pair from Brill-Noether theory

The above examples demonstrate how the machine learning intuition led us to a step-wise increase in the number of vector-like pairs by suitable tuning of the complex structure parameters. These jumps occur because the matter curve in question splits into several components. However, for this curve and line bundle, such splittings induce a jump from zero vector-like pairs to at least two (or three, or four, or five). If we are interested in models with a single vector-like pair — such as for the Higgs field in MSSM realizations — then we need to look for other effects than curve splitting.

As we have seen earlier, such effects are related to the cases not predicted by the trained decision tree. Here, the jumps in cohomology are not due to the curve becoming non-generic, but rather the line bundle. In fact, Brill-Noether theory (to be discussed in the next section, see also subsection A.1) tells us that for the matter curve $C_{\mathbf{5}_3}$ of genus 24, we expect that a scenario with a single vector-like pair — i.e., one having $h^i = (16, 1)$ — to occur on a subvariety of dimension $\rho = g - h^0 \cdot h^1 = 8$ of the space $\mathrm{Jac}(C_{\mathbf{5}_3})$ which parametrizes the line bundles on $C_{\mathbf{5}_3}$. Note that the same formula would yield $\rho = -10$ for jumps by two, and hence no such jumps can occur *for a generic $C_{\mathbf{5}_3}$*. This agrees with the above instances, as each of those requires the curve to become non-generic.

Because of this, engineering the jump by 1 becomes more challenging, and in particular requires additional tools from algebraic geometry. We defer the details of the relevant computations to section A and simply remark here that the necessary tuning is

$$
\begin{aligned}
& c_1 = c_2 = c_3 = c_4 = c_5 = c_7 = c_8 = c_9 = c_{10} = c_{35} = c_{36} = c_{37} = c_{38} = 1\,, \\
& c_{40} = c_{41} = c_{42} = c_{43} = c_{44} = 1, \quad c_{11} = c_{34} = -1, \quad c_6 = c_{39} = 2\,.
\end{aligned}
\tag{3.13}
$$

One can easily verify that the polynomial $a_{3,2}$ in eq. (A.67) does not factorize in this case, and that the curve $C_{\mathbf{5}_3}$ remains smooth. Therefore, the enhancement in cohomology in this case is indeed of Brill-Noether type.

4 Cohomology jumps throughout the moduli space

To put the intuition we gained from machine learning onto more solid grounds, we now apply tools from algebraic geometry to develop a more complete, "microscopic" understanding for the various sources of jumps we encountered in our data. As we will see, the resulting insights lead to a diagrammatic representation of a *stratification* of the complex structure moduli space of F-theory compactifications induced by vector-like spectra.

As we have alluded to in section 2, based on our database we can essentially distinguish two types of jumps:

1. Jumps due to a non-generic line bundle.

2. Jumps due to a non-generic curve.

This shows that our samplings are very atypical. Namely, true jump loci have lower dimensionality than the full set of parameters. Therefore, jump loci form sets of measure 0 and should never be encountered by a genuinely random sample.

It is central to our discussion that algebraic geomemtry can bound from below the 'size' of such jump loci. In particular, this is true for jumps due to non-generic line bundles. Such jumps have been analyzed since 1874 in the context of *Brill-Noether theory*[9] [47]. Given a generic curve C_g of genus g and an integer d, Brill-Noether theory provides an integer $\rho(r, g, d)$ which measures how likely it is that a line bundle $\mathcal{L}_d$ of degree d on C_g has $r + 1$ independent non-trivial global sections, i.e., has $h^0(C_g, \mathcal{L}_d) = r + 1$.

To formulate this more precisely, first recall that the Jacobian $\mathrm{Jac}(C_g)$ of the curve C_g is isomorphic to $\mathbb{C}^g/\Lambda$ where Λ is the full-dimensional period lattice of C_g. By the Abel-Jacobi map, equivalence classes of line bundles of degree d form a copy of the Jacobian $\mathrm{Jac}(C_g)$. Let us focus on the subset of the Jacobian formed by all equivalence classes of line bundles of degree d which admit exactly $r + 1$ global sections. Then a lower bound on the dimension of this space is given by the integer

$$\rho(r, g, d) = g - (r + 1) \cdot (r + 1 - (d - g + 1)) \equiv g - n^0 \cdot n^1 \,. \tag{4.1}$$

In the last equality we use the intuitive notation $n^0 = r + 1$. Furthermore, we have used that by the Riemann-Roch theorem, $n^1 \equiv n^0 - (d - g + 1)$ is equal to $h^1(C_g, \mathcal{L}_d)$ if $h^0(C_g, \mathcal{L}_d) = n^0$. Further details on Brill-Noether theory can be found in appendix A.1, and a more complete presentation is given in [48, 49].

An important result follows from [50]: if the curve is generic, then lines bundles of degree d only admit numbers $r + 1$ of global sections for which $\rho(r, g, d)$ is non-negative. Put differently, there are no line bundles on generic curves with $r + 1$ global sections with $\rho(r, g, d) < 0$. Furthermore, the value of ρ gives a very clear notion of the likelihood to have $r + 1$ sections in terms of a dimension on the "moduli" space of line bundles.

Let us demonstrate this for a line bundle $\mathcal{L}$ of degree $d = 2$ on a curve C_g of genus $g = 3$. By general theory, the number of section of this line bundle cannot exceed its degree. Hence, it has 0, 1 or 2 sections. With this information, let us compute $\rho(r, d, g)$:

r	h^i	$\rho(r, d, g)$
-1	$(0, 0)$	3
0	$(1, 1)$	2
1	$(2, 2)$	-1

$$\tag{4.2}$$

From this we learn, that most line bundles $\mathcal{L}$ of degree $d = 2$ on a genus $g = 3$ curve C_3 satisfy $h^0(C_3, \mathcal{L}) = 0$. Since for these bundles ρ matches the dimension of the Jacobian of C_3, we can say that these line bundles are associated to generic points of the Jacobian. Furthermore, we learn that there are line bundles with $h^0(C_3, \mathcal{L}) = 1$. However, these are special in the sense that they are associated to a codimension-1 locus in the Jacobian $\mathrm{Jac}(C_3)$.

Finally, $\rho = -1$ for $r = 1$ begs for an explanation. This explanation follows from work of Griffiths and Harris [50]:

On **generic** curves, $\dim(G_d^{r+1}) = \rho(r, d, g)$.

---

[9] The physics community may find it entertaining to learn that this theory is named after Max Noether, the father of Emmy Noether.

So in particular, on generic curves it holds $G_d^{r+1} = \emptyset$ if and only if $\rho(r,d,g) < 0$. Consequently, we conclude from eq. (A.14), that on generic genus $g = 3$ curve, there is no line bundle $\mathcal{L}$ of degree 2 such that $h^0(C_3, \mathcal{L}) = 2$.

Note however, that this does not rule out the possibility that non-generic curves may host such line bundles. In the case at hand, it follows from the theorem of Clifford [50] that hyperelliptic curves H_3 of genus $g = 3$ admit line bundles $\mathcal{L}$ of degree $d = 2$ and $h^0(H_3, \mathcal{L}) = 2$. Note that hyperelliptic curves of genus $g > 2$ are non-generic. Hence, this points us to jumps of the vector-like spectrum, which originate from non-generic deformations of the curve.

Let us give another such example, which illustrates a jump on a singular curve. To this end, let us consider a line bundle $\mathcal{L}$ of degree $d = 5$ on a genus $g = 2$ curve. Then $\chi(\mathcal{L}) = 4$ and $h^0(C_2, \mathcal{L}) \in \{4, 5\}$. Let us compute $\rho(r, d, g)$ for these two values of global sections:

$$
\begin{array}{c c | c}
r & h^i(C_2, \mathcal{L}) & \rho(r,d,g) \\
\hline
3 & (4,0) & 2 \\
4 & (5,1) & -3
\end{array}
\tag{4.3}
$$

Thus, on a generic curve of genus $g = 2$, any line bundle of degree $d = 5$ has 4 global sections. Even more, since the degree d is in the stable range, we find 4 global sections for this line bundle on every smooth curve of genus $g = 2$ — generic or not. Hence, 5 sections can only be realized on a singular curve.

This can be achieved by choosing the curve parameters (which model the complex structure moduli of global F-theory models) such that the curve becomes reducible, and factors into various components which intersect transversely in a number of points. A way to construct global sections on such curves is then as follows: first, consider each component individually and identify which sections they support. Then, by demanding that these sections agree at the intersection points, we glue these local sections to global sections. We will return to this gluing procedure in more detail in section 5.

In this section, we will take a closer look at the interplay of jumps that occur due to non-genericity both of the line bundle and the curve. In particular, since in global F-theory models, both the bundle and the curve depend on the complex structure parameters of the elliptic fibration in the same fashion (namely through the coefficients of its defining polynomials), they should be treated on the same footing, which we can summarize diagrammatically. The following analysis requires, at a technical level, a working understanding of the Koszul resolution of a pullback bundle, its associated long exact sequence in sheaf cohomology, inferring the maps in this long exact sequence from Čech ochomology as well as a basic understanding of non-reduced curves. For convenience of the reader, further details are provided in section A.

4.1 Jumps from curve splittings

We first analyze examples with jumps from curve splittings. We will see that rigidity of the components that split off play no role in the section counting. The reason, why we found in earlier sections that rigid divisors lead to jumps, is due to our special choice of

setting all coefficients in the polynomial that specify the curve in dP_3 to either zero or one. Namely, this factors the curve into a combination of toric divisors — which happen to be rigid in dP_3 — and remaining curve components.

4.1.1 Example: one additional section

Setup. Let us return to the example of a line bundle on a genus 2 curve discussed above. In more detail, the curve and line bundle are given by

$$D_C = (4; -1, -2, -1), \qquad D_L = (3; -3, -1, -2). \tag{4.4}$$

The curve $C(\mathbf{c}) = V(P(\mathbf{c}))$ is defined by a polynomial $P(\mathbf{c}) \in H^0(dP_3, \mathcal{O}_{dP3}(D_C)) \cong \mathbb{C}^{10}$ with

$$\begin{aligned} P(\mathbf{c}) = {} & c_1 x_1^3 x_2^3 x_3 x_4 + c_2 x_1^2 x_2^3 x_4^2 x_6 + c_3 x_1^3 x_2^2 x_3^2 x_5 + c_4 x_1^2 x_2^2 x_3 x_4 x_5 x_6 + c_5 x_1 x_2^2 x_4^2 x_5 x_6^2 \\ & + c_6 x_1^2 x_2 x_3^2 x_5^2 x_6 + c_7 x_1 x_2 x_3 x_4 x_5^2 x_6^2 + c_8 x_2 x_4^2 x_5^2 x_6^3 + c_9 x_1 x_3^3 x_5^3 x_6^2 + c_{10} x_3 x_4 x_5^3 x_6^3, \end{aligned} \tag{4.5}$$

where the coefficients $\mathbf{c} \in \mathbb{C}^{10}$ form the parameter space of this genus $g = 2$ setup. The line bundle $\mathcal{L}(\mathbf{c}) = \mathcal{O}_{dP3}(D_L)|_{C(\mathbf{c})}$ satisfies $\deg(\mathcal{L}(\mathbf{c})) = 5$. Hence, on smooth curves, the theorem of Riemann-Roch tells us

$$\chi(\mathcal{L}(\mathbf{c})) = \deg(\mathcal{L}(\mathbf{c})) - g + 1 = 5 - 2 + 1 = 4. \tag{4.6}$$

Moreover, since $\deg(\mathcal{L}(\mathbf{c})) = 5 > 2g-2$, we know that for smooth curves $h^1(C(\mathbf{c}), \mathcal{L}(\mathbf{c})) = 0$. Hence, $h^0(\mathcal{L}(\mathbf{c})) = 5$ is only possible on non-smooth curves.

Comparison with database. In our database, we have considered choices of parameters $\mathbf{c} \in \{-1, 0, 1\}^{10} - \mathbf{0}$. On about 96% of these 59048 curves, $\mathcal{L}(\mathbf{c})$ has 4 sections. This fits with the above picture, that generically we expect 4 sections. However, we also find 2186 curves for which $\mathcal{L}(\mathbf{c})$ has 5 sections. Those curves satisfy $c_3 = c_6 = c_9 = 0$, which means that $C(\mathbf{c}) = V(x_4) \cup B$, where

$$\begin{aligned} B = V(&c_1 x_1^3 x_2^3 x_3 + c_2 x_1^2 x_2^3 x_4 x_6 + c_4 x_1^2 x_2^2 x_3 x_5 x_6 + c_5 x_1 x_2^2 x_4 x_5 x_6^2 \\ &+ c_7 x_1 x_2 x_3 x_5^2 x_6^2 + c_8 x_2 x_4 x_5^2 x_6^3 + c_{10} x_3 x_5^3 x_6^3) \end{aligned} \tag{4.7}$$

is a genus-0 curve with $V(x_4) \cdot B = 3$. We will now argue that $\mathcal{L}(\mathbf{c})$ admits 5 sections if and only if $C(\mathbf{c})$ decomposes in this way.

Classification of jump geometries. To this end, we consider the Koszul resolution

$$0 \to \mathcal{O}_{dP_3}(D_L - D_C) \xrightarrow{\alpha} \mathcal{O}_{dP_3}(D_L) \to \mathcal{L}(\mathbf{c}) \to 0. \tag{4.8}$$

Its associated long exact sequence in sheaf cohomology takes the form

$$\begin{aligned} 0 \longrightarrow {} & 0 \longrightarrow H^0(dP_3, D_L) \cong \mathbb{C}^1 \longrightarrow H^0(C(\mathbf{c}), \mathcal{L}(\mathbf{c})) \\ \longrightarrow {} & H^1(dP_3, D_L - D_C) \cong \mathbb{C}^4 \xrightarrow{\varphi} H^1(dP_3, D_L) \cong \mathbb{C}^1 \longrightarrow H^1(C(\mathbf{c}), \mathcal{L}(\mathbf{c})) \\ \longrightarrow {} & 0 \longrightarrow 0 \longrightarrow 0 \longrightarrow 0. \end{aligned} \tag{4.9}$$

The exactness of this sequence implies that

$$h^0(C(\mathbf{c}), \mathcal{L}(\mathbf{c})) = 5 - \dim(\operatorname{im}\varphi) = 5 - \operatorname{rk}(M_\varphi) \,, \tag{4.10}$$

where $M_\varphi = (c_3, c_6, c_9, 0)$. We explain the construction of the mapping matrix M_φ in more detail in section A.

Obviously, M_φ has rank 1 iff $(c_3, c_6, c_9) \neq \mathbf{0}$ and its rank vanishes iff $(c_3, c_6, c_9) = \mathbf{0}$. This immediately leads to the following classification of curve geometries:

$\operatorname{rk}(M_\varphi)$	explicit condition	curve splitting
1	$(c_3, c_6, c_9) \neq \mathbf{0}$	C
0	$(c_3, c_6, c_9) = \mathbf{0}$	$V(x_4) \cup B$

(4.11)

showing that we obtain one additional vector-like pair if and only if the curve factors as $V(x_4) \cup B$. We illustrate this result in the following diagram:

$$\mathcal{F}_1 \colon (h^0, \rho) = (4, 1) \qquad \boxed{C}$$

$$\downarrow$$

$$\mathcal{F}_2 \colon (h^0, \rho) = (5, -3) \qquad \boxed{V(x_4) \cup B}$$

(4.12)
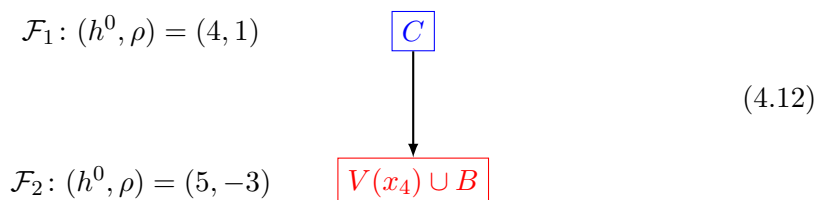
In this diagram, the ath node represents a family $\mathcal{F}_a$ of curves, for which we give the generic element in this family.

For example, the family $\mathcal{F}_1$ of curves at the first node is defined by the condition $(c_3, c_6, c_9) \neq \mathbf{0}$ and has the curve C as its generic element, which is a smooth, irreducible curve of genus $g = 2$. Note that (non-generic) members of $\mathcal{F}_1$ can also be singular curves with several components. For example, the curve $V(x_1^3 x_2^2 x_3^2 x_5)$ is defined by the condition that all c_i but c_3 vanish. This curve is clearly singular and has several connected components. Recall that $\mathcal{F}_1$ is the family of curves on which the line bundle in question admits four global sections. Hence, the statement is that even on such a very singular curve, the bundle in question admits exactly four sections.

This feature changes exactly on the family of curves $\mathcal{F}_2$, which are defined by $(c_3, c_6, c_9) \equiv \mathbf{0}$. Its generic element is a curve of the form $V(x_4) \cup B$, where B is a smooth genus $g = 0$ curve touching $V(x_4)$ in 3 distinct points. We can also view $\mathcal{F}_1 = \{\mathbf{c} \,|\, (c_3, c_6, c_9) \neq \mathbf{0}\}$ and $\mathcal{F}_2 = \{\mathbf{c} \,|\, (c_3, c_6, c_9) = \mathbf{0}\}$ as subspaces of the parameter space $\mathbb{C}^{10} \ni \mathbf{c}$. In this case it is trivial to see that

$$\mathcal{F}_1 \cap \mathcal{F}_2 = \emptyset \,, \qquad \mathcal{F}_2 \subset \overline{\mathcal{F}_1} \,, \tag{4.13}$$

where $\overline{\mathcal{F}_1}$ the closure with respect to the standard topology on $\mathbb{C}^{10}$. We will come back to this property shortly.

4.1.2 An h^0-gap

Whilst factoring-off curve components typically increases the number of global sections, this effect need not necessarily generate exactly one additional section, as we have already seen in subsection 3.1. Rather, it can force multiple additional sections to appear simultaneously. An example of this sort is

$$D_C = (3; -1, -1, -1), \qquad D_L = (1; -1, -3, -1).$$ (4.14)

In this case, $C(\mathbf{c}) = V(P(\mathbf{c}))$ is a genus 1 curve defined by

$$
\begin{aligned}
P(\mathbf{c}) = {}& c_1 x_1^2 x_2^2 x_3 x_4 + c_2 x_1^2 x_2 x_3^2 x_5 + c_3 x_1 x_2^2 x_4^2 x_6 + c_4 x_1 x_2 x_3 x_4 x_5 x_6 \\
& + c_5 x_1 x_3^2 x_5^2 x_6 + c_6 x_2 x_4^2 x_5 x_6^2 + c_7 x_3 x_4 x_5^2 x_6^2.
\end{aligned}
$$ (4.15)

Moreover, $\mathcal{L}$ is a line bundle of degree $d = -2$. Hence, its degree is in the stable regime and on any smooth curve we find $h^0(C, \mathcal{L}) = 0$. Still, as demonstrated in subsection 4.1.1, non-smooth curves can admit higher numbers of global sections. Here, we will argue, that even on singular curves, the pullback line bundle $\mathcal{L}$ can never have exactly one section.

To see this, let us look at the long exact sequence in sheaf cohomology associated to the Koszul resolution of the setup:

$$
\begin{aligned}
& 0 \longrightarrow 0 \longrightarrow 0 \longrightarrow H^0(D_C, \mathcal{L}) \\
& \hookrightarrow H^1(dP_3, D_L - D_C) \cong \mathbb{C}^3 \xrightarrow{\varphi} H^1(dP_3, D_L) \cong \mathbb{C}^5 \longrightarrow H^1(C, \mathcal{L}) \\
& \hookrightarrow 0 \longrightarrow 0 \longrightarrow 0 \longrightarrow 0.
\end{aligned}
$$ (4.16)

The exactness of this sequence implies $h^0(C, \mathcal{L}) = 3 - \dim(\mathrm{im}\,\varphi) = 3 - \mathrm{rk}(M_\varphi)$ with

$$M_\varphi = \begin{pmatrix} c_6 & 0 & 0 \\ c_7 & c_6 & 0 \\ c_3 & 0 & 0 \\ c_1 & 0 & c_3 \\ c_4 & c_3 & c_6 \end{pmatrix}.$$ (4.17)

Consequently, the statement that on the curves in class D_C the pullback of D_L never has exactly one section is equivalent to saying that M_φ never has rank 2. We see this by studying the four non-trivial and independent 3×3-minors of M_φ:

$$m_1 = c_6^2 c_3, \qquad m_2 = c_6^3 \qquad m_3 = c_3^3, \qquad m_4 = c_6 c_4 c_3 - c_7 c_3^2 - c_6^2 c_1.$$ (4.18)

Now, $\mathrm{rk}(M_\varphi) < 3$ requires $m_1 = m_2 = m_3 = m_4 = 0$. This is equivalent to $c_3 = c_6 = 0$ and

$$M_\varphi|_{c_3 = c_6 = 0} = \begin{pmatrix} 0 & 0 & 0 \\ c_7 & 0 & 0 \\ 0 & 0 & 0 \\ c_1 & 0 & 0 \\ c_4 & 0 & 0 \end{pmatrix},$$ (4.19)

which can have at most rank 1. More generally, we can classify the rank of M_φ and thereby summarize the curve geometry as follows:

$\mathrm{rk}(M_\varphi)$	explicit condition $(\mathcal{F}_i)$	splitting of curve
3	$c_3, c_6 \neq 0$	C
1	$c_3 = c_6 = 0$	$E_2 \cup B$
0	$c_1 = c_3 = c_4 = c_6 = c_7 = 0$	$E_6 \cup E_4 \cup E_2^{(2)} \cup A$

(4.20)

Observe again that within the parameter space of $\mathbf{c}$, we have

$$\mathcal{F}_i \cap \mathcal{F}_j = \emptyset, \quad \overline{\mathcal{F}_1} \supset \mathcal{F}_2, \quad \overline{\mathcal{F}_2} \supset \mathcal{F}_3. \tag{4.21}$$

The corresponding diagram is

$$
\begin{array}{ll}
\mathcal{F}_1: (h^0, \rho) = (0, 1) & \boxed{C} \\[2em]
\mathcal{F}_2: (h^0, \rho) = (2, -7) & \boxed{E_2 \cup B} \\[2em]
\mathcal{F}_3: (h^0, \rho) = (3, -14) & \boxed{E_6 \cup E_4 \cup E_2^{(2)} \cup A}
\end{array}
\tag{4.22}
$$

4.1.3 Jump from non-rigid curve splitting

We now address the bias in our data, and provide a concrete example of jumps from a curve splitting for which none of the components is rigid. To this end, we consider $D_C = (2; -1, -1, 0)$ and $D_L = (-2, 0, 4, 0)$. This curve is thus given by

$$P = c_1 x_4 x_5 x_6^2 + c_2 x_1 x_3 x_5 x_6 + c_3 x_1 x_2 x_4 x_6 + c_4 x_1^2 x_2 x_3. \tag{4.23}$$

For generic coefficients c_i, the curve C is a smooth curve of genus $g = 0$ and $\mathcal{L}$ has degree $d = 0$. Hence we conclude $h^0(C, \mathcal{L}) = 1$.

To understand jumps at special coefficients, we employ the Koszul resolution and find $h^0(C(\mathbf{c}), \mathcal{L}(\mathbf{c})) = 7 - \mathrm{rk}(M)$ where

$$
M = \begin{pmatrix}
0 & 0 & 0 & c_1 & 0 & c_3 & 0 \\
c_4 & 0 & 0 & 0 & c_1 & c_2 & c_3 \\
c_3 & 0 & 0 & 0 & 0 & c_1 & 0 \\
0 & c_1 & 0 & c_2 & c_3 & c_4 & 0 \\
0 & 0 & c_1 & c_3 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & c_2 & 0 & c_4
\end{pmatrix}.
\tag{4.24}
$$

The rank drops of this matrix include both cases of rigid and non-rigid splittings. Explicitly, let us set $A_i = V(x_i)$, which are rigid components. Moreover, let us consider the following non-rigid $g = 0$ curves:

$$
\begin{aligned}
D_1 &= V(c_2 x_3 x_5 x_6 + c_3 x_2 x_4 x_6 + c_4 x_1 x_2 x_3), \\
D_2 &= V(c_3 x_4 x_6 + c_4 x_1 x_3), \\
D_3 &= V(c_4 x_1 x_2 + c_2 x_5 x_6), \\
D_4 &= V(c_2 x_1 x_3 + c_1 x_4 x_6), \\
D_5 &= V(c_3 x_1 x_2 + c_1 x_5 x_6).
\end{aligned}
\tag{4.25}
$$

With these, we can then summarize the rank drops as follows:

rk(M)	explicit condition	curve splitting
6	generic	C
5	$c_1 = 0$	$A_1 \cup D_1$
5	$c_1 c_4 = c_2 c_3$	$D_2 \cup D_3$
3	$c_1 = c_3 = 0$	$A_1 \cup A_3 \cup D_3$

(4.26)

The corresponding diagram is of the form

$\mathcal{F}_1 : (h^0, \rho) = (1, 0)$

$\mathcal{F}_2 : (h^0, \rho) = (2, -2)$

$\mathcal{F}_4 : (h^0, \rho) = (4, -12)$

(4.27)
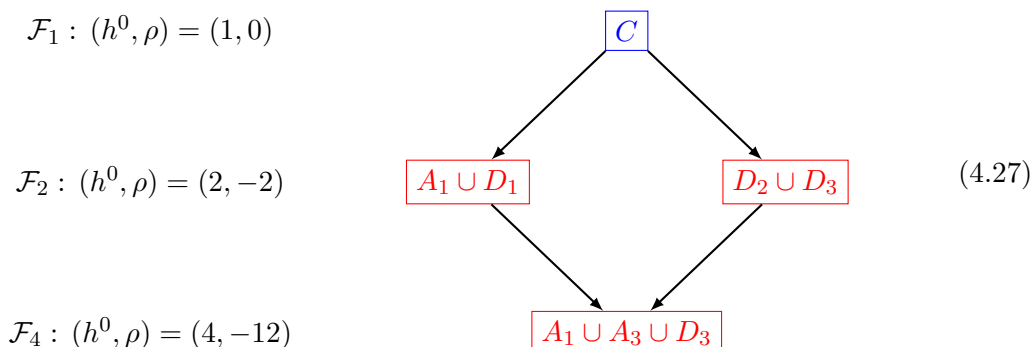
Similar to our discussion in subsubsection 4.1.2, there is a gap at $h^0 = 3$. Crucially, since D_2 and D_3 are non-rigid, the deformation $C \to D_2 \cup D_3$ provides an explicit example of a jump associated to curve splitting with no rigid components.

4.2 Jumps from non-generic line bundles

We now turn to jumps due to special alignments of the points that define a line bundle divisor. These phenomena are described by Brill-Noether theory.

4.2.1 Additional section due to special divisors

Let us consider the pair

$$D_C = (4; -1, -1, -1), \qquad D_L = (1; 2, -2, -1). \tag{4.28}$$

This genus $g = 3$ curve $C(\mathbf{c}) = V(P(\mathbf{c}))$ is defined by

$$
\begin{aligned}
P(\mathbf{c}) =\ & c_1 x_1^3 x_2^3 x_3^2 x_4 + c_2 x_1^2 x_2^3 x_3 x_4^2 x_6 + c_3 x_1 x_2^3 x_4^3 x_6^2 + c_4 x_1^3 x_2^2 x_3^3 x_5 + c_5 x_1^2 x_2^2 x_3^2 x_4 x_5 x_6 \\
& + c_6 x_1 x_2^2 x_3 x_4^2 x_5 x_6^2 + c_7 x_2^2 x_4^3 x_5 x_6^3 + c_8 x_1^2 x_2 x_3^3 x_5^2 x_6 + c_9 x_1 x_2 x_3^2 x_4 x_5^2 x_6^2 \\
& + c_{10} x_2 x_3 x_4^2 x_5^2 x_6^3 + c_{11} x_1 x_3^3 x_5^3 x_6^2 + c_{12} x_3^2 x_4 x_5^3 x_6^3 .
\end{aligned}
\tag{4.29}
$$

Brill-Noether theory implies

curve	g	$\mathcal{L}$	χ	d	BN-theory		
					h^0	h^1	ρ
$C = V(P)$	3	$\mathcal{O}_{dP3}(D_L)|_C$	1	3	1	0	3
					2	1	1
					3	2	-3

(4.30)

Hence, a jump on the generic curve — a Brill-Noether jump — to $h^0(C(\mathbf{c}), \mathcal{L}(\mathbf{c})) = 2$ is possible. To explicitly construct such curves, we again inspect the long exact sequence, associated to the Koszul resolution of $\mathcal{L}(\mathbf{c})$, which is given by

$$
\begin{aligned}
&0 \longrightarrow 0 \longrightarrow 0 \longrightarrow H^0\left(D_C, \mathcal{L}\right) \\
&\hookrightarrow H^1\left(dP_3, D_L - D_C\right) \cong \mathbb{C}^3 \xrightarrow{\varphi} H^1\left(dP_3, D_L\right) \cong \mathbb{C}^2 \longrightarrow H^1\left(C, \mathcal{L}\right) \\
&\hookrightarrow 0 \longrightarrow 0 \longrightarrow 0 \longrightarrow 0 \, .
\end{aligned}
\tag{4.31}
$$

From the exactness of this sequence, we learn that $h^0(C(\mathbf{c}), \mathcal{L}(\mathbf{c})) = 3 - \mathrm{rk}\left(M_\varphi\right)$ with

$$
M_\varphi = \begin{pmatrix} c_3 & c_2 & c_1 \\ 0 & c_{12} & c_{11} \end{pmatrix}.
\tag{4.32}
$$

We set $\mathbb{P}^1_a = V(x_2)$, $\mathbb{P}^1_b = V(x_5)$. Then the possible h^0 jumps are classified as

$\mathrm{rk}(M_\varphi)$	explicit condition	curve splitting
2	$(c_3 c_{11}, c_3 c_{12}, c_2 c_{11} - c_1 c_{12}) \neq \mathbf{0}$	C^1
1	$c_3 = 0, \, c_2 c_{11} - c_1 c_{12} = 0$	C^2
1	$c_1 = c_2 = c_3 = 0$	$B_2 \cup \mathbb{P}^1_b$
1	$c_{11} = c_{12} = 0$	$\mathbb{P}^1_a \cup B_1$
0	$c_1 = c_2 = c_3 = c_{11} = c_{12} = 0$	$\mathbb{P}^1_a \cup A \cup \mathbb{P}^1_b$

(4.33)

The corresponding diagram is of the form

$$\mathcal{F}_1 : (h^0, \rho) = (1, 3)$$

$$C^1$$

$$\mathcal{F}_2 : (h^0, \rho) = (2, 1)$$

$$\mathbb{P}^1_a \cup B_2 \qquad C^2 \qquad B_1 \cup \mathbb{P}^1_b$$

$$\mathcal{F}_3 : (h^0, \rho) = (3, -3)$$

$$\mathbb{P}^1_a \cup A \cup \mathbb{P}^1_b$$

(4.34)

The change of coefficients

$$
\mathbf{c} = (1,1,1,1,1,1,1,1,1,1,1,1) \quad \rightarrow \quad \mathbf{c} = (1,1,0,1,1,1,1,1,1,1,1,1)
\tag{4.35}
$$

leads to a transition $C^1 \to C^2$ of smooth, irreducible curves. Since the topology of the curve does not change for this choice of parameters, such a transition cannot be detected from the topological data which we used for our machine learning. Therefore, such transitions are the major source of error in our decision trees.

On smooth curves C^i, the nature of the jump $C^1 \to C^2$ can be analyzed by using Serre duality:

$$
\begin{aligned}
h^1\left(C, \mathcal{O}_C\left(D_L|_C\right)\right) > 0 \quad &\Leftrightarrow \quad h^0\left(C, \mathcal{O}_C\left(K_C - D_L|_C\right)\right) > 0 \\
&\Leftrightarrow \quad K_C - D_L|_C \text{ effective} \\
&\Leftrightarrow \quad \exists p \in C \colon K_C - p \sim D_L|_C .
\end{aligned}
\tag{4.36}
$$

Hence, the origin of this jump is that K_C and the line bundle divisor differ, modulo linear equivalence, only by a point on C. Such a divisor is known as a *special divisor*. Loosely speaking, we may thus say that the origin of this one additional sections is that the points, which define the line bundle on the curve, move into a special alignment.

Note that also in this case, the diagram in eq. (4.34) encodes a hierarchy $\overline{\mathcal{F}_1} \supset \mathcal{F}_2$, $\overline{\mathcal{F}_2} \supset \mathcal{F}_3$. This is a generic feature of the parameter space and reflects a *stratification* induced by the vector-like spectrum.

4.3 h^0-stratification of the parameter space

A stratification of a topological space X is a decomposition $X = \bigcup_i \mathcal{F}_i$ into locally closed subspaces $\mathcal{F}_i$ such that

1. $\mathcal{F}_i \cap \mathcal{F}_j = \emptyset$ if $i \neq j$,

2. if $\mathcal{F}_i \cap \overline{\mathcal{F}_j} \neq \emptyset$, then $\mathcal{F}_i \subset \overline{\mathcal{F}_j}$.

Intuitively speaking, a feature associated to a subspace $\mathcal{F}_i$ — a so-called stratum — becomes "less likely" with increasing codimension of $\mathcal{F}_i$, and being contained in (the closure of) a higher dimensional stratum $\mathcal{F}_j$ implies a "specialization" of the feature when going from $\mathcal{F}_i$ to $\mathcal{F}_j$ with $j > i$. The second defining property has a convenient diagrammatic representation: let the strata $\mathcal{F}_i$ form vertices of a graph, then there is a directed edge going from j to i if $\mathcal{F}_i \subset \overline{\mathcal{F}_j}$. This is precisely the structure of the diagrams (4.12), (4.22), (4.27), and (4.34). Here, the stratified X is the parameter space $\{\mathbf{c}\}$ associated with a pair (D_C, D_L), and the strata are defined by the value of $h^0(C(\mathbf{c}), \mathcal{L}(\mathbf{c}))$ in the notation of the previous subsections. Hence, we call these diagrams h^0-*stratification*, or in short, stratification diagrams.

Note that Brill-Noether theory basically provides an analog description of the moduli space of line bundles/divisors on a smooth curve. In particular, it provides lower bounds on the dimension of the strata in terms of ρ. For F-theory models, where also deformations of the curve's topology become relevant, we see that the stratification by h^0 can be extended to the enlarged moduli space.

We observe that in this generalized setting, a stratum associated to a certain value of h^0 can consist of several disjoint subfamilies of different dimensions. In the example (4.34), the stratum $\mathcal{F}_2$ associated with $h^0 = 2$ decomposes as $\mathcal{F}_2 = \mathcal{F}_2^{(a)} \cup \mathcal{F}_2^{(s)} \cup \mathcal{F}_2^{(b)}$ with

$$
\begin{aligned}
\mathcal{F}_2^{(a)} &= \{\mathbf{c} \,|\, c_{11} = c_{12} = 0 \,, c_1 \neq 0, c_2 \neq 0, c_3 \neq 0\} \,, \\
\mathcal{F}_2^{(b)} &= \{\mathbf{c} \,|\, c_1 = c_2 = c_3 = 0, c_{11} \neq 0 \neq c_{12}\} \,, \\
\mathcal{F}_2^{(s)} &= \{\mathbf{c} \,|\, c_3 = 0 = c_2 c_{11} - c_1 c_{12}, c_1 \neq 0 \neq c_2 \,, c_{11} \neq 0 \neq c_{12}\} \,.
\end{aligned}
\tag{4.37}
$$

It is easy to see that each of these components also satisfies the axioms for strata (since they satisfy $\mathcal{F}_2^{(x)} \cap \overline{\mathcal{F}}_2^{(y)} = \emptyset$ for $x \neq y$). Furthermore, their closure contains the common stratum $\mathcal{F}_3 = \{\mathbf{c} \,|\, c_1 = \ldots = c_{12} = 0\}$ of higher codimension with $h^0 = 3$, as can be seen from the arrows connecting the three subfamilies of the stratum $\mathcal{F}_2$ to $\mathcal{F}_3$ in eq. (4.34).

In general, a stratification diagram can be roughly divided into three regions. At low values of h^0, jumps typically occur for divisor alignment, i.e., are allowed by Brill-Noether theory on a smooth curve. To get to high h^0, i.e., many vector-like pairs, the curve typically needs to factorize into many components. In the middle regime, we can have a mixture, meaning in particular that a jump occurs due to divisor alignment on a split component.

To illustrate such a "typical" case, consider

$$D_C = (5; -1, -1, -2), \qquad D_L = (1; 1, -4, 1). \tag{4.38}$$

This genus $g = 5$ curve is given by $C(\mathbf{c}) = V(P(\mathbf{c}))$ with

$$
\begin{aligned}
P := \ & c_1 x_1^3 x_2^4 x_3^2 x_4^2 + c_2 x_1^2 x_2^4 x_3 x_4^3 x_6 + c_3 x_1 x_2^4 x_4^4 x_6^2 + c_4 x_1^3 x_2^3 x_3^3 x_4 x_5 + c_5 x_1^2 x_2^3 x_3^2 x_4^2 x_5 x_6 \\
& + c_6 x_1 x_2^3 x_3 x_4^3 x_5 x_6^2 + c_7 x_2^3 x_4^4 x_5 x_6^3 + c_8 x_1^3 x_2^2 x_3^4 x_5^2 + c_9 x_1^2 x_2^2 x_3^3 x_4 x_5^2 x_6 \\
& + c_{10} x_1 x_2^2 x_3^2 x_4^2 x_5^2 x_6^2 + c_{11} x_2^2 x_3 x_4^3 x_5^2 x_6^3 + c_{12} x_1^2 x_2 x_3^4 x_5^3 x_6 + c_{13} x_1 x_2 x_3^3 x_4 x_5^3 x_6^2 \\
& + c_{14} x_2 x_3^2 x_4^2 x_5^3 x_6^3 + c_{15} x_1 x_3^4 x_5^4 x_6^2 + c_{16} x_3^3 x_4 x_5^4 x_6^3 .
\end{aligned}
\tag{4.39}
$$

From Brill-Noether theory, we then find

curve	g	$\mathcal{L}$	χ	d	BN-theory		
					h^0	h^1	ρ
$C = V(P)$	5	$\mathcal{O}_{dP3}(D_L)|_C$	0	4	0	0	5
					1	1	4
					2	2	1

$$\tag{4.40}$$

The stratification of curve geometries follows from the long exact sequence

$$
\begin{aligned}
0 \longrightarrow\ & 0 \longrightarrow 0 \longrightarrow H^0\left(C(\mathbf{c}), \mathcal{L}(\mathbf{c})\right) \\
\hookrightarrow\ & H^1\left(dP_3, D_L - D_C\right) \cong \mathbb{C}^7 \xrightarrow{\varphi} H^1\left(dP_3, D_L\right) \cong \mathbb{C}^7 \longrightarrow H^1\left(C(\mathbf{c}), \mathcal{L}(\mathbf{c})\right) \\
\hookrightarrow\ & 0 \longrightarrow 0 \longrightarrow 0 \longrightarrow 0 .
\end{aligned}
\tag{4.41}
$$

Consequently $h^0\left(C(\mathbf{c}), \mathcal{L}(\mathbf{c})\right) = 7 - \mathrm{rk}(M_\varphi)$ and we find

$$M_\varphi = \begin{pmatrix} c_{15} & c_{11} & c_7 & 0 & 0 & 0 & 0 \\ 0 & c_{10} & c_6 & c_3 & c_{11} & c_7 & 0 \\ c_{12} & c_6 & c_3 & 0 & c_7 & 0 & 0 \\ 0 & c_5 & c_2 & 0 & c_6 & c_3 & c_7 \\ c_8 & c_2 & 0 & 0 & c_3 & 0 & 0 \\ 0 & c_{14} & c_{11} & c_7 & 0 & 0 & 0 \\ 0 & c_1 & 0 & 0 & c_2 & 0 & c_3 \end{pmatrix}. \tag{4.42}$$

We list the curve strata in table 1 and display the corresponding stratification diagram in figure 3.

Of particular interest is the transition $A_3 \cup D_1 \to A_3 \cup D_2$. The former curve admits 3, the latter 4 sections. This change in the number of sections is due to a Brill-Noether jump on the curve components D_i:

curve	class	genus	d	h^0	h^1	ρ
D_i	$(5, -1, -2, -2)$	4	0	0	3	4
				1	4	0

(4.43)

Hence, provided that the line bundle divisor is chosen such that $K_{D_i} - D_L|_{D_i}$ is effective, we find an additional section on D_i, due to a Brill-Noether effect. More explicitly, in the case at hand this condition states that the line bundle divisor is linearly equivalent to the trivial divisor, i.e. $D_L|_{D_i} \sim \emptyset$. This condition is satisfied on D_2 but not on D_1. For this reason we find one additional section on $A_3 \cup D_2$.

5 Local to global section counting

In this section, we provide an in-depth analysis of the procedure of gluing local sections on reducible curves. As a result, we can place a lower bound on the number of global sections. We find sufficient topological conditions for a jump of h^0 to occur. This further allows us to formulate an algorithm to estimate the possible numbers of vector-like pairs on the moduli space of F-theory compactifications.

5.1 Gluing local sections to global sections

5.1.1 Trivial boundary conditions

Let us start by looking at a simple example. To this end, we go back to the geometry discussed in subsubsection 4.1.2, i.e.

$$D_C = (3; -1, -1, -1), \qquad D_L = (1; -1, -3, -1). \tag{5.1}$$

Recall that in this case, $C(\mathbf{c}) = V(P(\mathbf{c}))$ is a genus 1 curve defined by

$$\begin{aligned} P(\mathbf{c}) = {}& c_1 x_1^2 x_2^2 x_3 x_4 + c_2 x_1^2 x_2 x_3^2 x_5 + c_3 x_1 x_2^2 x_4^2 x_6 + c_4 x_1 x_2 x_3 x_4 x_5 x_6 \\ & + c_5 x_1 x_3^2 x_5^2 x_6 + c_6 x_2 x_4^2 x_5 x_6^2 + c_7 x_3 x_4 x_5^2 x_6^2 \,. \end{aligned} \tag{5.2}$$

$\mathrm{rk}(M_\varphi)$	explicit condition	curve splitting
7	$\det(M_\varphi) \neq 0$	C^0
6	$\det(M_\varphi) = 0$	C_1^1
5	$c_3 c_7 c_{12} = c_{15} c_3^2 + c_8 c_7^2, \quad c_{11} c_3^2 = c_3 c_6 c_7 - c_2 c_7^2$ $c_1 c_7^3 + c_{10} c_3^2 c_7 = c_{14} c_3^3 + c_3 c_5 c_7^2$	C_1^2
4	$c_3 = c_7 = 0$	$A_3 \cup D_1$
3	$c_3 = c_7 = 0 \quad c_{11} c_8 = c_{15} c_2 \quad c_{11} c_{12} = c_{15} c_6$ $c_{11} c_2 c_5 = c_{14} c_2^2 + c_1 c_{11} c_6 \quad c_{10} c_{11} c_2 = c_1 c_{11}^2 + c_{14} c_2 c_6$	$A_3 \cup D_2$
3	$c_3 = c_7 = c_8 = c_{12} = c_{15} = 0$	$A_3 \cup A_4 \cup D_3$
2	$c_2 = c_3 = c_6 = c_7 = c_{11} = 0$	$A_3^{(2)} \cup D_4$
1	$c_1 = c_2 = c_3 = c_5 = c_6 = c_7 = c_{10} = c_{11} = c_{14} = 0$	$A_3^{(3)} \cup A_5 \cup D_5$
1	$c_2 = c_3 = c_6 = c_7 = c_8 = c_{11} = c_{12} = c_{15} = 0$	$A_3^{(2)} \cup A_4 \cup D_6$
0	$c_1 = c_2 = c_3 = c_5 = c_6 = c_7 = 0$ $c_8 = c_{10} = c_{11} = c_{12} = c_{14} = c_{15} = 0$	$A_3^{(3)} \cup A_4 \cup A_5 \cup D_7$

Table 1. The curve strata for $D_C = (5; -1, -1, -2)$ and $D_L = (1; 1, -4, 1)$.

We found that for $c_1 = c_3 = c_4 = c_6 = c_7 = 0$ we have 3 global sections. Furthermore, we have already seen that for this choice of parameters, the curve has 4 components

$$C(\mathbf{c}) = E_6 \cup E_4 \cup E_2^{(2)} \cup A. \tag{5.3}$$

These components have the following properties:

curve component	equation	class	g	d	$h^0(C_i, D_L)$
A	$V(c_2 x_1 x_2 + c_5 x_5 x_6)$	$(1; 0, -1, 0)$	0	-2	0
E_4	$V(x_1)$	$(1; -1, -1, 0)$	0	-3	0
E_6	$V(x_5)$	$(1; 0, -1, -1)$	0	-3	0
$E_2^{(2)}$	$V(x_3^2)$	$(0; 0, 2, 0)$	-2	6	9

(5.4)

In the last column we give the number of sections of the restriction of the bundle $\mathcal{O}_{dP_3}(D_L)$ to these curve components. We will refer to these sections in the following as the *local sections*.

We display this geometry in figure 4. Our task is to glue the *local sections* to global sections on the curve $C = E_6 \cup E_4 \cup E_2^{(2)} \cup A$. To this end, we work out the sections explicitly and then subject them to boundary conditions at the intersection points of the different curve components.

For the components A, E_4 and E_6 we already know that the only allowed local section vanishes identically. On $E_2^{(2)}$ however, the situation is a bit more involved since $E_2^{(2)}$ is a non-reduced curve. As a set, $E_2^{(2)}$ is the locus $V(x_3)$. Using the scaling relations of dP_3,
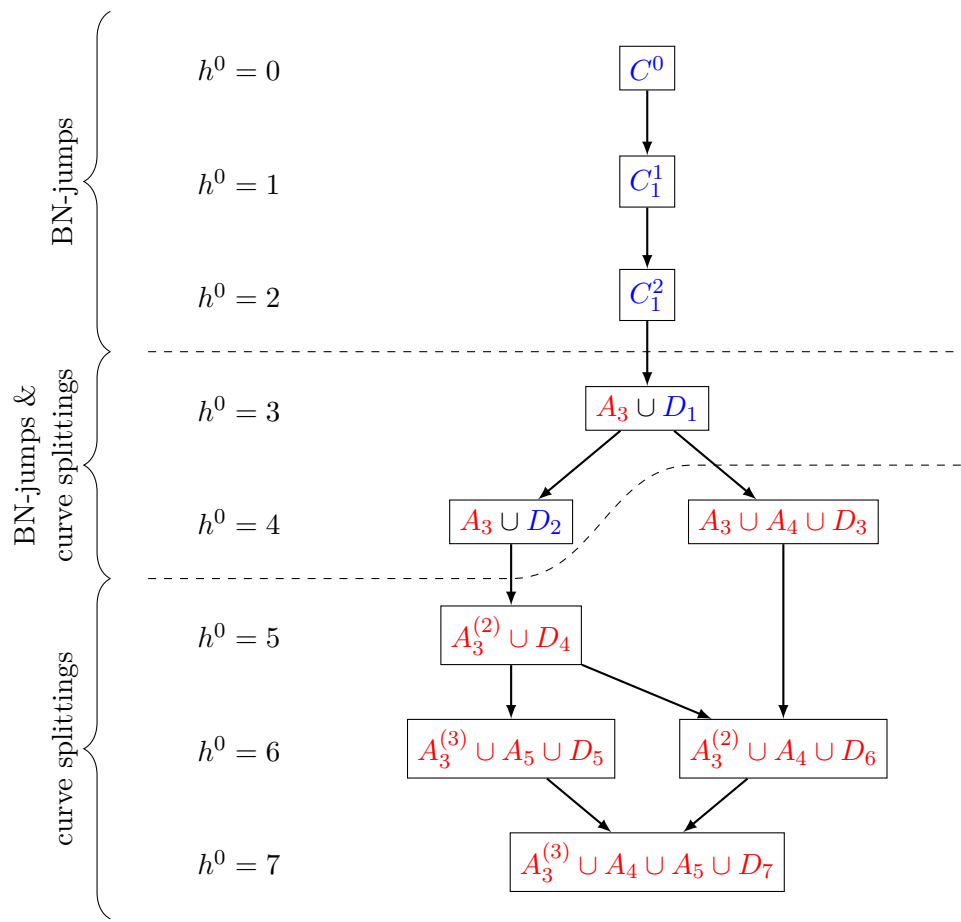
Figure 3. The stratification diagram for $D_C = (5; -1, -1, -2)$, $D_L = (1; 1, -4, 1)$.
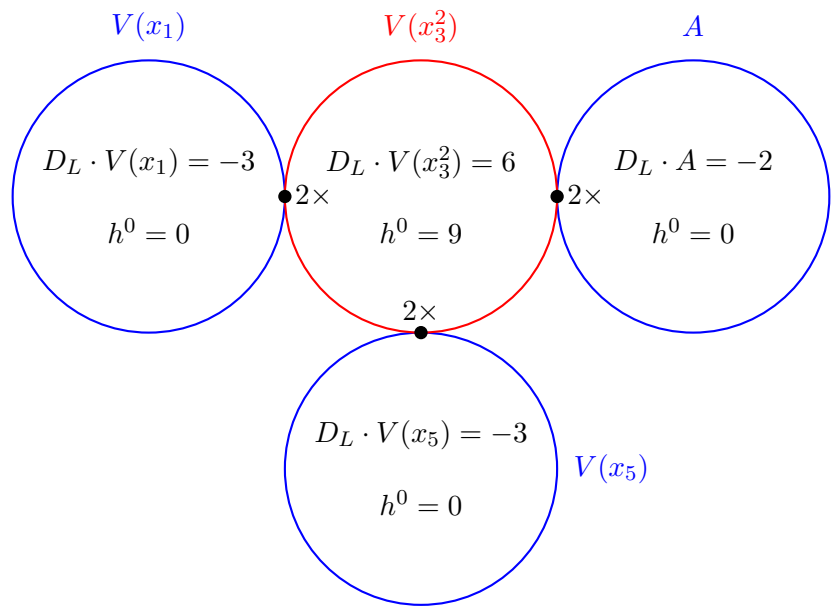

Figure 4. The 9 local sections on A lead to $9 - 3 \times 2 = 3$ global sections.

we can then set $x_2 = x_4 = x_6 = 1$ and thereby identify (x_1, x_5) as coordinates of $E_2^{(2)}$. Note, however, that since $E_2^{(2)}$ is a non-reduced curve, the polynomial x_3 is a non-trivial function on this curve component. These observations allow us to conclude

$$H^0\left(E_2^{(2)}, \mathcal{O}_{dP_3}(D_L)|_{E_2^{(2)}}\right) \cong P_3(x_1, x_5) \oplus x_3 \cdot P_4(x_1, x_5), \tag{5.5}$$

where $P_i(x_1, x_5)$ is the space of polynomials of degree i in x_1 and x_5. Upon homogenization with x_2, x_4, x_6, we can then write

$$H^0\left(E_2^{(2)}, \mathcal{O}_{dP_3}(D_L)|_{E_2^{(2)}}\right) \cong \mathrm{Span}_{\mathbb{C}}\left\{\frac{x_5^3}{x_2^3 x_4^2}, \frac{x_1 x_5^2}{x_2^2 x_4^2 x_6}, \frac{x_1^2 x_5}{x_2 x_4^2 x_6^2}, \frac{x_1^3}{x_4^2 x_6^3}\right\} \tag{5.6}$$
$$\oplus x_3 \cdot \mathrm{Span}_{\mathbb{C}}\left\{\frac{x_5^4}{x_2^4 x_4^3}, \frac{x_5^3 x_1}{x_2^3 x_4^3 x_6}, \frac{x_1^2 x_5^2}{x_2^2 x_4^3 x_6^2}, \frac{x_1^3 x_5}{x_2 x_4^3 x_6^3}, \frac{x_1^4}{x_4^3 x_6^4}\right\}.$$

From this, we learn that the only sections on $V(x_3^2)$, which vanish at $V(x_1)$, $V(x_5)$ and $V(c_2 x_1 x_2 + c_5 x_5 x_6)$, are linear combinations of the following three sections:

$$s_1 = c_5 \frac{x_1 x_5^2}{x_2^2 x_4^2 x_6} + c_2 \frac{x_1^2 x_5}{x_2 x_4^2 x_6^2} = \frac{x_1 x_5 \left(c_2 x_1 x_2 + c_5 x_5 x_6\right)}{x_2^2 x_4^2 x_6^2}, \tag{5.7}$$

$$s_2 = c_5 \frac{x_5^3 x_1}{x_2^3 x_4^3 x_6} + c_2 \frac{x_1^2 x_5^2}{x_2^2 x_4^3 x_6^2} = \frac{x_1 x_5^2 \left(c_2 x_1 x_2 + c_5 x_5 x_6\right)}{x_2^3 x_4^3 x_6^2}, \tag{5.8}$$

$$s_3 = c_5 \frac{x_1^2 x_5^2}{x_2^2 x_4^3 x_6^2} + c_2 \frac{x_1^3 x_5}{x_2 x_4^3 x_6^3} = \frac{x_1^2 x_5 \left(c_2 x_1 x_2 + c_5 x_5 x_6\right)}{x_2^2 x_4^3 x_6^3}. \tag{5.9}$$

Consequently, by extending these sections by zero outside of $V(x_3^2)$, we obtain 3 global sections.

5.1.2 Non-trivial boundary conditions

Let us consider $D_C = (3, -1, -1, -1)$ and $D_L = (5; -4, -4, 3)$. We pick special values for the parameters such that $C = V(x_1 x_2^2 x_4^2 x_6)$. The curve thus factors into four components, as displayed in figure 5. These components have the following properties:

curve	class	eq.	d	g	h^0	basis of sections
E_3	$(0; 0, 0, 1)$	$V(x_6)$	1	0	2	$\frac{x_4}{x_3^3}, \frac{x_5}{x_2 x_3^2}$
$E_5^{(2)}$	$(2; -2, 0, -2)$	$V(x_4^2)$	-2	-2	1	$\frac{x_4}{x_3^3}$
$E_1^{(2)}$	$(0; 2, 0, 0)$	$V(x_2^2)$	2	-2	5	$\frac{x_1}{x_3^2 x_6}, \frac{x_4}{x_3^3}, \frac{x_2 x_4^2}{x_3^4 x_5}, \frac{x_2 x_1 x_4}{x_3^3 x_5 x_6}, \frac{x_2 x_1^2}{x_3^2 x_5 x_6^2}$
E_4	$(1; -1, -1, 0)$	$V(x_1)$	-3	0	0	0

$$\tag{5.10}$$

We have also listed bases for the sections on the individual curve components. By starting in E_3, we see that there is a unique section which extends to $E_5^{(2)}$ and then to $E_1^{(2)}$ — this section is $\frac{x_4}{x_3^3}$. However, this section fails to vanish on $V(x_1)$. Consequently, this geometry only admits the global section which is identically zero.
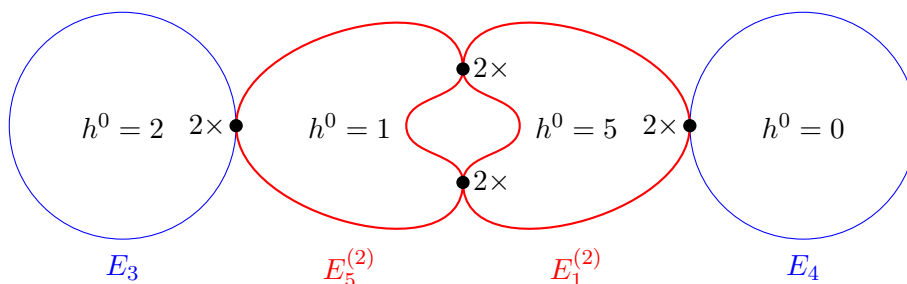
Figure 5. A non-trivial gluing example which gives no global sections.

5.1.3 From trivial to non-trivial boundary conditions

We have seen an interesting geometric transition when we discussed $D_C = (5, -1, -1, -2)$ and $D_L = (1; 1, -4, 1)$ in subsubsection 4.3. Namely, the transition

$$A_3 \cup D_1 \to A_3 \cup D_2 \tag{5.11}$$

enforces a Brill-Noether jump on D_2. Whilst D_1 only supports the trivial section, D_2 supports a one-dimensional space of non-trivial sections. As a consequence, $A_3 \cup D_2$ admits one additional section as compared to $A_3 \cup D_1$. Let us investigate this finding in more detail. We depict this geometry in figure 6 and recall the following information:

curve	class	degree	genus	h^0
A_3	$(0; 0, 1, 0)$	4	0	5
D_1	$(5, -1, -2, -2)$	0	4	0
D_2	$(5, -1, -2, -2)$	0	4	1

$$\tag{5.12}$$

To simplify our analysis, let us work with a particular class of curves D_1 and D_2, for which the transition $D_1 \to D_2$ is particularly simple:

$$
\begin{aligned}
D_1 = V \Big(& c_{12} x_1^2 x_2 x_3^3 x_5^3 x_6 + c_{13} x_1 x_2 x_3^2 x_4 x_5^3 x_6^2 + c_{16} x_3^2 x_4 x_5^4 x_6^3 + c_4 x_1^3 x_2^3 x_3^2 x_4 x_5 \\
& + c_9 x_1^2 x_2^2 x_3^2 x_4 x_5^2 x_6 + x_1^3 x_2^4 x_3 x_4^2 + x_1^3 x_2^2 x_3^3 x_5^2 + x_1^2 x_2^4 x_4^3 x_6 \\
& - x_1^2 x_2^3 x_3 x_4^2 x_5 x_6 - x_1 x_2^2 x_3 x_4^2 x_5^2 x_6^2 - x_1 x_3^3 x_5^4 x_6^2 - x_2^2 x_4^3 x_5^2 x_6^3 \\
& + x_2 x_3 x_4^2 x_5^3 x_6^3 \Big) \,,
\end{aligned}
\tag{5.13}
$$

$$D_2 = D_1 \big|_{c_{12}=0} \,.$$

Next, we turn to the sections on $A_3 \cong \mathbb{P}^1$. We note that the homogeneous coordinates are $[x_1 : x_5]$. Hence, the line bundle sections at hand are of the form $(\lambda = x_2 x_6^{-1})$:

$$H^0 \left(A_3, \mathcal{L}|_{A_3} \right) = \frac{1}{x_4^3} \cdot \mathrm{Span}_{\mathbb{C}} \left\{ x_1^4 \cdot \lambda^2, \ x_1^3 x_5 \cdot \lambda, x_1^2 x_5^2, \ x_1 x_5^3 \cdot \lambda^{-1}, \ x_5^4 \cdot \lambda^{-2} \right\} \,. \tag{5.14}$$

At $x_3 = 0$, we may set $x_2 = x_4 = x_6 = 1$ by the scaling relations of dP_3. In terms of these inhomogeneous coordinates, we find

$$A_3 \cap D_i = V(x_3, x_1 - x_5) \cup V(x_3, x_1 + x_5) \,. \tag{5.15}$$
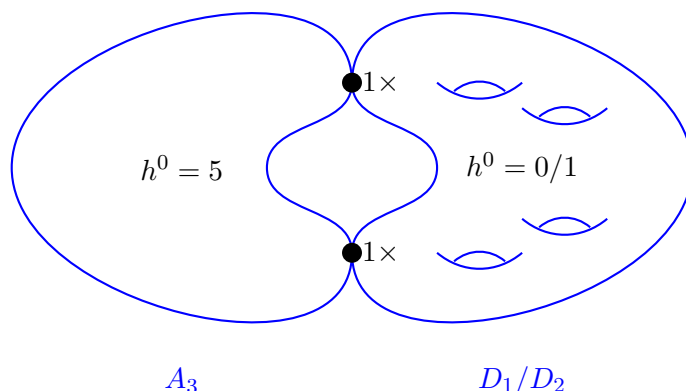
Figure 6. A Brill-Noether jump $D_1 \to D_2$ generates one additional global section.

That all said, we can discuss the global sections on $A_3 \cup D_1$ and $A_3 \cup D_2$:

- On D_1, the only supported section vanishes identically. Hence, we may only consider sections on A_3, which vanish at $A_3 \cap D_1$. It is not too hard to see that the space of these sections is generated by

$$s_1 = -x_1^4 + x_5^4, \qquad s_2 = -x_1^3 x_5 + x_1 x_5^3, \qquad s_3 = -x_1^4 + x_1^2 x_5^2. \tag{5.16}$$

- On D_2 however, the line bundle divisor is special. In fact, since it is a divisor of degree zero, this divisor must be the trivial divisor. Consequently, the sections on D_2 are the constant ones. It is not too hard to see that the sections on A_3, which have value 1 at the intersection points $A_3 \cap D_2$, are generated by

$$t_1 = x_1^4, \qquad t_2 = t_1 + s_1, \qquad t_3 = t_1 + s_2, \qquad t_4 = t_1 + s_3. \tag{5.17}$$

This explains the one additional section on $A_3 \cap D_2$ as opposed to $A_3 \cap D_1$.

5.1.4 Overcounting boundary conditions

As a final example, let us look at $D_C = (4; -1, -1, -1)$ and $D_L = (1, 1, -3, 0)$. Let us deform the curve C such that it is given by

$$P = x_1 \cdot Q, \qquad Q = x_1^2 x_2^2 x_3^3 x_5 + x_2^3 x_4^3 x_6^3 + x_3^3 x_5^3 x_6^2. \tag{5.18}$$

We display this curve geometry in figure 7. The two curve components have the following properties:

component	equation	class	g	d	h^0	
C_1	$V(x_1)$	$(1; -1, -1, 0)$	0	-1	0	(5.19)
C_2	$V(Q)$	$(3; 0, 0, -1)$	1	3	3	
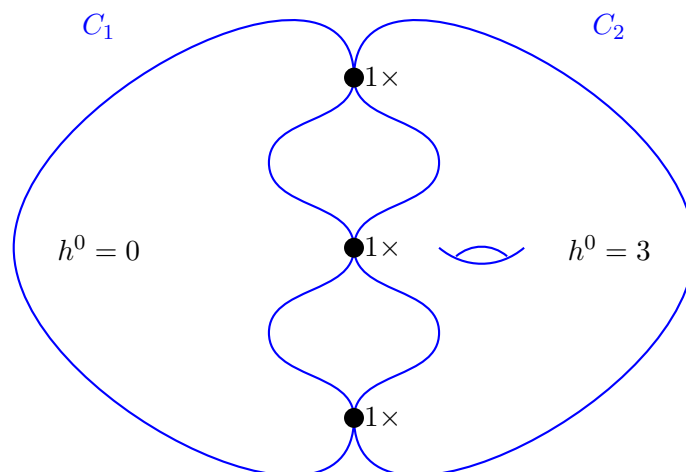
Figure 7. Naively, we expect $3 - 3 = 0$ global sections. However, one section on C_2 automatically vanishes at $C_1 \cap C_2$, leading to $h^0 (C_1 \cup C_2, \mathcal{L}) = 1$.

Up to canonical isomorphism (induced from the connection homomorphism), we find a basis of the sections on C_2 as

$$\mathcal{B} = \left\{ \frac{1}{x_2 x_3^3 x_4^2 x_6}, \frac{x_5}{x_2^2 x_3^2 x_4^3 x_6}, \frac{x_1}{x_2 x_3^2 x_4^3 x_6^2} \right\} . \tag{5.20}$$

From this we can see that the third section automatically vanishes at the intersection $C_1 \cap C_2$, whilst the other two sections do not vanish there. Consequently, and in agreement with the computational results by *gap*, we find $h^0 (C_1 \cup C_2, \mathcal{L}) = 1$.

Importantly, a naive guess cannot predict this number. In this case, we would have counted as follows: 3 sections on C_2 subject to vanishing conditions at the 3 intersection points $C_1 \cap C_2$ should leave us only with the trivial section. Hence, in this example, a naive counting fails. Such phenomena were originally studied more generally in [51, 52] — see also [53] for a more modern exposition of the material.

5.2 Sufficient jump condition and algorithmic section estimate

As demonstrated in the previous section, gluing local sections to global sections is a non-trivial task. The exact details depend, among other things, on the relative position of the line bundle divisor and the intersection points of the curve components: the results change when some of these intersection points coincide and when the bundle divisor is special on some curve components.

In the following, we will propose a counting mechanism with the following key properties:

- It relies mostly on topological data.

- It provides a lower bound on the number of global sections.

Of course, such a simplified counting procedure will fail to predict intricate geometries as discussed in [51–53]. Still, it has two distinct advantages. First, since it relies mostly on topological data, it is very fast. Given a curve C and a line bundle $\mathcal{L}$ on C, we can apply the strategy to place a lower bound on $h^0\left(C(\mathbf{c}), \mathcal{L}(\mathbf{c})\right)$ for many different choices of parameters $\mathbf{c}$ of C. The collection of these lower bounds can then serve as an estimate of the vector-like spectrum of $(C, \mathcal{L})$ over the parameter space. Note that obtaining such an estimate is unfeasible with existing exact algorithms, e.g., those implemented in [42], since these algorithms require extensive computational resources and often take a long time to finish. The second advantage results from the fact that our counting procedure systematically underestimates the actual number of global sections. Therefore, it allows us to formulate sufficient conditions for a jump in the vector-like spectrum to happen.

5.2.1 Counting procedure

Let us consider a curve C with

$$C = \bigcup_{i=1}^{N} C_i \,, \tag{5.21}$$

i.e., C has N components C_i. For our counting procedure to be as simple and reliable as possible, let us avoid setups of the type discussed in subsubsection 5.1.2 and subsubsection 5.1.3. Hence, for simplicity, let us consider a line bundle $\mathcal{L}$ on C such that neighboring curve components do not support non-trivial sections simultaneously.[10] Put different, let us focus on setups where for all curve components C_i the following holds true:

$$h^0\left(C_i, \mathcal{L}|_{C_i}\right) > 0 \quad \Rightarrow \quad h^0\left(C_j, \mathcal{L}|_{C_j}\right) = 0 \quad \forall \, C_j \text{ with } C_i \cap C_j \neq \emptyset \,. \tag{5.22}$$

Let us denote by b_i the number of intersection points of C_i with the other curve components. Generically, we then impose b_i conditions on the "local" sections in $H^0(C_i, \mathcal{L}|_{C_i})$. Consequently,

$$n_i(C_i) = \begin{cases} h^0(C_i, \mathcal{L}|_{C_i}) - b_i & \text{if } h^0(C_i, \mathcal{L}|_{C_i}) \geq b_i \\ 0 & \text{else} \end{cases} \tag{5.23}$$

is a lower bound to the number of sections on C_i which satisfy the gluing boundary conditions. The sum of these contributions over all curve components places a lower bound on $h^0(C, \mathcal{L})$:

$$\sum_{i=1}^{N} n_i(C_i) \leq h^0(C, \mathcal{L}) \,. \tag{5.24}$$

We expect that equality holds in generic situations and that only fairly tuned geometries, in the spirit of [51–53], will lead to a proper inequality.

---

[10]Let us emphasize that this is a simplifying assumption. We will eventually drop it.

As simple demonstration, let us apply this procedure to the geometry in subsubsection 5.1.1:

component C_i	$h^0(C_i, \mathcal{L}|_{C_i})$	b_i	n_i
$V(x_1)$	0	2	0
$V(x_3^2)$	9	6	3
$V(x_5)$	0	2	0
A	0	2	0

$$(5.25)$$

Indeed, $\sum_{i=1}^{3} n_i = 3$ in agreement with our discussion in subsubsection 4.1.2. However, if we apply this counting to $A_3 \cup D_2$, as discussed in subsubsection 5.1.3, then we find the inequality

$$n_1 + n_2 = (5 - 2) + 0 = 3 < 4 = h^0(A_3 \cup D_2, \mathcal{L}). \tag{5.26}$$

Likewise, for the geometry in subsubsection 5.1.4, we find $n_1 + n_2 = (3 - 3) + 0 = 0 < 1 = h^0(C_1 \cup C_2, \mathcal{L})$. In both cases, special alignments of the line bundle divisor/the intersection points of the curve components happen. These intricate phenomena are not captured by our counting procedure, so that it can only provide a lower bound.

5.2.2 Accuracy on our database

Let us now apply this counting procedure to our database [43] to obtain an estimate of how often the inequality is satisfied. To this end, we need to identify the number of local sections, which can be challenging for complicated curve geometries and could call for an application of, e.g., the exact methods implemented in [42]. However, given the vast number of curve components in our database, we find it more appealing to focus on those curves for which we can identify the number of local sections quicker. To this end, we focus on the following two types of curves:

- Smooth curves:
  We consider the line bundle degree $d = \deg(\mathcal{L}|_{C_i})$. Provided that $d < 0$, we know that $\mathcal{L}|_{C_i}$ does not admit non-trivial sections. Conversely, if $d > 2g(C_i) - 2$, then it follows from application of the Kodaira vanishing theorem, that $h^0(C_i, \mathcal{L}|_{C_i}) = d - g + 1$. If none of these conditions is satisfied, we discard the curve for this test.

- Non-split curves:
  For these curves, we can simply read off the number of local sections from our database.

Based on these local section counts, we have then applied the counting procedure presented in subsubsection 5.2.1. Recall that a large number of curves in our database do neither consist of smooth curve components nor are non-split. For simplicity, let us furthermore subject the curve geometry to the condition that neighboring components do not support non-trivial sections simultaneously. We find that we can apply this restrictive counting procedure to roughly 60% of the cases in our database. For these, we predict the correct number of global sections with an accuracy of more than 99%, i.e. our counting procedure works remarkably well. We list the detailed results in subsubsection B.2.1.

5.2.3 Sufficient conditions for jumps in cohomology

These insights of gluing local sections to form global sections, imply sufficient conditions for jumps in cohomology. First, we have the following

Lemma 1. *Let S be a smooth surface, $L \in \mathrm{Pic}(S)$ a line bundle, and $|C|$ a linear system of curves on S. Consider a special member $C_1 \cup C_2$ such that the curves C_1, C_2 meeting transversely in $C_1 \cdot C_2 > 0$ distinct points. Let $N_1 = h^0(C_1, \mathcal{L}|_{C_1})$ and $N_2 = h^0(C_2, \mathcal{L}|_{C_2})$. Then*

$$h^0\left(C_1 \cup C_2, L|_{C_1 \cup C_2}\right) \geq N_1 + N_2 - C_1 \cdot C_2 . \tag{5.27}$$

Proof. We consider the short exact sequence $0 \to L|_{C_1 \cup C_2} \to L|_{C_1 \sqcup C_2} \to L|_{C_1 \cap C_2} \to 0$. The associated long exact sequence in sheaf cohomology begins with

$$0 \to h^0\left(C_1 \cup C_2, L|_{C_1 \cup C_2}\right) \to h^0\left(C_1 \sqcup C_2, L|_{C_1 \sqcup C_2}\right) \to h^0\left(C_1 \cap C_2, L|_{C_1 \cap C_2}\right) \to \ldots \tag{5.28}$$

Now, since $h^0(C_1 \sqcup C_2, L|_{C_1 \sqcup C_2}) = N_1 + N_2$ and $h^0(C_1 \cap C_2, L|_{C_1 \cap C_2}) = C_1 \cdot C_2$, the statement follows. ∎

   We can use this result, together with the insights on gluing local sections to global sections, to derive the following

Corollary 1. *Let S be a smooth surface, $L \in \mathrm{Pic}(S)$ a line bundle, and $|C|$ a linear system of curves on S with smooth general member C and special member $C_1 \cup C_2$ where C_1, C_2 are smooth curves of genera g_1, g_2 meeting transversely in $C_1 \cdot C_2 > 0$ distinct points. We assume $h^1(C, L|_C) = 0$, $\deg\left(L|_{C_2}\right) > 2g_2 - 2$ and $\deg\left(L|_{C_1}\right) < \min\{0, g_1 - 1\}$. Then*

$$h^0\left(C_1 \cup C_2, L|_{C_1 \cup C_2}\right) - h^0\left(C, L|_C\right) \geq g_1 - 1 - \deg\left(L|_{C_1}\right) . \tag{5.29}$$

Proof. Since $\deg\left(L|_{C_1}\right) < 0$, there are no sections on C_1. Hence, from Lemma 1 we obtain the inequality

$$h^0\left(C_1 \cup C_2, L|_{C_1 \cup C_2}\right) \geq h^0\left(C_2, L|_{C_2}\right) - C_1 \cdot C_2 . \tag{5.30}$$

Note that $g_C = g_1 + g_2 + C_1 \cdot C_2 - 1$. Consequently, since $\deg\left(L|_{C_2}\right) > 2g_2 - 2$, we can write

$$
\begin{aligned}
h^0(C_2, L|_{C_2}) &= \deg\left(L|_{C_2}\right) - g_2 + 1 \\
&= \deg\left(L|_{C_2}\right) - (g_C - g_1 - C_1 \cdot C_2 + 1) + 1 \\
&= (\deg\left(L|_C\right) - g_C + 1) + C_1 \cdot C_2 + g_1 - 1 - \deg\left(L|_{C_1}\right) \\
&= h^0(C, L|_C) + C_1 \cdot C_2 + g_1 - 1 - \deg\left(L|_{C_1}\right) .
\end{aligned}
\tag{5.31}
$$

Hence, we conclude

$$
\begin{aligned}
h^0(C_1 \cup C_2, L|_{C_1 \cup C_2}) &\geq h^0\left(C_2, L|_{C_2}\right) - C_1 \cdot C_2 \\
&= h^0(C, L|_C) + g_1 - 1 - \deg\left(L|_{C_1}\right) , \qquad (5.32)
\end{aligned}
$$

$$\Leftrightarrow \quad h^0(C_1 \cup C_2, L|_{C_1 \cup C_2}) - h^0(C, L|_C) \geq g_1 - 1 - \deg\left(L|_{C_1}\right) .$$

Finally, since we assume $\deg\left(L|_{C_1}\right) < \min\{0, g_1 - 1\}$, the number of additional sections on $C_1 \cup C_2$ is bounded from below by the positive integer $g_1 - \deg\left(L|_{C_1}\right) - 1$. ∎

We expect that equality holds in generic situations and that only special setups in the spirit of [51, 52] lead to a proper inequality. Still, our result is powerful enough to give a sufficient condition for a jump. Let us demonstrate this in the geometries discussed for the F-theory toy model in subsection 3.1. There, we looked at $S = dP_3$ and considered

$$D_C = (10; -3, -3, -4), \qquad D_L = (5; -4, -4, 3). \tag{5.33}$$

We found that on the generic $g = 24$ curve C it holds $h^0(C, \mathcal{L}|_C) = 15$. We then considered the splitting $C \to C_1 \cup C_2$ where $C_1 = V(x_6)$. The curve components have the following properties:

curve	class	degree	genus	h^0
C_1	(0;0,0,1)	-3	0	0
C_2	(10;-3,-3,-5)	41	20	22

(5.34)

From this we see that Corollary 1 applies to this geometry and implies

$$h^0\left(C_1 \cup C_2, \mathcal{L}\right) - h^0\left(C, \mathcal{L}\right) \geq g_1 - 1 - \deg\left(\mathcal{L}|_{C_1}\right) = 0 - 1 - (-3) = 2. \tag{5.35}$$

This is in agreement with our discussion in subsection 3.1.

In many string theory constructions, it is important to engineer exactly one additional vector-like pair. This is particularly true when generating exactly one Higgs pair in MSSM constructions. It is intuitive, that such a minimal change in the vector-like spectrum, requires only mild changes in the geometry. As long as Corollary 1 applies, a necessary condition for such a mild change is to merely split off either a $\mathbb{P}^1$ or a torus — $g_1 \geq 2$ implies $h^0\left(C_1 \cup C_2, \mathcal{L}\right) - h^0\left(C, \mathcal{L}\right) \geq 2$.

More generally, it is of interest to identify the allowed numbers of global sections on a given curve. Therefore, we will now describe an estimate for these values, which is based on the counting procedure presented in subsubsection 5.2.1, Lemma 1 and Corollary 1.

5.2.4 Algorithmic spectrum estimates

We can use our results to formulate an algorithmic estimate for the vector-like spectrum over the parameter space of a given setup (D_C, D_L) in a global model. For the time being, our algorithm is focused on the case of a curve in dP_3 defined by $\{P = 0\}$ and pullback line bundles on these curves. We have implemented this algorithm in the package *H0Approximator* [44] as part of [42]. Our algorithm proceeds as follows:

1. Input: curve class D_C and line bundle class D_L

2. Identify all combinations of toric $\mathbb{P}^1$s that can be split off from the curve D_C.

3. Identify the *generic* number of sections of D_L on each curve component.

4. Use the counting procedure presented in subsubsection 5.2.1 as well as Lemma 1 and Corollary 1 to place a lower bound on the number of global sections.

$\Rightarrow$ The collection of all these global section estimates forms an estimate for h^0 of D_L on the parameter space of the curve D_C.

Let us emphasize a couple of important points of this counting procedure. First, in the second step we do not apply exact methods, such as [42], to find the exact number of local sections. Rather, we identify the generic number of sections, by which we mean $h^0(C, \mathcal{L}) = \chi(\mathcal{L})$ if $\chi(\mathcal{L}) \geq 0$ and $h^0(C, \mathcal{L}) = 0$ otherwise. The advantage of this is, that the chiral index can be obtained from topology only. Hence, the number of global sections can be estimated very quickly. Furthermore, this strategy does not violate our *lower bound philosophy*, since the generic number of sections is never larger than the actual number of sections. Consequently, this strategy allows us to quickly identify a lower bound to the actual number of global sections.

Secondly, let us point out that one disadvantage of our approach of *generic local sections* is that we are unable to identify Brill-Noether jumps on the curve components in this way. However, since such a quick spectrum estimate over the entire parameter space of the curve is currently unfeasible or impossible to obtain with the fully accurate methods, we accept this minor drawback.

Finally, note that upon splitting off $\mathbb{P}^1$s from the curve, the curve could (accidentally) factor further. Computing these further factorizations requires a primary ideal decomposition of the corresponding principal ideal. Currently, this is the most time consuming operation in our algorithm. We reserve optimizations for future work.

This algorithm correctly predicts all the possible values of h^0 for 67 of the 83 pairs (D_C, D_L) in our database [43]. Only for one pair (D_C, D_L), our prediction misses more than 2 values of the exact spectrum. Given the simplicity of our approximation, which means that we cannot detect intricate Brill-Noether jumps and effects discussed in [51, 52], we consider this a very positive result. We list the details in subsubsection B.2.2.

Let us complete this section by applying our procedure to estimate the vector-like spectrum of the F-theory setup discussed in section 3. Recall that in this case we are looking at $D_C = (10; -3, -3, -4)$, i.e., a complicated genus 24 curve. The line bundle in this case is $D_L = (5; -4, -4, 3)$. Even though this geometry is fairly involved, our approximator can estimate the spectrum in a couple of minutes:[11]

```
gap> LoadPackage( "H0Approximator" );
true
gap> FineApproximation( [10,-3,-3,-4],[5,-4,-4,3] );
(*) 56 rough approximations
(*) Rough spectrum estimate: [ 15, 17, 18, 19, 20, 21 ]
    (x) h0 = 15: 9
    (x) h0 = 17: 18
    (x) h0 = 18: 4
    (x) h0 = 19: 9
    (x) h0 = 20: 12
    (x) h0 = 21: 4
```

---

[11]In this case, this long run time is mostly attributed to the primary decomposition, which we perform to check irreducibility of the curve components.

```
(*) Checking irreducibility of curves...
(*) 26 fine approximations
(*) Fine spectrum estimate: [ 15, 17, 18, 19, 20, 21 ]
     (x) h0 = 15: 3
     (x) h0 = 17: 6
     (x) h0 = 18: 4
     (x) h0 = 19: 3
     (x) h0 = 20: 6
     (x) h0 = 21: 4
[ 15, 17, 18, 19, 20, 21 ]
```

Hence, we have identified 26 curve splittings into irreducible components, for which our counting procedure can estimate the spectrum. Based on this, we expect $h^0 \in \{15, 17, 18, 19, 20, 21\}$. As we know from our analysis in section 3, indeed $15 \leq h^0 \leq 21$ and $h^0 = 16$ is only possible by a Brill-Noether jump. The latter cannot be predicted by this method. More information on this implementation can be found in [42].

6 Conclusion and outlook

Motivated by a better understanding of the exact massless spectra of 4d F-theory compactifications, we have analyzed in this work families of curves $C(\mathbf{c})$ in a complex surface and line bundles $\mathcal{L}(\mathbf{c})$ on these. Our focus has been on the interplay between changes in the cohomology $h^0(C(\mathbf{c}), \mathcal{L})$ and variations of the parameters $\mathbf{c}$, which play the role of complex structure moduli in the context of global F-theory models. To gain insights on how these two are related, we have used two approaches.

To begin with, we first used ideas from Big Data and machine learning to gain some intuitions, based on computationally simpler examples, under what circumstances the cohomology may jump, leading to additional vector-like pairs in the F-theory interpretation. To this end we have generated, in section 2, a database [43] of cohomologies for pairs $(C(\mathbf{c}), \mathcal{L}(\mathbf{c}))$ by varying the parameters $\mathbf{c}$, where the curves are of genus $1 \leq g \leq 6$, and the line bundles were pullback bundles from a dP_3 surface. For these less complex examples, the cohomologies can be computed using the computer implementations in [42]. We then use supervised learning on decision trees to predict jumps in the value of h^0. Using different features for training, we find that, while not performing perfectly, topological criteria are surprisingly well-suited (reaching about 95% accuracy) for distinguishing cases with generic vs. enhanced h^0. In particular, the algorithm learns from the data a strong correlation between jumps and curves $C(\mathbf{c})$ which split into various components. This intuition can be applied, without any detailed understanding of the origin of the jumps, directly to find complex structure tunings targeted at generating additional vector-like pairs in F-theory model building. We demonstrate this in section 3 with an F-theory toy model containing a curve of genus 24, for which a scan over the relevant parameter space would be computationally infeasible. Nevertheless, we found that we can use curve splittings alone to easily engineer 2 to 5 additional vector-like pairs. This highlights the effectiveness of the machine learning approach to learn certain features from simpler examples, and without any previ-

ous knowledge. However, we also saw there that by curve splitting alone, a spectrum with just one vector-like pair is impossible to achieve.

To overcome this obstacle, we have employed well-known techniques in algebraic geometry, such as the Koszul resolution and Čech cohomology, which also helps to explain our findings from the machine learning approach in more detail. We conclude that deformations of the parameters $\mathbf{c}$ leading to a jump in cohomology can be largely classified as either the curve $C(\mathbf{c})$ or the line bundle $\mathcal{L}(\mathbf{c})$ becoming non-generic. While the former comes from curve splittings and is thus topological,[12] the latter is due to special alignments of the points on $C(\mathbf{c})$ defining $\mathcal{L}(\mathbf{c})$, and not visible just from topological criteria. The fact that the learner performed so well with the topological criteria is due to a bias in the dataset, which contains only a small number of instances with non-generic line bundles. Such jumps can never be predicted by the learner based just on split type and intersection numbers. However, as we discussed in section 4, we find in general "equally likely" jumps due to non-generic line bundles. The likeliness can be quantified by comparing the dimension of the corresponding subspace of the parameter space on which the jumps occur, which for non-generic line bundles is the subject of *Brill-Noether* theory. This is generalized in the F-theoretic setup, where complex structure deformations affect genericity of the curve and line bundle democratically. This leads to a *stratification* of the parameter space by the values of h^0. That is, the complex structure moduli space of global F-theory models decomposes into disjoint subspaces labelled by the vector-like spectrum. The relationship between the strata can be represented by a Hasse-type diagram, which we term h^0-*stratification diagrams*.

The connection between decision trees, which are also Hasse diagrams, and the stratification diagrams is rather intriguing. While they bear some resemblance with decision trees, a key difference is that, unlike in decision trees, nodes can have more than one *incoming* edge. It would be interesting to investigate whether other graph-based machine learning techniques, such as Graph NNs, can be used to train algorithms that can predict the presence of jumps more accurately than the decision trees. Furthermore, recall that global F-theory models typically contain more than one matter curve. The complex structures of these curves are determined by the global moduli of the elliptic fibration, and it is in general not possible to tune the complex structures of all of these curves independently. Therefore, it would be important to extend our analysis to a simultaneous h^0-stratification of the moduli space by all the matter curves in a global F-theory model.

In section 5, we have then investigated the "microscopic" origins of jumps due to curve-splittings. It follows a simple counting procedure of local sections on individual curve components, which we then glue to global contributions to h^0 on the whole curve. Depending on the boundary conditions imposed by the intersection patterns of the components, this can lead to a net-increase of global sections on the reducible curve compared to the generic case. We have used this understanding to formulate sufficient conditions for a jump in the vector-like spectrum to occur as a result of a curve splitting. These criteria are purely topological, and combine the gluing arguments with vanishing theorems on

---

[12]More generally, a curve can also remain smooth while being non-generic, e.g., if it becomes hyperelliptic. Such transitions are of non-topological nature, and therefore more subtle to detect. We have neglected them for simplicity in our discussions.

individual components. Let us stress that this in general provides only a lower bound for h^0 for the split curve, because it does not take into account alignments of the intersection points of the components and divisors on the individual components. It will be interesting to investigate, if these bounds can be further improved by topological considerations.

Despite these simplifications, we found these criteria extremely useful to provide a rough estimate of the possible spectrum of h^0 on the moduli space of F-theory compactifications, and implemented the algorithm in [44]. To fully appreciate this implementation, let us mention that to the best knowledge of the authors, the exact algorithms implemented in [42, 54, 55] do not allow for a parametric cohomology computation. Rather, they will focus on one particular point in the complex structure moduli space and provide the exact answer at this very point. Since each of these computations requires huge amounts of computational resources and runtime, it is impractical to repeat such computations for many points in the complex structure moduli space. In contrast, the new algorithm yields an approximate, but oftentimes sufficiently accurate, estimate — even for complicated examples such as the genus 24 curve discussed in section 3 — within minutes. We leave generalizations of this counting algorithm, as well as extensions to other toric surfaces, for future work.

Another limitation of our approach is that we have only considered pullback line bundles so far. However, as already alluded to in the introduction, vector-like spectra in F-theory are oftentimes encoded in line bundles described by a formal weighted sum of points. Such a description is computationally harder for two main reasons. First, it takes much longer to compute line bundle cohomologies of non-pullback bundles with the technologies of [42]. This makes it more challenging to generate a sufficiently large database to apply ideas from Big Data and machine learning. The second obstacle is the parametrization of the line bundles. Namely, distinct point configuration can encode equivalent line bundles if their difference is the divisor of a meromorphic function. To have a better handle on tracking how these equivalences change with complex structure deformations, we need a better understanding of meromorphic functions on higher genus curves. The crucial tool in this direction is the *Abel-Jacobi map*, which also plays a similar role in the hyperelliptic curve cryptography. It would be interesting to see to what extent machine learning ideas can be beneficial here.

A related issue arises for *fractional bundles* or *root bundles*. These appear frequently in explicit global F-theory constructions that engineer a three-generation Standard-Model-like particle physics sector [16, 21, 23–25]. The constraint to have chiral indices with $|\chi| = 3$ in these models lead to line bundles $\mathcal{L}$ on curves C which satisfy $\mathcal{L}^{\otimes n} = L|_C$, where L is a line bundle on the base $\mathcal{B}_3$ of the elliptic fibration. In case $n = 2$ and $L|_C = K_C$ is the canonical bundle of the curve, the bundle $\mathcal{L}$ can be understood as a spin bundle on C. However, for general F-theory constructions, also 3rd and higher roots of bundles $\mathcal{L} \neq K_C$ appear. An understanding of which line bundles $\mathcal{L}$ on C satisfy such an equation again requires a detailed understanding of which points — in this case the intersection points of C with the divisor on $\mathcal{B}_3$ dual to L — on the curve define equivalent divisors. We expect that this will also be intimately related to satisfying the quantization condition [56] for the gauge flux background.

Finally, it is important to point out that the complex structure parameters of the elliptic fibration are not the only parameters of the physical theory. Rather, a large part of this parameter space, which we have not touched upon, is in the parametrization of all possible gauge backgrounds. This includes in particular backgrounds with so-called *non-vertical G_4-flux* [57, 58], for which explicit construction methods in global models are largely unknown. While these typically do not contribute to the chiral index, it is not clear at the moment if they could modify the flux-induced line bundles on the matter curves. However, since non-vertical fluxes contribute prominently to a superpotential for the moduli, their presence will dynamically select points in the moduli space that can be a vacuum for the theory, thus have a very different, but direct influence on the vector-like spectrum. We will therefore need a much better handle on these gauge backgrounds first before we can develop a full understanding for the space of 4d F-theory vacua.

Acknowledgments

A Tools: Koszul resolution, Brill-Noether theory and fat points

The purpose of this appendix is to cover some of the necessary mathematical backgrounds and to provide more details on computations carried out throughout the paper.

A.1 Brill-Noether theory

Our exposition of Brill-Noether theory is based on [48, 49]. We refer the interested reader to these references for more details.

A.1.1 The Jacobian of Riemann surfaces

To each smooth Riemann surface C_g one can associate a Jacobian variety $\mathrm{Jac}(C_g)$. This variety is of dimension g and classifies equivalence classes of line bundle divisors of degree 0:

$$\mathrm{Jac}(C_g) = \mathrm{Div}_0(C_g)/\mathrm{Prin}(C_g)\,. \tag{A.1}$$

In this expression $\mathrm{Div}^0(C_g)$ is the group of all divisors of degree 0 and $\mathrm{Prin}(C_g)$ the group of all principal divisors on C_g. Line bundles on C_g are isomorphic iff their divisors differ by

a divisor in $\mathrm{Prin}(C_g)$. Hence, sheaf cohomologies of line bundles can only differ if the line bundles are not isomorphic, or equivalently if their divisors differ by more than elements of $\mathrm{Prin}(C_g)$. Consequently, the Jacobian of C_g plays an important role for our analysis and in Brill-Noether theory. Let us therefore introduce the Jacobian in more detail.

Historically, the Jacobian of a curve C_g of genus g was discovered by investigating integrals $\int_{\mathcal{P}} \omega$ where $\mathcal{P} \subset C_g$ is a (not necessarily closed) path and ω a holomorphic differential. More generally, mark a point $p_0 \in C_g$, let $(\omega_1, \ldots, \omega_g)$ be a basis of the holomorphic differentials on C_g and consider the map

$$\phi \colon C_g \to \mathbb{C}^g \, , \, p \mapsto \left(\int_{p_0}^{p} \omega_1 \, , \, \ldots \, , \, \int_{p_0}^{p} \omega_g \right) . \tag{A.2}$$

The value of this map strongly depends on the path $\mathcal{P} \subset C_g$ which we choose to connect p_0 and p. This redundancy can be removed by taking the period lattice of C_g into account. To this end, recall that there are $2g$ homologically distinct closed 1-cycles in C_g, i.e., $H_1(C_g, \mathbb{Z})$ is a $2g$-dimensional vector space over $\mathbb{Z}$.[13] We now consider the map

$$\phi \colon H_1(C_g, \mathbb{Z}) \to \mathbb{C}^g \, , \, \alpha \mapsto \left(\int_{\alpha} \omega_1 \, , \, \ldots \, , \, \int_{\alpha} \omega_g \right) , \tag{A.3}$$

where ω_i denote the above basis of holomorphic differentials on C_g. Hence, for every of the $2g$-basis elements of $H_1(C_g, \mathbb{Z})$, we obtain an element $\phi(\alpha) \in \mathbb{C}^g$. It turns out that these $2g$ elements span a full-dimensional lattice Λ in $\mathbb{C}^g$ — the period lattice of C_g. By virtue of this lattice, we obtain a well-defined map

$$\phi \colon C_g \to \mathbb{C}^g/\Lambda \, , \, p \mapsto \left(\int_{p_0}^{p} \omega_1 \, , \, \ldots \, , \, \int_{p_0}^{p} \omega_g \right) . \tag{A.4}$$

This map is known as the *Abel-Jacobi map*. It can easily be extended to divisors in C_g. Namely, for a divisor

$$D = \sum_{i=1}^{N} \lambda_i \cdot p_i \, , \, \lambda_i \in \mathbb{Z}, \qquad p_i \in C_g \, , \tag{A.5}$$

we define

$$\phi \colon \mathrm{Div}(C_g) \to \mathbb{C}^g/\Lambda \, , \, D \mapsto \sum_{i=1}^{N} \lambda_i \cdot \phi(p_i) . \tag{A.6}$$

The theorem of Abel (see [59] and references therein) states that two effective divisors D and E satisfy $\phi(D) = \phi(E)$ iff D and E are linearly equivalent. Consequently, we obtain an injective group homomorphism

$$\Phi \colon \mathrm{Div}_0(C_g)/\mathrm{Prin}(C_g) \to \mathbb{C}^g/\Lambda \, , \, [D] \mapsto \sum_{i=1}^{N} \lambda_i \cdot \phi(p_i) , \tag{A.7}$$

of divisor classes of degree 0. It turns out that this map is also surjective (see [59] for a proof). Hence, there is a natural isomorphism

$$\mathrm{Jac}(C_g) = \mathrm{Div}_0(C_g)/\mathrm{Prin}(C_g) \cong \mathbb{C}^g/\Lambda \, . \tag{A.8}$$

---

[13]See e.g. [48] for an explicit construction of the $2g$-generators A_i, B_i of $H_1(C_g, \mathbb{Z})$.

A.1.2 Central results

For ease of notation let $\text{Div}(C_g)_d$ denote all divisors of degree d. Then, let us consider the restriction of eq. (A.6) to $\text{Div}(C_g)_d$, i.e.

$$\Phi_d \colon \text{Div}(C_g)_d/\text{Prin}(C_g) \to \mathbb{C}^g/\Lambda \,, \ D \mapsto \sum_{i=1}^{N} \lambda_i \cdot \phi\,(p_i)\,. \tag{A.9}$$

Let us pick an integer $r \geq -1$ and study the subvariety of $\text{Jac}(C_g)$

$$G_d^r = \left\{ p \in \text{im}\,(\Phi_d) \,,\, h^0\left(C_g, \mathcal{O}_{C_g}(\Phi_d^{-1}(p))\right) = r+1 \right\}\,. \tag{A.10}$$

Then, the central result of Brill-Noether theory states [47]

$$\dim(G_d^r) \geq \rho\,(r,d,g) \equiv g - (r+1) \cdot ((r+1) - (d-g+1))\,. \tag{A.11}$$

By use of the Riemann-Roch theorem

$$h^0\left(C_g, \mathcal{O}_{C_g}(D)\right) - h^1\left(C_g, \mathcal{O}_{C_g}(D)\right) = \deg\left(\mathcal{O}_{C_g}(D)\right) - g + 1 = d - g + 1\,, \tag{A.12}$$

we can rewrite this results in the suggestive form

$$\dim(G_d^r) \geq \rho\,(r,d,g) \equiv g - n^0 \cdot n^1\,, \tag{A.13}$$

with $n^0 \equiv r+1$ and $n^1 = r+1 - (d-g+1)$. We may thus use $\rho\,(r,d,g)$ as a measure for how likely it is that a line bundle of degree d on a genus g curve C_g has $n^0 = r+1$ global sections.

Let us demonstrate this for degree $d = 2$ bundles on a genus-3 curve. By general theory, the number of section of a line bundle on a curve C_g with $g \geq 1$ can never exceed its degree. Hence $n^0 \in \{0,1,2\}$. With this information, let us compute $\rho\,(r,d,g)$ for the admissible values of r:

r	(n^0, n^1)	$\rho(r,d,g)$
-1	$(0,0)$	3
0	$(1,1)$	2
1	$(2,2)$	-1

$$\tag{A.14}$$

From this we learn, that most line bundles $\mathcal{L}$ of degree 2 on a genus-3 curve C_3 satisfy $h^0\,(C_3, \mathcal{L}) = 0$. Since for these bundles ρ matches the dimension of the Jacobian of C_3, we can say that these line bundles are associated to generic points of the Jacobian. Furthermore, we learn that there are such line bundles with $h^0\,(C_3, \mathcal{L}) = 1$. However, these are special in the sense that they are associated to a codimension-1 locus in the Jacobian $\text{Jac}(C_3)$.

Finally, $\rho = -1$ for $r = 1$ begs for an explanation. This explanation follows from work of Griffiths and Harris [50]:

On **generic** curves, $\dim(G_d^r) = \rho\,(r,d,g)$.

So in particular, on generic curves it holds $G_d^r = \emptyset$ if and only if $\rho\,(r,d,g) < 0$. Consequently, we conclude from eq. (A.14), that on generic genus $g = 3$ curve, there is no line bundle $\mathcal{L}$ of degree 2 such that $h^0(C_3, \mathcal{L}) = 2$.

Note however, that this does not rule out the possibility that non-generic curves may host such line bundles. In the case at hand, it follows from the theorem of Cliffford [50] that hyperelliptic curves H_3 of genus $g = 3$ admit line bundles $\mathcal{L}$ of degree 2 and $h^0(H_3, \mathcal{L}) = 2$.

A.1.3 Brill-Noether jump

As we see from eq. (A.14), we can in general modify a line bundle on a generic curve such that it admits additional sections. A jump from $r = r_{\text{generic}}$ to $r > r_{\text{generic}} + 1$ is equivalent to saying that the Serre-dual bundle admits a section, i.e., becomes effective:

$$K_C - D > 0 \quad \Leftrightarrow \quad \exists p_i \colon K_C - D \sim \sum_i p_i \,. \tag{A.15}$$

where $\sim$ represents linear equivalence of divisors. Obviously, this requires the line bundle divisor D to move into special alignment relative to K_C. Such a divisor is termed a *special* divisor. We term a change in h^0, which is solely attributed to a special alignment of the line bundle divisor, a *Brill-Noether jump*.

A.2 Koszul resolution

A.2.1 Generalities

Given a curve $C \subset S$ in a smooth surface S and a line bundle $\mathcal{L} = L|_C$ where $L \in \text{Pic}(S)$, we wish to identify which deformations of the curve C lead to an increased number of global sections for $\mathcal{L}$. For hypersurface curves in dP_3, the answer follows from a study of the Koszul resolution. In this case $C(\mathbf{c}) = V(P(\mathbf{c}))$ for a polynomial $P(\mathbf{c})$. Then, the Koszul resolution of $\mathcal{L}$ is given by the short-exact sequence

$$0 \to \mathcal{O}_{dP_3}(D_L - D_C) \xrightarrow{\alpha} \mathcal{O}_{dP_3}(D_L) \to \mathcal{L}(\mathbf{c}) \to 0 \,. \tag{A.16}$$

The map α is induced by the polynomial $P(\mathbf{c})$. Namely, for $U \subseteq dP_3$ open, α is given by

$$s \in \mathcal{O}_{dP_3}(D_L - D_C)(U) \mapsto s \cdot P(\mathbf{c}) \in \mathcal{O}_{dP_3}(D_L)(U) \,. \tag{A.17}$$

The Koszul resolution then induces the following long exact sequence in sheaf cohomology:

$$
\begin{aligned}
0 \longrightarrow H^0(dP_3, D_L - D_C) &\xrightarrow{\varphi_0} H^0(dP_3, D_L) \longrightarrow H^0(C(\mathbf{c}), \mathcal{L}(\mathbf{c})) \\
\hookrightarrow H^1(dP_3, D_L - D_C) &\xrightarrow{\varphi_1} H^1(dP_3, D_L) \longrightarrow H^1(C(\mathbf{c}), \mathcal{L}(\mathbf{c})) \\
\hookrightarrow H^2(dP_3, D_L - D_C) &\xrightarrow{\varphi_2} H^2(dP_3, D_L) \longrightarrow 0 \longrightarrow 0 \,.
\end{aligned}
\tag{A.18}
$$

The maps $\varphi_i = \varphi_i(\mathbf{c})$ are induced from multiplication with $P(\mathbf{c})$. Therefore, these maps are sensitive to the choice of parameters $\mathbf{c}$ for the curve $C(\mathbf{c})$. Explicitly, the maps φ_i are vector-space morphisms and the entries of their defining matrices are functions of the parameters c_i. Provided that we know these mapping matrices, we may thus use the exactness of the Koszul resolution of infer $h^i(C(\mathbf{c}), \mathcal{L}(\mathbf{c}))$ as a function of the coefficients c_i in $P(\mathbf{c})$.

For example, in subsection 4.1.1, we consider $D_C = (4; -1, -2, -1)$ and $D_L = (3; -3, -1, -2)$. In this case, the Koszul resolution simplifies and takes the form

$$
\begin{array}{ccccccc}
0 & \longrightarrow & 0 & \longrightarrow & 0 & \longrightarrow & H^0\left(C(\mathbf{c}), \mathcal{L}(\mathbf{c})\right) \\
\end{array}
$$
$$
\hookrightarrow H^1\left(dP_3, D_L - D_C\right) \cong \mathbb{C}^4 \xrightarrow{\varphi} H^1\left(dP_3, D_L\right) \cong \mathbb{C}^1 \longrightarrow H^1\left(C(\mathbf{c}), \mathcal{L}(\mathbf{c})\right)
$$
$$
\hookrightarrow 0 \longrightarrow 0 \longrightarrow 0 \longrightarrow 0 \,.
$$
(A.19)

Then it follows

$$
\begin{aligned}
H^1\left(C(\mathbf{c}), \mathcal{L}(\mathbf{c})\right) &\cong \operatorname{coker}\varphi \,, \\
h^1(C(\mathbf{c}), \mathcal{L}(\mathbf{c})) &= 1 - \dim\left(\operatorname{im}\varphi\right) \,.
\end{aligned}
$$
(A.20)

A detailed study of Čech cohomology [60] shows that in this geometry we have $M_\varphi = (c_3, c_6, c_9, 0)$. Hence, $h^1(C(\mathbf{c}), \mathcal{L}(\mathbf{c})) = 1$ on curves with $c_3 = c_6 = c_9 = 0$ and otherwise $h^1(C(\mathbf{c}), \mathcal{L}(\mathbf{c})) = 0$. Along these lines, we classify the curve geometries according to their admitted number of global sections.

Recall that Čech cohomology expresses $H^i(dP_3, \mathcal{O}_{dP_3}(D_L - D_C))$ and $H^i(dP_3, \mathcal{O}_{dP_3}(D_L))$ as collections of local sections, which are called Čech cochains. The mappings of these Čech cochains follow from eq. (A.17), i.e., are given by multiplication of the local sections with the polynomial $P(\mathbf{c})$ which defines the curve $C(\mathbf{c})$. Importantly, the Čech cochains define elements in Čech cohomology once considered modulo equivalence relations induced from Čech coboundaries. Identification of the Čech cochains and coboundaries is typically tedious, which makes these computations hard in practice.

Oftentimes, *cohomCalg* [61–67] can help to simplify this task. Namely, it identifies bases of $H^i(dP_3, \mathcal{O}_{dP_3}(D_L - D_C))$ and $H^i(dP_3, \mathcal{O}_{dP_3}(D_L))$ in terms of rationoms — quotients of monomials in the homogeneous coordinates — and therefore simplifies the task to identify the Čech cochains. Even more, we may be tempted to simply multiply the basis elements identifed by *cohomCalg* [61–67] with the polynomial $P(\mathbf{c})$ and ignore all image rationoms that have not been identified as bases for $H^i(dP_3, \mathcal{O}_{dP_3}(D_L))$ by *cohomCalg* under the assumption that they correspond to Čech coboundaries.

This procedure fails whenever *Čech cohomology chamber factors* greater than 1 appear. In this case, *cohomCalg* finds that one rationom R spans a vector space of dimension greater than 1 in sheaf cohomology. The interpretation of this is, that there are at least two distinct Čech cochains, i.e., collections of local sections, in which the rationom R is the only non-trivial entry. Hence, these distinct Čech cochains are both canonically isomorphic to R. However, to identify the mapping matrices of the line bundle cohomologies correctly, the information about R is insufficient. Rather, the corresponding Čech cochains need to be identified explicitly.

Given these insights, we have taken extra care, to work out the mappings presented in this work carefully with Čech cohomology. We present such a computation in large detail in the following section.

Before we come to this, let us mentioned that a detailed study of the Koszul resolution is not original to this work. For example, in the context of heterotic compactifications, these resolutions — including the mappings in the induced long exact sequence — have been studied extensively [68–72]. However, to the best of our knowledge, chamber factor greater than 1 do not show in products of projective spaces. Hence, this complication does not arise in existing heterotic compactifications with CICYs.

A.2.2 Čech cohomologies for section 4.2.1

Here, we present a more detailed computation of the example discussed in subsubsection 4.2.1. Recall that the curve and line bundle in question are given by

$$D_C = (4; -1, -1, -1), \qquad D_L = (1; 2, -2, -1). \tag{A.21}$$

Moreover, recall that in this case $h^0\left(C(\mathbf{c}), \mathcal{L}(\mathbf{c})\right)$ is uniquely determined by the mapping

$$\varphi\colon H^1\left(dP_3, \mathcal{O}_{dP_3}\left(D_L - D_C\right)\right) \xrightarrow{\cdot P(\mathbf{c})} H^1\left(dP_3, \mathcal{O}_{dP_3}\left(D_L\right)\right), \tag{A.22}$$

where

$$\begin{aligned}
P(\mathbf{c}) = {}& c_1 x_1^3 x_2^3 x_3^2 x_4 + c_2 x_1^2 x_2^3 x_3 x_4^2 x_6 + c_3 x_1 x_2^3 x_3^3 x_6^2 + c_4 x_1^3 x_2^2 x_3^3 x_5 + c_5 x_1^2 x_2^2 x_3^2 x_4 x_5 x_6 \\
& + c_6 x_1 x_2^2 x_3 x_4^2 x_5 x_6^2 + c_7 x_2^2 x_4^3 x_5 x_6^3 + c_8 x_1^2 x_2 x_3^3 x_5^2 x_6 + c_9 x_1 x_2 x_3^2 x_4 x_5^2 x_6^2 \\
& + c_{10} x_2 x_3 x_4^2 x_5^2 x_6^3 + c_{11} x_1 x_3^3 x_5^3 x_6^2 + c_{12} x_3^2 x_4 x_5^3 x_6^3.
\end{aligned} \tag{A.23}$$

Namely, $h^0(C(\mathbf{c}), \mathcal{L}(\mathbf{c})) = 3 - \mathrm{rk}\,(M_\varphi)$. With *cohomCalg* [61–67], we obtain a basis of the line bundle cohomologies:

$$H^1(D_L - D_C) \cong \mathrm{Span}_{\mathbb{C}}\left\{\frac{1}{x_3 x_4^3 x_6^3}, \frac{1}{x_1 x_3^2 x_4^2 x_6^2}, \frac{1}{x_1^2 x_3^3 x_4 x_6}\right\} \cong \mathbb{C}^3, \tag{A.24}$$

$$H^1(D_L) \cong \mathrm{Span}_{\mathbb{C}}\left\{\frac{x_5^3 x_6}{x_1 x_4}, \frac{x_1 x_2^3}{x_3 x_6}\right\} \cong \mathbb{C}^2. \tag{A.25}$$

By polynomial multiplication we then have

$$\frac{1}{x_3 x_4^3 x_6^3} \cdot P(\mathbf{c}) = c_3 \frac{x_1 x_2^3}{x_3 x_6} + \dots, \tag{A.26}$$

$$\frac{1}{x_1 x_3^2 x_4^2 x_6^2} \cdot P(\mathbf{c}) = c_2 \frac{x_1 x_2^3}{x_3 x_6} + c_{12} \frac{x_5^3 x_6}{x_1 x_4} + \dots, \tag{A.27}$$

$$\frac{1}{x_1^2 x_3^3 x_4 x_6} \cdot P(\mathbf{c}) = c_1 \frac{x_1 x_2^3}{x_3 x_6} + c_{11} \frac{x_5^3 x_6}{x_1 x_4} + \dots. \tag{A.28}$$

On the r.h.s. of these equations, we have omitted all rationoms which cannot be expressed as $\mathbb{Z}$-linear combinations of those listed in eq. (A.25). The remainder of this section will justify that we can indeed omit these terms. For the time being, note that this leads to

$$M_\varphi = \begin{pmatrix} c_3 & c_2 & c_1 \\ 0 & c_{12} & c_{11} \end{pmatrix}, \tag{A.29}$$

which is the matrix analyzed in subsubsection 4.2.1.

Strategy. In order to justify that all omitted terms in eq. (A.28) can be ignored, we will now analyse $H^1(dP_3, \mathcal{O}_{dP_3}(D_L))$ and $H^1(dP_3, \mathcal{O}_{dP_3}(D_L - D_C))$ from the perspective of Čech cohomology. For additional background we refer the interested reader to [60]. Recall that for $H^1(dP_3, \mathcal{O}_{dP_3}(D_L))$ it holds

$$H^1(dP_3, \mathcal{O}_{dP_3}(D_L)) \cong \check{H}^1(\mathcal{U}, \mathcal{O}_{dP_3}(D_L)) = \ker(\delta_1)/\operatorname{im}(\delta_0). \tag{A.30}$$

In this expression, $\mathcal{U}$ is the affine open cover of the dP_3 surface — we will discuss this momentarily — and the maps δ_i are the boundary morphisms in the Čech complex

$$0 \to \check{C}^0(\mathcal{U}, \mathcal{O}_{dP_3}(D_L)) \xrightarrow{\delta_0} \check{C}^1(\mathcal{U}, \mathcal{O}_{dP_3}(D_L)) \xrightarrow{\delta_1} \dots \tag{A.31}$$

Thereby, let us specify our statement regarding the r.h.s. of eq. (A.28). We claim that all omitted terms are in $\operatorname{im}(\delta_0)$, i.e., are Čech coboundaries. To justify this statement, we proceed by investigating the following objects:

1. $\operatorname{im}(\delta_0(D_L))$,

2. $\ker(\delta_1(D_L))$,

3. $\ker(\delta_1(D_L - D_C))$,

4. the map $\ker(\delta_1(D_L - D_C)) \to \ker(\delta_1(D_L))$.

Čech 0-cocycles of D_L. To understand $\check{C}^0(\mathcal{U}, \mathcal{O}_{dP_3}(D_L))$, recall that dP_3 has 6 homogeneous variables x_i. These correspond to the ray generators

$$\begin{aligned} u_1 &= (0, -1), & u_2 &= (-1, 0), & u_3 &= (1, -1), \\ u_4 &= (-1, 1), & u_5 &= (1, 0), & u_6 &= (0, 1). \end{aligned} \tag{A.32}$$

In terms of these, the maximal cones in the fan of dP_3 are given by

$$\begin{aligned} U_1 &= \operatorname{Span}_{\geq 0}\{u_1, u_3\}, & U_2 &= \operatorname{Span}_{\geq 0}\{u_3, u_5\}, & U_3 &= \operatorname{Span}_{\geq 0}\{u_5, u_6\}, \\ U_4 &= \operatorname{Span}_{\geq 0}\{u_6, u_4\}, & U_5 &= \operatorname{Span}_{\geq 0}\{u_4, u_2\}, & U_6 &= \operatorname{Span}_{\geq 0}\{u_2, u_1\}. \end{aligned} \tag{A.33}$$

These cones correspond to open affine subsets of the dP_3, namely the subsets $\{x_i \neq 0\}$. Collectively, $\mathcal{U} = \{U_i\}_{1 \leq i \leq 6}$ is the open affine cover of dP_3. To compute $\check{C}^0(\mathcal{U}, \mathcal{O}_{dP_3}(D_L))$ with respect to this open affine cover $\mathcal{U}$, we note

$$D_L = (1; 2, -2, -1) = H + 2E_1 - 2E_2 - E_3 = \sum_{i=1}^{6} a_i V(x_i), \tag{A.34}$$

with $a_1 = a_4 = a_6 = 0$ and $a_2 = 2$, $a_3 = -1$, $a_5 = 1$. Now, we can quote from [60] that

$$\check{C}^0(\mathcal{U}, \mathcal{O}_{X_\Sigma}(D_L)) = \bigoplus_{1 \leq i \leq 6} H^0(U_i, \mathcal{O}_{X_\Sigma}(D_L)|_{U_i}), \tag{A.35}$$

$$H^0(U_i, \mathcal{O}_{X_\Sigma}(D_L)|_{U_i}) \cong \left(\prod_{j=1}^{6} x_j^{a_j}\right) \cdot \bigoplus_{m \in P_D(U_i)} \mathbb{C} \cdot \left(\prod_{j=1}^{6} x_j^{\langle m, u_j \rangle}\right), \tag{A.36}$$

$$P_D(U_i) = \{m \in \mathbb{Z}^2, \langle m, u_\rho \rangle \geq -a_\rho \; \forall \rho \in \sigma(1)\}. \tag{A.37}$$

The normalization in eq. (A.36) ensures that we are looking at rationoms of degree D_L, as analysed by *cohomCalg*. Explicitly, it holds

$$P_D(U_1) = \{m \in \mathbb{Z}^2, -m_2 \geq 0 \text{ and } m_1 - m_2 \geq 1\}, \tag{A.38}$$

$$P_D(U_2) = \{m \in \mathbb{Z}^2, m_1 - m_2 \geq 1 \text{ and } m_1 \geq -1\}, \tag{A.39}$$

$$P_D(U_3) = \{m \in \mathbb{Z}^2, m_1 \geq -1 \text{ and } m_2 \geq 0\}, \tag{A.40}$$

$$P_D(U_4) = \{m \in \mathbb{Z}^2, m_2 \geq 0 \text{ and } -m_1 + m_2 \geq 0\}, \tag{A.41}$$

$$P_D(U_5) = \{m \in \mathbb{Z}^2, -m_1 + m_2 \geq 0 \text{ and } -m_1 \geq -2\}, \tag{A.42}$$

$$P_D(U_6) = \{m \in \mathbb{Z}^2, -m_1 \geq -2 \text{ and } -m_2 \geq 0\}. \tag{A.43}$$

To express these polytopes in simpler terms, we define the regions A, B, C, D, E, F, G, H:
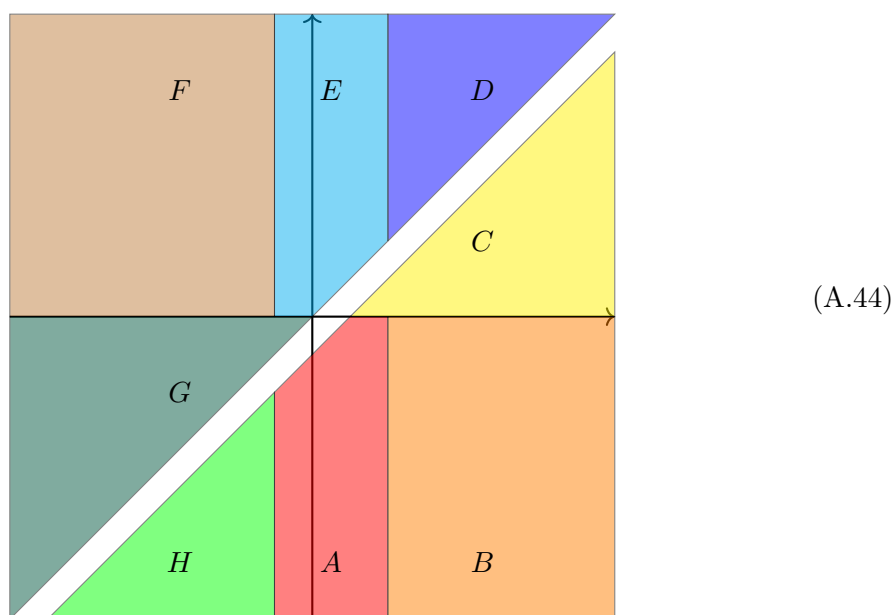
$$\tag{A.44}$$

In an abuse of terminology, we use A to denote all polynomials formed from linear combination of the Laurent monomials associated to the lattice points of the region A. Similarly, we use the names for the other regions. Thereby, we can write

$$\check{C}^0(\mathcal{U}, \mathcal{O}_{X_\Sigma}(D_L)) = \frac{x_2^2 x_5}{x_3} \cdot (H + A + B, A + B + C, C + D + E,$$
$$D + E + F, E + F + G, G + H + A). \tag{A.45}$$

Finally note that the map $\delta_0 \colon \check{C}^0(\mathcal{U}, \mathcal{O}_{X_\Sigma}(D_L)) \to \check{C}^1(\mathcal{U}, \mathcal{O}_{X_\Sigma}(D_L))$ is given by multiplication with the following matrix:

$$
M_{\delta_0} = \begin{pmatrix}
-1 & 1 & 0 & 0 & 0 & 0 \\
-1 & 0 & 1 & 0 & 0 & 0 \\
-1 & 0 & 0 & 1 & 0 & 0 \\
-1 & 0 & 0 & 0 & 1 & 0 \\
-1 & 0 & 0 & 0 & 0 & 1 \\
0 & -1 & 1 & 0 & 0 & 0 \\
0 & -1 & 0 & 1 & 0 & 0 \\
0 & -1 & 0 & 0 & 1 & 0 \\
0 & -1 & 0 & 0 & 0 & 1 \\
0 & 0 & -1 & 1 & 0 & 0 \\
0 & 0 & -1 & 0 & 1 & 0 \\
0 & 0 & -1 & 0 & 0 & 1 \\
0 & 0 & 0 & -1 & 1 & 0 \\
0 & 0 & 0 & -1 & 0 & 1 \\
0 & 0 & 0 & 0 & -1 & 1
\end{pmatrix} . \tag{A.46}
$$

Čech 1-cocycles of D_L. We repeat this analysis for $\check{C}^1(\mathcal{U}, \mathcal{O}_{dP_3}(D_L))$. The elements in this Čech cohomology are given by local sections on pairwise intersections of the U_i which form the affine open cover of dP_3. These pairwise intersections and the corresponding polytopes are as follows:

Intersection	Cone	$P_D(U_{ij})$	Presentation
$U_1 \cap U_2$	$\mathrm{Span}_{\geq 0}(u_3)$	$\{m \in \mathbb{Z}^2 ,\, m_1 \geq 1 + m_2\}$	B, C, D, E, F, N
$U_1 \cap U_3$	$\mathrm{Span}_{\geq 0}(0)$	$\mathbb{Z}^2$	
$U_1 \cap U_4$	$\mathrm{Span}_{\geq 0}(0)$	$\mathbb{Z}^2$	
$U_1 \cap U_5$	$\mathrm{Span}_{\geq 0}(0)$	$\mathbb{Z}^2$	
$U_1 \cap U_6$	$\mathrm{Span}_{\geq 0}(u_1)$	$\{m \in \mathbb{Z}^2 ,\, m_1 \leq 0\}$	A, B, C, I, K, L
$U_2 \cap U_3$	$\mathrm{Span}_{\geq 0}(u_5)$	$\{m \in \mathbb{Z}^2 ,\, m_1 \geq -1\}$	$C, D, E, F, G, H, I, L, M, N$
$U_2 \cap U_4$	$\mathrm{Span}_{\geq 0}(0)$	$\mathbb{Z}^2$	
$U_2 \cap U_5$	$\mathrm{Span}_{\geq 0}(0)$	$\mathbb{Z}^2$	
$U_2 \cap U_6$	$\mathrm{Span}_{\geq 0}(0)$	$\mathbb{Z}^2$	
$U_3 \cap U_4$	$\mathrm{Span}_{\geq 0}(u_6)$	$\{m \in \mathbb{Z}^2 ,\, m_2 \geq 0\}$	$A, B, C, D, E, L,\text{M}$
$U_3 \cap U_5$	$\mathrm{Span}_{\geq 0}(0)$	$\mathbb{Z}^2$	
$U_3 \cap U_6$	$\mathrm{Span}_{\geq 0}(0)$	$\mathbb{Z}^2$	
$U_4 \cap U_5$	$\mathrm{Span}_{\geq 0}(u_4)$	$\{m \in \mathbb{Z}^2 ,\, m_2 \geq m_1\}$	A, G, H, I, K, L
$U_4 \cap U_6$	$\mathrm{Span}_{\geq 0}(0)$	$\mathbb{Z}^2$	
$U_5 \cap U_6$	$\mathrm{Span}_{\geq 0}(u_2)$	$\{m \in \mathbb{Z}^2 ,\, m_1 \leq 2\}$	$A, B, C, D, H, I, K, L, M, N$

(A.47)

In this table, we have use the following geometric loci to express the polytopes in question:
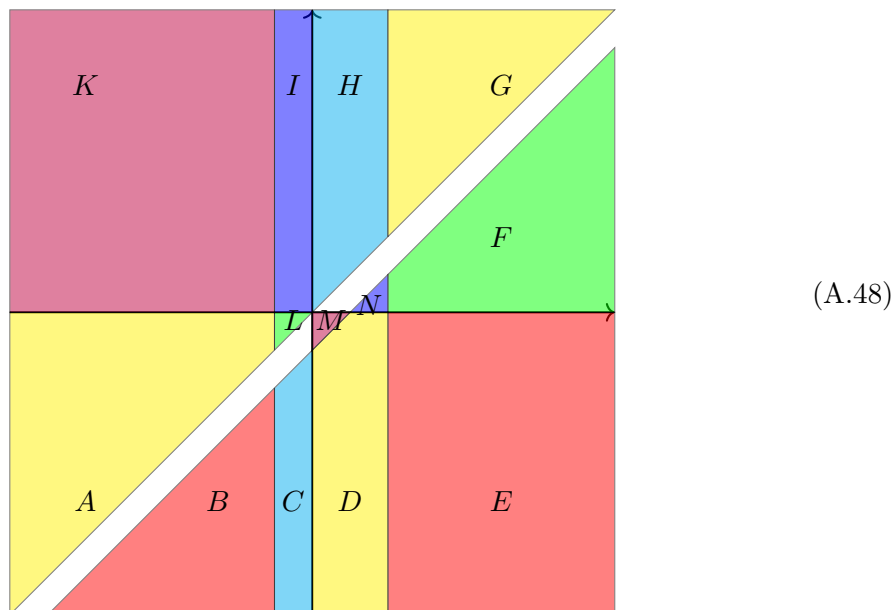
$$(A.48)$$

To identify a basis of $\ker(\delta_1)$, we look at the corresponding mapping matrix

$$M_{\delta_1} = \begin{pmatrix}
-1 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\
0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\
0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\
0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\
0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & -1
\end{pmatrix}. \qquad (A.49)$$

Let us introduce the points

$$p_2 = (2,1), \qquad p_9 = (-1,-1). \qquad (A.50)$$

The corresponding Laurent monomials, once multiplied by $x^a \equiv \prod_{j=1}^6 x_j^{a_j}$, are $\frac{x_5^3 x_6}{x_1 x_4}$, $\frac{x_1 x_2^3}{x_3 x_6}$, i.e., exactly those rationoms which *cohomCalg* identified in eq. (A.25) as basis of the cohomology:

$$H^1(D_L) \cong \mathrm{Span}_{\mathbb{C}} \left\{ \frac{x_5^3 x_6}{x_1 x_4}, \frac{x_1 x_2^3}{x_3 x_6} \right\}. \qquad (A.51)$$

However, here we can make this isomorphism explicit. In an abuse of terminology let p_2, p_9 denote their Laurent monomials. Then it is readily verified that the following Čech

1-cocycles furnish a basis of $\ker(\delta_1)$:

$$(0, -p_2, -p_2, -p_2, 0, -p_2, -p_2, -p_2, 0, 0, 0, p_2, 0, p_2, p_2) \cong \frac{x_5^3 x_6}{x_1 x_4}, \qquad \text{(A.52)}$$

$$(0, -p_9, -p_9, -p_9, 0, -p_9, -p_9, -p_9, 0, 0, 0, p_9, 0, p_9, p_9) \cong \frac{x_1 x_2^3}{x_3 x_6}. \qquad \text{(A.53)}$$

Čech 1-cocycles of $D_L - D_C$. Finally, let us identify $\check{C}^1(\mathcal{U}, \mathcal{O}_{dP_3}(D_L - D_C))$. We have

$$D_L - D_C = (-3; 3, -1, 0) = 3V(x_2) - 4V(x_3) - 3V(x_5) - 3V(x_6). \qquad \text{(A.54)}$$

Thus $a_1 = a_4 = 0$, $a_2 = 3$, $a_3 = -4$ and $a_5 = a_6 = -3$. The points associated to the Laurent monomials identified by *cohomCalg* in eq. (A.24) are:

$$\frac{1}{x_3 x_4^3 x_6^3} = \frac{x_2^3}{x_3^4 x_5^3 x_6^3} \cdot \frac{x_3^3 x_5^3}{x_2^3 x_4^3} \qquad \leftrightarrow \qquad q_1 = (3, 0), \qquad \text{(A.55)}$$

$$\frac{1}{x_1 x_3^2 x_4^2 x_6^2} = \frac{x_2^3}{x_3^4 x_5^3 x_6^3} \cdot \frac{x_3^2 x_5^3 x_6}{x_2^3 x_4^2 x_1} \qquad \leftrightarrow \qquad q_2 = (3, 1), \qquad \text{(A.56)}$$

$$\frac{1}{x_1^2 x_3^3 x_4 x_6} = \frac{x_2^3}{x_3^4 x_5^3 x_6^3} \cdot \frac{x_3 x_5^3 x_6^2}{x_4 x_2^3 x_1^2} \qquad \leftrightarrow \qquad q_3 = (3, 2). \qquad \text{(A.57)}$$

The relevant pairwise intersection and polytopes are as follows:

Intersection	Cone	$P_D(U_{ij})$	Points contained
$U_1 \cap U_2$	$\mathrm{Span}_{\geq 0}(u_3)$	$\{m \in \mathbb{Z}^2, \, m_1 - m_2 \geq 4\}$	$\emptyset$
$U_1 \cap U_3$	$\mathrm{Span}_{\geq 0}(0)$	$\mathbb{Z}^2$	q_1, q_2, q_3
$U_1 \cap U_4$	$\mathrm{Span}_{\geq 0}(0)$	$\mathbb{Z}^2$	q_1, q_2, q_3
$U_1 \cap U_5$	$\mathrm{Span}_{\geq 0}(0)$	$\mathbb{Z}^2$	q_1, q_2, q_3
$U_1 \cap U_6$	$\mathrm{Span}_{\geq 0}(u_1)$	$\{m \in \mathbb{Z}^2, \, -m_2 \geq 0\}$	q_1
$U_2 \cap U_3$	$\mathrm{Span}_{\geq 0}(u_5)$	$\{m \in \mathbb{Z}^2, \, m_1 \geq 3\}$	q_1, q_2, q_3
$U_2 \cap U_4$	$\mathrm{Span}_{\geq 0}(0)$	$\mathbb{Z}^2$	q_1, q_2, q_3
$U_2 \cap U_5$	$\mathrm{Span}_{\geq 0}(0)$	$\mathbb{Z}^2$	q_1, q_2, q_3
$U_2 \cap U_6$	$\mathrm{Span}_{\geq 0}(0)$	$\mathbb{Z}^2$	q_1, q_2, q_3
$U_3 \cap U_4$	$\mathrm{Span}_{\geq 0}(u_6)$	$\{m \in \mathbb{Z}^2, \, m_2 \geq 3\}$	$\emptyset$
$U_3 \cap U_5$	$\mathrm{Span}_{\geq 0}(0)$	$\mathbb{Z}^2$	q_1, q_2, q_3
$U_3 \cap U_6$	$\mathrm{Span}_{\geq 0}(0)$	$\mathbb{Z}^2$	q_1, q_2, q_3
$U_4 \cap U_5$	$\mathrm{Span}_{\geq 0}(u_4)$	$\{m \in \mathbb{Z}^2, \, -m_1 + m_2 \geq 0\}$	$\emptyset$
$U_4 \cap U_6$	$\mathrm{Span}_{\geq 0}(0)$	$\mathbb{Z}^2$	q_1, q_2, q_3
$U_5 \cap U_6$	$\mathrm{Span}_{\geq 0}(u_2)$	$\{m \in \mathbb{Z}^2, \, -m_1 \leq -3\}$	q_1, q_2, q_3

$$\text{(A.58)}$$

It is not hard to verify that $\ker(\delta_1) = \mathrm{Span}\{b_1, b_2, b_3\}$ where

$$b_1 = (0, q_1, q_1, q_1, 0, q_1, q_1, q_1, 0, 0, 0, -q_1, 0, -q_1, -q_1), \qquad \text{(A.59)}$$

$$b_2 = (0, q_2, q_2, q_2, 0, q_2, q_2, q_2, 0, 0, 0, -q_2, 0, -q_2, -q_2), \qquad \text{(A.60)}$$

$$b_3 = (0, q_3, q_3, q_3, 0, q_3, q_3, q_3, 0, 0, 0, -q_3, 0, -q_3, -q_3). \qquad \text{(A.61)}$$

Images of b_1, b_2, b_3 in $\check{C}^1(\mathcal{U}, D_L)$. The mapping between the Čech cocycles happens through the following mapping of complexes

$$
\begin{array}{ccccccc}
0 \to \check{C}^0(\mathcal{U}, D_L - D_C) & \xrightarrow{\delta_0} & \check{C}^1(\mathcal{U}, D_L - D_C) & \xrightarrow{\delta_1} & \check{C}^2(\mathcal{U}, D_L - D_C) \to \cdots \\
{\scriptstyle \cdot P(\mathbf{c})} \downarrow & & {\scriptstyle \cdot P(\mathbf{c})} \downarrow & & {\scriptstyle \cdot P(\mathbf{c})} \downarrow \\
0 \longrightarrow \check{C}^0(\mathcal{U}, D_L) & \xrightarrow{\delta_0} & \check{C}^1(\mathcal{U}, D_L) & \xrightarrow{\delta_1} & \check{C}^2(\mathcal{U}, D_L) \longrightarrow \cdots
\end{array}
\tag{A.62}
$$

where $P(\mathbf{c})$ is the global section of D_C in eq. (A.23). From this it is now readily verified, that the terms omitted on the r.h.s. of eq. (A.28) can be dropped. Namely, we map the basis elements b_i in eq. (A.61) to $\check{C}^1(\mathcal{U}, D_L)$ by multiplying them with the polynomial $P(\mathbf{c})$. Thereby, it is not to hard to see that those terms ignored on the r.h.s. of eq. (A.28) correspond to the elements of $\check{C}^1(\mathcal{U}, D_L)$ which are given by

$$
\varphi_i = (0, r_i, r_i, r_i, 0, r_i, r_i, r_i, 0, 0, 0, -r_i, 0, -r_i, -r_i) \,,
\tag{A.63}
$$

where r_i is the Laurent monomial associated — upon multiplication by $x^a = \frac{x_2^2 x_5}{x_3}$ — to

$$
\begin{aligned}
& r_1 = (-1, -3), \quad r_2 = (-1, -2), \quad r_3 = (2, -1), \quad r_4 = (-1, 0), \\
& r_5 = (2, 0), \qquad\;\; r_6 = (-1, 1), \qquad r_7 = (1, 1).
\end{aligned}
\tag{A.64}
$$

We claim that these elements φ_i are in the image of $\delta_0 \colon \check{C}^0(\mathcal{U}, D_L) \to \check{C}^1(\mathcal{U}, D_L)$ and do therefore not contribute to $\check{H}^1(\mathcal{U}, D_L) = \check{C}^1(\mathcal{U}, D_L)/\mathrm{im}(\delta_0)$. Hence, we claim that $\varphi_i = \delta_0(\mu_i)$ for $\mu_i \in \check{C}^0(\mathcal{U}, D_L)$. It is readily verified, that this is indeed satisfied for the following μ_i:

$$
\begin{array}{cc}
\hline
\varphi_i & \mu_i \\
\hline
\varphi_1 & (r_1, r_1, 0, 0, 0, r_1) \\
\varphi_2 & (r_2, r_2, 0, 0, 0, r_2) \\
\varphi_3 & (-r_3, -r_3, 0, 0, 0, -r_3) \\
\varphi_4 & (0, 0, r_4, r_4, r_4, 0) \\
\hline
\varphi_5 & (-r_5, -r_5, 0, 0, 0, -r_5) \\
\varphi_6 & (0, 0, r_6, r_6, r_6, 0) \\
\varphi_7 & (0, 0, r_7, r_7, r_7, 0) \\
\hline
\end{array}
\tag{A.65}
$$

Hence, we conclude

$$
\varphi(b_1) \cong c_3 \frac{x_1 x_2^3}{x_3 x_6}, \quad \varphi(b_2) \cong c_2 \frac{x_1 x_2^3}{x_3 x_6} + c_{12} \frac{x_5^3 x_6}{x_1 x_4}, \quad \varphi(b_3) \cong c_1 \frac{x_1 x_2^3}{x_3 x_6} + c_{11} \frac{x_5^3 x_6}{x_1 x_4}.
\tag{A.66}
$$

This justifies our analysis based on the matrix in eq. (A.29).

A.2.3 Application to GUT-example

In the example discussed in section 3 we consider $D_C = (10; -3, -3, -4)$ and $D_L = (5; -4, -4, 3)$. This curve $C_{\mathbf{5}_3}$ is cut-out by the following polynomial $a_{3,2}$:

$$
\begin{aligned}
a_{3,2} = {} & c_{44}x_1^6x_2^7x_3^3x_4^4 + c_{43}x_1^6x_2^6x_3^4x_4^3x_5 + c_{42}x_1^6x_2^5x_3^5x_4^2x_5^2 + c_{41}x_1^6x_2^4x_3^6x_4x_5^3 \\
& + c_{40}x_1^6x_2^3x_3^7x_5^4 + c_{39}x_1^5x_2^7x_3^2x_4^5x_6 + c_{38}x_1^5x_2^6x_3^3x_4^4x_5x_6 + c_{37}x_1^5x_2^5x_3^4x_4^3x_5^2x_6 \\
& + c_{36}x_1^5x_2^4x_3^5x_4^2x_5^3x_6 + c_{35}x_1^5x_2^3x_3^6x_4x_5^4x_6 + c_{34}x_1^5x_2^2x_3^7x_5^5x_6 + c_{33}x_1^4x_2^7x_3x_4^6x_6^2 \\
& + c_{32}x_1^4x_2^6x_3^2x_4^5x_5x_6^2 + c_{31}x_1^4x_2^5x_3^3x_4^4x_5^2x_6^2 + c_{30}x_1^4x_2^4x_3^4x_4^3x_5^3x_6^2 + c_{29}x_1^4x_2^3x_3^5x_4^2x_5^4x_6^2 \\
& + c_{28}x_1^4x_2^2x_3^6x_4x_5^5x_6^2 + c_{27}x_1^4x_2x_3^7x_5^6x_6^2 + c_{26}x_1^3x_2^7x_4^7x_6^3 + c_{25}x_1^3x_2^6x_3x_4^6x_5x_6^3 \\
& + c_{24}x_1^3x_2^5x_3^2x_4^5x_5^2x_6^3 + c_{23}x_1^3x_2^4x_3^3x_4^4x_5^3x_6^3 + c_{22}x_1^3x_2^3x_3^4x_4^3x_5^4x_6^3 + c_{21}x_1^3x_2^2x_3^5x_4^2x_5^5x_6^3 \\
& + c_{20}x_1^3x_2x_3^6x_4x_5^6x_6^3 + c_{19}x_1^3x_3^7x_5^7x_6^3 + c_{18}x_1^2x_2^6x_4^7x_5x_6^4 + c_{17}x_1^2x_2^5x_3x_4^6x_5^2x_6^4 \\
& + c_{16}x_1^2x_2^4x_3^2x_4^5x_5^3x_6^4 + c_{15}x_1^2x_2^3x_3^3x_4^4x_5^4x_6^4 + c_{14}x_1^2x_2^2x_3^4x_4^3x_5^5x_6^4 + c_{13}x_1^2x_2x_3^5x_4^2x_5^6x_6^4 \\
& + c_{12}x_1^2x_3^6x_4x_5^7x_6^4 + c_{11}x_1x_2^5x_4^7x_5^2x_6^5 + c_{10}x_1x_2^4x_3x_4^6x_5^3x_6^5 + c_9x_1x_2^3x_3^2x_4^5x_5^4x_6^5 \\
& + c_8x_1x_2^2x_3^3x_4^4x_5^5x_6^5 + c_7x_1x_2x_3^4x_4^3x_5^6x_6^5 + c_6x_1x_3^5x_4^2x_5^7x_6^5 + c_5x_2^4x_4^7x_5^3x_6^6 \\
& + c_4x_2^3x_3x_4^6x_5^4x_6^6 + c_3x_2^2x_3^2x_4^5x_5^5x_6^6 + c_2x_2x_3^3x_4^4x_5^6x_6^6 + c_1x_3^4x_4^3x_5^7x_6^6 \,.
\end{aligned}
\tag{A.67}
$$

Hence, the Koszul resolution of the line bundle $\mathcal{L} = \mathcal{O}_{dP_3}(D_L)|_{C_{\mathbf{5}_3}}$ is given by

$$
0 \to \mathcal{O}_{dP_3}(D_L - D_C) \xrightarrow{\phi} \mathcal{O}_{dP_3}(D_L) \to \mathcal{L} \to 0 \,,
\tag{A.68}
$$

and the map ϕ is induced from multiplication with $a_{3,2}$. The associated long exact sequence in sheaf cohomology is then:

$$
\begin{aligned}
& 0 \longrightarrow 0 \longrightarrow H^0(dP_3, D_L) \cong \mathbb{C}^4 \longrightarrow H^0(D_C, \mathcal{L}) \\
& \hookrightarrow H^1(dP_3, D_L - D_C) \cong \mathbb{C}^4 \xrightarrow{\varphi} H^1(dP_3, D_L) \cong \mathbb{C}^6 \longrightarrow H^1(C_{\mathbf{5}_3}, \mathcal{L}) \\
& \hookrightarrow 0 \longrightarrow 0 \longrightarrow 0 \longrightarrow 0 \,.
\end{aligned}
\tag{A.69}
$$

By exactness of this sequence, we have $h^1(C_{\mathbf{5}_3}, \mathcal{L}) = 6 - \mathrm{rk}(M_\varphi)$, where the mapping matrix M_φ is determined by the coefficients of $a_{3,2}$:

$$
M_\varphi = \begin{pmatrix}
0 & c_1 & 0 & 0 & c_2 & c_3 & c_4 & c_5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
c_5 & 0 & 0 & 0 & c_1 & c_2 & c_3 & c_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
c_{11} & c_6 & 0 & 0 & c_7 & c_8 & c_9 & c_{10} & c_1 & c_2 & c_3 & c_4 & c_5 & 0 & 0 & 0 \\
0 & 0 & c_{39} & c_{34} & 0 & 0 & 0 & 0 & c_{40} & c_{41} & c_{42} & c_{43} & c_{44} & c_{35} & c_{36} & c_{37} & c_{38} \\
0 & 0 & c_{44} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_{40} & c_{41} & c_{42} & c_{43} \\
0 & 0 & 0 & c_{40} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_{41} & c_{42} & c_{43} & c_{44}
\end{pmatrix} \,.
\tag{A.70}
$$

Some linear algebra yields that the rank of this map drops by one, if

$$c_1 = c_2 = c_3 = c_4 = c_5 = c_7 = c_8 = c_9 = c_{10} = c_{35} = c_{36} = c_{37} = c_{38} = 1$$
$$c_{40} = c_{41} = c_{42} = c_{43} = c_{44} = 1, \quad c_{11} = c_{34} = -1, \quad c_6 = c_{39} = 2\,. \tag{A.71}$$

One can easily verify that the polynomial in eq. (A.67) does not factorize for generic other coefficients not tuned above. Hence, the curve $C_{\mathbf{5}_3}$ remains irreducible. By applying *sage-math* [73], one can further justify the smoothness of $C_{\mathbf{5}_3}$. Therefore, this tuning condition leads to one additional section without topology change for $C_{\mathbf{5}_3}$. This is an example of jump from Brill-Noether theory.

A.3 The fat point

Finally, in our analysis, non-reduced curves feature prominently. Consequently, a basic understanding of such curves is required. Let us therefore briefly discuss the mother of all non-reduced varieties — the *fat point*. This is an example in non-compact affine space $\mathbb{C}^2$ with coordinates x, y. Most of this intuition carries over to compact curves. More details can for example be found in [49, 74].

Let us consider $V(x) \subseteq \mathbb{C}^2$. This is the complex (non-compact) curve with coordinate y. The difference between $V(x)$ and $V(x^2)$ is not the collection of points, which these vanishing sets contain, but rather the allowed functions on these spaces. Namely, recall that in the modern language of algebraic geometry, a scheme (or equivalently in the analytic regime — a *geometric space*) is a pair of a topological space and a structure sheaf. The difference between $V(x)$ and $V(x^2)$ is this very structure sheaf.

In staying within the regime of algebraic geometry, the structure sheaf of $\mathbb{C}^2$ is given by (the sheafification of) the total coordinate ring $\mathbb{C}[x, y]$ — the ring of all polynomials in the variables x and y. Likewise, we can understand the structure sheaf on $V(x)$ from its coordinate ring:

$$R_{V(x)} = \mathbb{C}[x, y]/\langle x \rangle = \mathbb{C}[y]\,. \tag{A.72}$$

Hence, functions on the variety $V(x)$ correspond to polynomials in y. How about $V(x^2)$? On this space it holds

$$R_{V(x^2)} = \mathbb{C}[x, y]/\langle x^2 \rangle = \mathbb{C}[y] \oplus \langle x \rangle\,. \tag{A.73}$$

Consequently, on $V(x^2)$, the polynomial x provides a *non-trivial* function! This is the difference between $V(x)$ and $V(x^2)$.

We can extend this example slightly by looking at $V(y, x^2)$. For this space we find

$$R_{V(y,x^2)} = \mathbb{C}[x, y]/\langle y, x^2 \rangle = \langle x \rangle\,. \tag{A.74}$$

Hence, on this point in the affine plane $\mathbb{C}$, the set of non-trivial functions is 1-dimensional and is generated by the polynomial x. This lends $V(y, x^2)$ its name — as point set it is just a single point, yet this point is large enough to admit non-trivial functions — it is a *fat point*.

B Collection of data

B.1 Curve splittings and jumps

Recall that the six toric $\mathbb{P}^1$s of dP_3 correspond to the exceptional divisors E_1, E_2, E_3 and the following three divisors:

$$E_4 = H - E_1 - E_2\,, \qquad E_5 = H - E_1 - E_3\,, \qquad E_6 = H - E_2 - E_3\,. \tag{B.1}$$

B.1.1 $D_C = (3; -1, -1, -1)$

For this genus-1 curve we find:

bundle	h^0-values	E_1-splits	E_2-splits	E_3-splits	E_4-splits	E_5-splits	E_6-splits
(2, 1, -4, 1)	(4, 5, 6)	(4, 5, 6)	(4, 5, 6)	(4, 5, 6)	(4, 5, 6)	(4, 5, 6)	(4, 5, 6)
(1, -3, -3, -2)	(0, 1, 2, 3, 4, 5)	(2, 3, 4, 5)	(2, 3, 4, 5)	(1, 3, 4, 5)	(0, 1, 2, 3, 4, 5)	(0, 1, 2, 3, 4, 5)	(0, 1, 2, 3, 4, 5)
(1, -1, -3, 0)	(0, 1, 2, 3)	(0, 1, 2, 3)	(2)	(0, 1, 2, 3)	(1, 2, 3)	(0, 1, 2)	(0, 2, 3)
(1, -2, -3, -2)	(0, 1, 2, 3, 4)	(1, 2, 3, 4)	(2, 3, 4)	(1, 2, 3, 4)	(0, 1, 2, 3, 4)	(0, 1, 2, 3, 4)	(0, 1, 2, 3, 4)
(1, -1, -3, -1)	(0, 2, 3)	(0, 2, 3)	(2)	(0, 2, 3)	(0, 2, 3)	(0, 2)	(0, 2, 3)
(1, -3, -4, -2)	(0, 1, 2, 3, 4) (5, 6, 7)	(2, 3, 5, 6, 7)	(3, 4, 5, 6)	(1, 3, 4, 6)	(0, 1, 2, 3, 4) (5, 6)	(0, 1, 2, 3, 4) (5, 6)	(0, 1, 2, 3, 4) (5, 6, 7)
(2, 1, -4, 2)	(5, 6, 7)	(5, 6, 7)	(5, 6, 7)	(6, 7)	(5, 6, 7)	(5, 6, 7)	(5, 6, 7)
(2, 2, -4, 2)	(6, 7, 8, 9)	(7, 8, 9)	(6, 7, 8)	(7, 8, 9)	(6, 7, 8, 9)	(6, 7, 8)	(6, 7, 8, 9)
(1, -1, -4, -1)	(0, 3, 5)	(0, 3, 5)	(3)	(0, 3, 5)	(0, 3, 5)	(0, 3)	(0, 3, 5)
(1, 1, -3, 1)	(2, 3, 4)	(2, 3, 4)	(2, 3, 4)	(2, 3, 4)	(2, 3, 4)	(2, 3, 4)	(2, 3, 4)
(1, 1, -3, 0)	(1, 2, 3)	(1, 2, 3)	(2, 3)	(1, 2, 3)	(1, 2, 3)	(1, 2, 3)	(2, 3)
(1, -1, -2, 0)	(0, 1)	(0, 1)	(1)	(0, 1)	(1)	(0, 1)	(0, 1)
(1, 1, -3, 2)	(3, 4, 5)	(3, 4, 5)	(3, 4, 5)	(4, 5)	(3, 4, 5)	(3, 4, 5)	(3, 4, 5)

B.1.2 $D_C = (4; -1, -2, 1)$

For this (generically disjoint) union of a genus-0 and a genus-2 curve, we find:

bundle	h^0-values	E_1-splits	E_2-splits	E_3-splits	E_4-splits	E_5-splits	E_6-splits
(2, -1, -2, 5)	(2, 5, 7, 8)	(2, 5, 7, 8)	(2, 5, 7, 8)	(2)	(2, 5, 7, 8)	(5, 7, 8)	(5, 7, 8)
(1, -1, -2, -1)	(2, 3)	(2, 3)	(2, 3)	(2)	(2, 3)	(3)	(3)
(1, -2, -2, -2)	(3, 4, 5, 6, 7)	(4, 6, 7)	(3, 4, 5, 6, 7)	(3, 4)	(3, 4, 5, 6, 7)	(5, 6, 7)	(5, 6, 7)
(2, -3, -2, -1)	(2, 3, 4, 5, 6)	(4, 5)	(2, 3, 4, 5, 6)	(2, 3, 4, 5)	(4, 5, 6)	(3, 5, 6)	(3, 4, 5, 6)
(1, -2, -1, 4)	(0, 1, 2, 3)	(1, 2, 3)	(0, 1, 2, 3)	(0, 1)	(1, 2, 3)	(1, 2, 3)	(1, 2, 3)
(1, -2, -2, -3)	(4, 5, 7, 8, 9) (10, 11)	(5, 8, 10, 11)	(4, 5, 7, 8, 9) (10, 11)	(4,5)	(4, 5, 7, 8, 9) (10, 11)	(7, 8, 9, 10, 11)	(7, 8, 9, 10, 11)
(2, -3, -2, -2)	(3, 4, 5, 6, 7) (8, 9)	(5, 6, 7, 8)	(3, 5, 6, 7, 8, 9)	(3, 4, 5, 6)	(5, 6, 7, 8, 9)	(5, 6, 7, 8, 9)	(5, 6, 7, 8, 9)
(1, -2, 1, -1)	(5, 6)	(5, 6)	(5, 6)	(5)	(5, 6)	(6)	(5, 6)
(2, -2, -1, -2)	(6, 7)	(6, 7)	(6, 7)	(6)	(6, 7)	(7)	(6, 7)
(2, -2, -2, 7)	(1, 2, 6, 7, 10) (11, 13, 14, 15)	(2, 6, 7, 10, 11) (11, 13, 14, 15)	(1, 2, 6, 7, 10) (13, 14, 15)	(1, 2)	(2, 7, 11, 14)	(6, 7, 10, 11, 13) (13, 14, 15)	(6, 7, 10, 11, 13) (14)
(3, -1, -2, 10)	(6, 14, 21, 27, 32)	(6, 14, 21, 27, 32)	(6, 14, 21, 27, 32)	(6)	(6, 14, 21, 27)	(14, 21, 27, 32)	(14, 21, 27)
(1, -3, 1, -1)	(4, 5, 6, 7)	(4, 5, 6, 7)	(4, 5, 6, 7)	(4, 5, 6)	(4, 5, 6, 7)	(6, 7)	(4, 5, 6, 7)

B.1.3 $D_C = (4; -1, -2, -1)$

For this genus-2 curve we find:

bundle	h^0-values	E_1-splits	E_2-splits	E_3-splits	E_4-splits	E_5-splits	E_6-splits
(2, 3, -3, 1)	(5, 7, 8)	(7)	(5, 7, 8)	(5, 7, 8)	(5, 7, 8)	(5, 7, 8)	(5, 7, 8)
(3, 1, -4, -1)	(3, 4)	(3, 4)	(3, 4)	(3, 4)	(3, 4)	(3, 4)	(4)
(2, 2, -4, 0)	(1, 2, 3, 4)	(2, 3, 4)	(2, 3, 4)	(1, 2, 3, 4)	(1, 2, 3, 4)	(2, 3, 4)	(2, 3, 4)
(2, 1, -4, -3)	(0, 1, 2, 3, 4) (5, 6)	(0, 1, 2, 3, 4) (5)	(2, 3, 4, 5, 6)	(2, 3, 4, 5)	(0, 1, 2, 3, 4) (5, 6)	(0, 1, 2, 3, 4) (5, 6)	(1, 2, 3, 4)
(1, -1, -3, -2)	(0, 1, 2)	(0, 1, 2)	(1, 2)	(1, 2)	(0, 1, 2)	(0, 1, 2)	(0, 1, 2)
(1, -2, -4, 2)	(0, 1, 2, 3, 4)	(1, 2, 3, 4)	(2, 3, 4)	(0, 1, 2, 3, 4)	(0, 1, 2, 3)	(0, 1, 2, 3, 4)	(0, 1, 2, 3, 4)
(4, 3, -3, -8)	(4, 5, 6, 7, 8) (10, 12, 13, 15) (16, 17, 18, 19)	(6, 7, 8, 9, 10) (12, 13, 15, 16) (17, 18)	(4, 5, 6, 7, 8) (9, 10, 12, 13) (15, 17, 18, 19)	(7, 10, 12, 13, 15) (17)	(4, 5, 6, 7, 8) (9, 10, 12, 13, 15) (16, 17, 18, 19)	(4, 7, 9, 10, 12) (13, 15, 16, 18, 19)	(10, 12, 13, 15, 16)
(1, 3, -4, -5)	(0, 1, 2, 4, 6) (7, 8, 9, 11)	(0, 2, 4, 6, 7) (8, 9)	(2, 4, 6, 8, 9) (11)	(4, 6, 8)	(0, 1, 2, 4, 6) (7, 8, 9, 11)	(0, 2, 4, 6, 7) (9,11)	(0, 1, 2, 4, 6) (7)
(3, 1, -4, -5)	(0, 1, 2, 4, 5) (6, 7, 8, 9, 11)	(0, 1, 2, 4, 5) (6, 7, 8, 9)	(2, 4, 6, 8, 9) (11)	(4, 5, 6, 8)	(0, 1, 2, 4, 5) (6, 7, 8, 9, 11)	(0, 1, 2, 4, 5) (6, 7, 8, 9, 11)	(4, 5, 6, 7)
(3, 2, -3, -7)	(0, 1, 2, 3, 4) (6, 7, 9, 10, 11) (12, 14, 15, 16)	(1, 2, 3, 4, 6) (7, 9, 10, 11) (12, 14, 15)	(1, 3, 4, 6, 7) (7, 9, 10, 11) (12, 14, 15, 16)	(6, 7, 9, 10, 11) (12)	(1, 2, 3, 4, 6) (7, 9, 10, 11) (12, 14, 15)	(1, 3, 4, 6, 7) (9, 10, 11, 12) (14, 15, 16)	(6, 7, 9, 10, 11) (12)
(3, 2, -3, -5)	(2, 3, 4, 5, 6) (7, 8, 9, 10, 11)	(3, 4, 5, 7, 8) (10)	(2, 3, 4, 6, 7) (8, 9, 10, 11)	(4, 5, 6, 7, 8) (9,10)	(2, 3, 4, 5, 6) (7, 8, 9, 10, 11)	(2, 3, 4, 5, 6) (7, 8, 9, 10, 11)	(6, 7, 8)
(1, 1, -4, 2)	(0, 1, 2, 3, 4)	(0, 1, 2, 3, 4)	(2, 3, 4)	(0, 1, 2, 3, 4)	(0, 1, 2, 3, 4)	(2, 3, 4)	(0, 1, 2, 3, 4)
(1, 0, -4, -1)	(0, 2, 3)	(0, 2, 3)	(2)	(0, 2, 3)	(0, 2, 3)	(0, 2)	(0, 2, 3)
(3, -3, -1, -2)	(4, 5)	(4, 5)	(4, 5)	(4, 5)	(4, 5)	(4, 5)	(4, 5)
(4, -7, -1, -3)	(3, 4, 5, 6, 7) (8, 10, 11, 12) (13, 15, 17)	(3, 4, 5, 6, 7) (8, 10, 11, 12)	(3, 4, 5, 6, 7) (8, 10, 11, 13, 15)	(3, 4, 5, 6, 8) (10, 11, 12, 13) (15, 17)	(3, 4, 5, 6, 7) (8, 11, 13, 15)	(3, 4, 5, 6, 8) (10, 11, 13, 15)	(3, 4, 5, 6, 7) (10, 11, 12, 13, 15)

B.1.4 $D_C = (4; -1, -2, 0)$

For this genus-2 curve we find:

bundle	h^0-values	E_1-splits	E_2-splits	E_3-splits	E_4-splits	E_5-splits	E_6-splits
(1, -2, -1, 4)	(0, 1, 2, 3, 4, 5, 6)	(1, 2, 3, 4, 5, 6)	(0, 1, 2, 3, 4, 5, 6)	(2, 3)	(1, 3, 5, 6)	(2, 3, 4, 5, 6)	(2, 3, 4, 5, 6)

B.1.5 $D_C = (4; -1, -1, -1)$

On this genus-3 curve we find:

bundle	h^0-values	E_1-splits	E_2-splits	E_3-splits	E_4-splits	E_5-splits	E_6-splits
(1, -2, -3, -1)	(0, 1, 2, 3, 4)	(1, 3, 4)	(2, 3)	(0, 1, 2, 3, 4)	(0, 1, 2, 3, 4)	(0, 1, 2, 3, 4)	(0, 1, 2, 3, 4)
(1, -3, -4, -3)	(0, 2, 3, 4, 5) (6, 7, 8, 9, 10)	(2, 4, 5, 7, 8) (9, 10)	(3, 5, 6, 7, 8) (9)	(2, 4, 5, 7, 8) (9, 10)	(0, 2, 3, 4, 5) (6, 7, 8, 9, 10)	(0, 2, 3, 4, 5) (6, 7, 8, 9, 10)	(0, 2, 3, 4, 5) (6, 7, 8, 9, 10)
(1, 1, -3, 0)	(0, 1, 2, 3)	(0, 1, 2, 3)	(2)	(0, 1, 2, 3)	(1, 2, 3)	(0, 1, 2, 3)	(1, 2, 3)
(1, -3, -3, -3)	(0, 2, 3, 4, 5) (6, 7, 8)	(2, 4, 5, 6, 7) (8)	(2, 4, 5, 6, 7) (8)	(2, 4, 5, 6, 7) (8)	(0, 2, 3, 4, 5) (6, 7, 8)	(0, 2, 3, 4, 5) (6, 7, 8)	(0, 2, 3, 4, 5) (6, 7, 8)
(1, -3, -2, -3)	(0, 1, 2, 3, 4) (5, 6, 7)	(2, 3, 4, 5, 6) (7)	(1, 3, 4, 5, 6) (7)	(2, 3, 4, 5, 6)	(0, 1, 2, 3, 4) (5, 6, 7)	(0, 1, 2, 3, 4) (5, 6)	(0, 1, 2, 3, 4) (5, 6, 7)
(1, 2, -2, -1)	(1, 2, 3)	(2, 3)	(1, 2, 3)	(1, 2, 3)	(1, 2, 3)	(1, 2, 3)	(2, 3)
(1, 1, -3, -3)	(0, 1, 2, 3, 4) (5, 6)	(0, 1, 2, 3, 4) (5, 6)	(2, 4, 5)	(2, 4, 5)	(0, 2, 3, 4, 5) (6)	(0, 2, 3, 4, 5) (6)	(1, 2, 3, 4, 5)
(2, 3, -4, -1)	(4, 5, 6, 7, 8, 9)	(6, 7, 8, 9)	(4, 5, 6, 7, 8, 9)	(4, 5, 6, 7, 8, 9)	(4, 5, 6, 7, 8, 9)	(4, 5, 6, 7, 8, 9)	(6, 7, 8, 9)
(1, 2, -4, 2)	(2, 3, 4, 5, 6) (7, 8)	(3, 4, 5, 6, 7) (8)	(3, 4, 5, 6, 7) (8)	(3, 4, 5, 6, 7) (8)	(2, 3, 4, 5, 6) (7, 8)	(3, 4, 5, 6, 7) (8)	(2, 3, 4, 5, 6) (7, 8)
(1, -2, -3, -2)	(0, 1, 2, 3, 4, 5)	(1, 2, 3, 4, 5)	(2, 3, 4)	(1, 2, 3, 4, 5)	(0, 1, 2, 3, 4, 5)	(0, 1, 2, 3, 4, 5)	(0, 1, 2, 3, 4, 5)
(1, 3, -3, 1)	(3, 4, 5, 6, 7)	(5, 6, 7)	(3, 4, 5, 6, 7)	(3, 4, 5, 6, 7)	(3, 4, 5, 6, 7)	(3, 4, 5, 6, 7)	(3, 4, 5, 6, 7)
(1, -1, -3, 0)	(0, 1, 2, 3)	(0, 1, 2, 3)	(2)	(0, 1, 2, 3)	(1, 2, 3)	(0, 1, 2, 3)	(0, 2, 3)

B.1.6 $D_C = (5; -2, -2, -1)$

On this genus-4 curve we find:

bundle	h^0-values	E_1-splits	E_2-splits	E_3-splits	E_4-splits	E_5-splits	E_6-splits
(2, -2, -4, -2)	(0, 1, 2, 3, 4)	(0, 1, 2, 3, 4)	(2, 3)	(1, 3, 4)	(0, 1, 2, 3, 4)	(0, 1, 2, 3, 4)	(0, 1, 2, 3, 4)
(1, -1, -3, 0)	(0, 1)	(0, 1)	(1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)
(1, 2, -2, 0)	(2, 3)	(3)	(2, 3)	(2, 3)	(2, 3)	(2, 3)	(2, 3)
(1, 2, -2, 1)	(3, 4)	(4)	(3, 4)	(3, 4)	(3, 4)	(3, 4)	(3, 4)
(1, 1, -4, -1)	(0, 2, 3)	(0, 2, 3)	(2)	(0, 2, 3)	(0, 2, 3)	(0, 2, 3)	(0, 2, 3)
(1, -1, -4, -1)	(0, 2, 3)	(0, 2, 3)	(2)	(0, 2, 3)	(0, 2, 3)	(0, 2, 3)	(0, 2, 3)
(1, -2, -4, 2)	(0, 2, 3)	(0, 2, 3)	(2)	(0, 2, 3)	(0, 2, 3)	(0, 2, 3)	(0, 2, 3)
(1, 1, -4, 1)	(0, 1, 2, 3, 4)	(0, 1, 2, 3, 4)	(2, 3)	(0, 1, 2, 3, 4)	(0, 1, 2, 3, 4)	(1, 3, 4)	(0, 1, 2, 3, 4)
(1, -1, -2, 3)	(0, 1, 2, 3)	(0, 1, 2, 3)	(0, 1, 2, 3)	(1, 2)	(0, 1, 2, 3)	(1, 2, 3)	(0, 1, 2, 3)
(2, -1, -4, 1)	(0, 1, 2, 3)	(0, 1, 2, 3)	(2)	(0, 1, 2, 3)	(0, 1, 2, 3)	(0, 1, 2, 3)	(0, 1, 2, 3)
(1, 2, -3, 1)	(1, 2, 3, 4)	(2, 3, 4)	(1, 2, 3, 4)	(1, 2, 3, 4)	(1, 2, 3, 4)	(2, 3, 4)	(1, 2, 3, 4)
(1, 1, -4, 0)	(0, 2, 3)	(0, 2, 3)	(2)	(0, 2, 3)	(0, 2, 3)	(0, 2, 3)	(0, 2, 3)
(1, 2, -2, -2)	(0, 1, 2, 3, 4)	(1, 2, 3, 4)	(0, 1, 2, 3, 4)	(1, 2, 3, 4)	(0, 1, 2, 3, 4)	(0, 1, 2, 3, 4)	(2, 3)

B.1.7 $D_C = (5; -1, -1, -2)$

On this genus-5 curve we find:

bundle	h^0-values	E_1-splits	E_2-splits	E_3-splits	E_4-splits	E_5-splits	E_6-splits
(1, -2, -2, -3)	(0, 1, 2, 3)	(1, 2, 3)	(1, 2, 3)	(1, 2, 3)	(0, 1, 2, 3)	(0, 1, 2, 3)	(0, 1, 2, 3)
(1, 1, -4, 2)	(2, 3, 4, 5, 6) (7, 8)	(2, 3, 4, 5, 6) (7, 8)	(3, 4, 5, 6, 7)	(3, 4, 5, 6, 7) (8)	(3, 4, 5, 6, 7) (8)	(2, 3, 4, 5, 6) (7, 8)	(2, 3, 4, 5, 6) (7, 8)
(1, 1, -4, 1)	(0, 1, 2, 3, 4) (5, 6, 7)	(0, 1, 2, 3, 4) (5, 6, 7)	(3, 4)	(0, 1, 2, 3, 4) (5, 6, 7)	(1, 2, 3, 4, 5) (5, 6, 7)	(1, 2, 4, 6, 7)	(1, 2, 3, 4, 5) (6, 7)
(1, -1, -3, -2)	(0, 2, 3)	(0, 2, 3)	(2)	(0, 2, 3)	(0, 2, 3)	(0, 2, 3)	(0, 2, 3)
(1, 1, -3, -1)	(0, 1, 2, 3)	(0, 1, 2, 3)	(2)	(0, 1, 2, 3)	(0, 2, 3)	(0, 1, 2, 3)	(1, 2, 3)
(1, 1, -3, -2)	(0, 2, 3)	(0, 2, 3)	(2)	(0, 2, 3)	(0, 2, 3)	(0, 2, 3)	(0, 2, 3)
(1, 2, -2, -1)	(0, 1)	(0, 1)	(1)	(0, 1)	(0, 1)	(0, 1)	(1)
(1, 1, -4, 0)	(0, 1, 3, 5, 6)	(0, 1, 3, 5, 6)	(3)	(0, 1, 3, 5, 6)	(1, 3, 5, 6)	(0, 1, 3, 5, 6)	(1, 3, 5, 6)
(1, -2, -1, -3)	(0, 1, 2)	(1, 2)	(0, 1, 2)	(1, 2)	(0, 1, 2)	(0, 1, 2)	(0, 1, 2)
(1, 1, -3, 1)	(1, 2, 3, 4)	(1, 2, 3, 4)	(2, 3)	(1, 2, 3, 4)	(1, 2, 3, 4)	(1, 2, 3, 4)	(1, 2, 3, 4)
(1, -1, -2, -2)	(0, 1)	(0, 1)	(1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)
(1, -2, -3, -3)	(0, 1, 2, 3, 4, 5)	(1, 2, 3, 4, 5)	(2, 3, 4)	(1, 2, 3, 4, 5)	(0, 1, 2, 3, 4, 5)	(0, 1, 2, 3, 4, 5)	(0, 1, 2, 3, 4, 5)
(1, 1, -4, -1)	(0, 1, 3, 5, 6)	(0, 1, 3, 5, 6)	(3)	(0, 1, 3, 5, 6)	(0, 3, 5, 6)	(0, 1, 3, 5, 6)	(1, 3, 5, 6)

B.1.8 $D_C = (6; -3, -2, -1)$

On this genus-6 curve we find:

bundle	h^0-values	E_1-splits	E_2-splits	E_3-splits	E_4-splits	E_5-splits	E_6-splits
(1, 1, -4, 1)	(0, 1, 2, 3, 4)	(0, 1, 2, 3, 4)	(2, 3)	(0, 1, 2, 3, 4)	(0, 1, 2, 3, 4)	(1, 3, 4)	(0, 1, 2, 3, 4)
(1, 0, -3, 1)	(0, 1)	(0, 1)	(1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)

B.2 Local to global section counting applied to our database

In this section, we list results which quantify how good the counting procedure proposed in subsubsection 5.2.1 works, when applied to our database. We have preformed two tests:

1. We consider those curves in our data, for which we can quickly identify the exact number of sections on all curve components. This can be done quickly for non-split curves and for curves with only smooth components. For the latter curves, we have read-off the genus g and the line bundle degree d from our database. If $d < 0$, we know that there are no non-trivial sections on this curve component. However, if $d > 2g - 2$, then $h^0(C, \mathcal{L}) = d - g + 1$. Based on these exact local section counts, we have then tried to predict the number of global sections. The accuracy for this is listed in subsubsection B.2.1.

2. Our second test is based on our *H0Approximator*-program [44], which is part of [75]. This program considers curve degeneration, which split-off combinations of the 6 toric $\mathbb{P}^1$s in dP_3. For each such curve splitting, the program assumes that the number of local sections on each curve component is generic. Since this generic value is a lower bound to the actual number of local sections, we can use these estimates to derive a lower bound on the number of global sections. By repeating this strategy for many curve splittings, we obtain an estimate for the allowed h^0-values over the parameter space of the curve in question. We list the so-obtained results for all pairs (D_C, D_L) in our database [43] in subsubsection B.2.2.

B.2.1 Accuracy

D_C	D_L	Applicable data sets [%]	Accuracy [%]
(3, -1, -1, -1)	(1, 1, -3, 0)	62.2	100
(3, -1, -1, -1)	(1, 1, -3, 1)	71.6	100
(3, -1, -1, -1)	(1, 1, -3, 2)	52.7	100
(3, -1, -1, -1)	(1, -1, -2, 0)	52.7	100
(3, -1, -1, -1)	(1, -1, -3, 0)	66.9	100
(3, -1, -1, -1)	(1, -1, -3, -1)	76.4	100
(3, -1, -1, -1)	(1, -1, -4, -1)	76.4	100
(3, -1, -1, -1)	(1, -2, -3, -2)	90.5	100
(3, -1, -1, -1)	(1, -3, -3, -2)	90.5	100
(3, -1, -1, -1)	(1, -3, -4, -2)	90.5	100
(3, -1, -1, -1)	(2, 1, -4, 1)	62.2	100
(3, -1, -1, -1)	(2, 1, -4, 2)	48.0	100
(3, -1, -1, -1)	(2, 2, -4, 2)	37.0	100

Continued on next page.

Table 2 — *Continued from previous page.*

$\mathbf{D_C}$	$\mathbf{D_L}$	Applicable data sets [%]	Accuracy [%]
(4, -1, -2, 1)	(1, -1, -2, 0)	38.7	100
(4, -1, -2, 1)	(1, -1, -2, -1)	38.7	100
(4, -1, -2, 1)	(1, -2, 1, -1)	26.9	100
(4, -1, -2, 1)	(1, -2, -1, 4)	12.6	65.1
(4, -1, -2, 1)	(1, -2, -2, -2)	43.4	100
(4, -1, -2, 1)	(1, -2, -2, -3)	43.4	100
(4, -1, -2, 1)	(1, -3, 1, -1)	9.2	100
(4, -1, -2, 1)	(2, -1, -2, 5)	4.3	100
(4, -1, -2, 1)	(2, -2, -1, -2)	28.3	100
(4, -1, -2, 1)	(2, -2, -2, 7)	4.4	100
(4, -1, -2, 1)	(2, -3, -2, -1)	12.6	100
(4, -1, -2, 1)	(2, -3, -2, -2)	12.6	100
(4, -1, -2, 1)	(3, -1, -2, 10)	23.9	100
(4, -1, -2, -1)	(1, 0, -4, -1)	80.4	100
(4, -1, -2, -1)	(1, 3, -4, -5)	83.4	99
(4, -1, -2, -1)	(1, -1, -3, -2)	88.3	100
(4, -1, -2, -1)	(1, -2, -4, 2)	84.2	100
(4, -1, -2, -1)	(3, 2, -3, -7)	71.8	100
(4, -1, -2, -1)	(2, 1, -4, -3)	76.4	100
(4, -1, -2, -1)	(2, 2, -4, 0)	50.6	100
(4, -1, -2, -1)	(2, 3, -3, 1)	44.8	100
(4, -1, -2, -1)	(3, 1, -4, -1)	45.4	100
(4, -1, -2, -1)	(3, 1, -4, -5)	69.4	100
(4, -1, -2, -1)	(3, 2, -3, -5)	54.3	100
(4, -1, -2, -1)	(1, 1, -4, 2)	76.3	98.6
(4, -1, -2, -1)	(4, 3, -3, -8)	60.6	100
(4, -1, -2, -1)	(3, -3, -1, -2)	66.5	98.7
(4, -1, -2, -1)	(4, -7, -1, -3)	74.1	92.6
(4, -1, -2, 0)	(1, -2, -1, 4)	58.7	92.5
(4, -1, -1, -1)	(1, 1, -3, 0)	52.2	95.8
(4, -1, -1, -1)	(1, 1, -3, -1)	56.5	100
(4, -1, -1, -1)	(1, 1, -3, -3)	73.8	100
(4, -1, -1, -1)	(1, 2, -2, -1)	45.6	100

Table 2 — *Continued from previous page.*

D_C	D_L	Applicable data sets [%]	Accuracy [%]
(4, -1, -1, -1)	(1, 2, -4, 2)	65.3	100
(4, -1, -1, -1)	(1, 3, -3, 1)	56.9	100
(4, -1, -1, -1)	(1, -1, -3, 0)	64.3	96.6
(4, -1, -1, -1)	(1, -2, -3, -1)	82.4	100
(4, -1, -1, -1)	(1, -2, -3, -2)	87.5	100
(4, -1, -1, -1)	(1, -3, -2, -3)	85.8	100
(4, -1, -1, -1)	(1, -3, -3, -3)	84.0	100
(4, -1, -1, -1)	(1, -3, -4, -3)	86.2	100
(4, -1, -1, -1)	(2, 3, -4, -1)	45.5	100
(5, -2, -2, -1)	(1, 1, -4, 0)	62.0	100
(5, -2, -2, -1)	(1, 1, -4, 1)	58.4	99.7
(5, -2, -2, -1)	(1, 1, -4, -1)	67.9	100
(5, -2, -2, -1)	(1, 2, -2, 0)	45.2	100
(5, -2, -2, -1)	(1, 2, -2, 1)	50.8	100
(5, -2, -2, -1)	(1, 2, -2, -2)	46.7	98.9
(5, -2, -2, -1)	(1, 2, -3, 1)	48.9	99.0
(5, -2, -2, -1)	(1, -1, -2, 3)	45.1	99.9
(5, -2, -2, -1)	(1, -1, -3, 0)	72.7	100
(5, -2, -2, -1)	(1, -1, -4, -1)	88.6	100
(5, -2, -2, -1)	(1, -2, -4, 2)	77.3	100
(5, -2, -2, -1)	(2, -1, -4, 1)	51.4	97.9
(5, -2, -2, -1)	(2, -2, -4, -2)	75.2	100
(5, -1, -1, -2)	(1, -2, -2, -3)	88.1	100
(5, -1, -1, -2)	(1, -2, -1, -3)	84.4	100
(5, -1, -1, -2)	(1, -1, -3, -2)	82.7	100
(5, -1, -1, -2)	(1, -1, -2, -2)	79.3	100
(5, -1, -1, -2)	(1, 2, -2, -1)	42.1	100
(5, -1, -1, -2)	(1, 1, -4, 2)	54.3	99.2
(5, -1, -1, -2)	(1, 1, -4, 1)	47.5	99.2
(5, -1, -1, -2)	(1, 1, -4, 0)	56.8	95.0
(5, -1, -1, -2)	(1, 1, -3, -2)	65.9	100
(5, -1, -1, -2)	(1, 1, -3, -1)	55.5	98.6
(5, -1, -1, -2)	(1, 1, -3, 1)	46.1	99.4

Table 2 — *Continued from previous page.*

D_C	D_L	Applicable data sets [%]	Accuracy [%]
(5, -1, -1, -2)	(1, -2, -3, -3)	88.1	100
(5, -1, -1, -2)	(1, 1, -4, -1)	64.4	98.8
(6, -3, -2, -1)	(1, 0, -3, 1)	51.8	100
(6, -3, -2, -1)	(1, 1, -4, 1)	52.4	99.7

Table 2. Accuracy of counting procedure for exact numbers of local sections.

B.2.2 Spectrum estimate

D_C	D_L	Predicted spectrum	Missing values
(5, -2, -2, -1)	(2, -2, -4, -2)	(0, 1, 2, 3, 4)	—
(5, -2, -2, -1)	(1, -1, -3, 0)	(0, 1)	—
(5, -2, -2, -1)	(1, 2, -2, 0)	(2, 3)	—
(5, -2, -2, -1)	(1, 2, -2, 1)	(3, 4)	—
(5, -2, -2, -1)	(1, 1, -4, -1)	(0, 2, 3)	—
(5, -2, -2, -1)	(1, -1, -4, -1)	(0, 2, 3)	—
(5, -2, -2, -1)	(1, -2, -4, 2)	(0, 2, 3)	—
(5, -2, -2, -1)	(1, 1, -4, 1)	(0, 1, 2, 3, 4)	—
(5, -2, -2, -1)	(1, -1, -2, 3)	(0, 1, 2, 3)	—
(5, -2, -2, -1)	(2, -1, -4, 1)	(0, 1, 2, 3)	—
(5, -2, -2, -1)	(1, 2, -3, 1)	(1, 2, 3, 4)	—
(5, -2, -2, -1)	(1, 1, -4, 0)	(0, 2, 3)	—
(5, -2, -2, -1)	(1, 2, -2, -2)	(0, 1, 2, 3, 4)	—
(3, -1, -1, -1)	(2, 1, -4, 1)	(4, 5, 6)	—
(3, -1, -1, -1)	(1, -3, -3, -2)	(0, 1, 2, 3, 4)	(5)
(3, -1, -1, -1)	(1, -1, -3, 0)	(0, 1, 2, 3)	—
(3, -1, -1, -1)	(1, -2, -3, -2)	(0, 1, 2, 3)	(4)
(3, -1, -1, -1)	(1, -1, -3, -1)	(0, 2, 3)	—
(3, -1, -1, -1)	(1, -3, -4, -2)	(0, 1, 2, 3, 4, 5)	(6, 7)
(3, -1, -1, -1)	(2, 1, -4, 2)	(5, 6, 7)	—
(3, -1, -1, -1)	(2, 2, -4, 2)	(6, 7, 8, 9)	—
(3, -1, -1, -1)	(1, -1, -4, -1)	(0, 3, 5)	—
(3, -1, -1, -1)	(1, 1, -3, 1)	(2, 3, 4)	—
(3, -1, -1, -1)	(1, 1, -3, 0)	(1, 2, 3)	—
(3, -1, -1, -1)	(1, -1, -2, 0)	(0, 1)	—
(3, -1, -1, -1)	(1, 1, -3, 2)	(3, 4, 5)	—

Continued on next page.

Table 3 — *Continued from previous page.*

D_C	D_L	Predicted spectrum	Missing values
(5, -1, -1, -2)	(1, -2, -2, -3)	(0, 1, 2, 3)	—
(5, -1, -1, -2)	(1, 1, -4, 2)	(2, 3, 4, 5, 6, 7, 8)	—
(5, -1, -1, -2)	(1, 1, -4, 1)	(0, 1, 2, 3, 4, 5, 6, 7)	—
(5, -1, -1, -2)	(1, -1, -3, -2)	(0, 2, 3)	—
(5, -1, -1, -2)	(1, 1, -3, -1)	(0, 2, 3)	(1)
(5, -1, -1, -2)	(1, 1, -3, -2)	(0, 2, 3)	—
(5, -1, -1, -2)	(1, 2, -2, -1)	(0, 1)	—
(5, -1, -1, -2)	(1, 1, -4, 0)	(0, 3, 5, 6)	(1)
(5, -1, -1, -2)	(1, -2, -1, -3)	(0, 1, 2)	—
(5, -1, -1, -2)	(1, 1, -3, 1)	(1, 2, 3, 4)	—
(5, -1, -1, -2)	(1, -1, -2, -2)	(0, 1)	—
(5, -1, -1, -2)	(1, -2, -3, -3)	(0, 1, 2, 3, 4, 5)	—
(5, -1, -1, -2)	(1, 1, -4, -1)	(0, 3, 5, 6)	(1)
(4, -1, -1, -1)	(1, -2, -3, -1)	(0, 1, 2, 3, 4)	—
(4, -1, -1, -1)	(1, -3, -4, -3)	(0, 2, 3, 4, 5, 6, 7, 8, 9)	(10)
(4, -1, -1, -1)	(1, 1, -3, 0)	(0, 1, 2, 3)	—
(4, -1, -1, -1)	(1, -3, -3, -3)	(0, 2, 3, 4, 5, 6, 7)	(8)
(4, -1, -1, -1)	(1, -3, -2, -3)	(0, 1, 2, 3, 4, 5, 6)	(7)
(4, -1, -1, -1)	(1, 2, -2, -1)	(1, 2, 3)	—
(4, -1, -1, -1)	(1, 1, -3, -3)	(0, 1, 2, 3, 4, 5, 6)	—
(4, -1, -1, -1)	(1, 1, -3, -3)	(0, 1, 2, 3, 4, 5, 6)	—
(4, -1, -1, -1)	(2, 3, -4, -1)	(4, 5, 6, 7, 8, 9)	—
(4, -1, -1, -1)	(1, 2, -4, 2)	(2, 3, 4, 5, 6, 7, 8)	—
(4, -1, -1, -1)	(1, -2, -3, -2)	(0, 1, 2, 3, 4, 5)	—
(4, -1, -1, -1)	(1, 3, -3, 1)	(3, 4, 5, 6, 7)	—
(4, -1, -1, -1)	(1, -1, -3, 0)	(0, 2, 3)	(1)
(4, -1, -2, 0)	(1, -2, -1, 4)	(0, 1, 2, 3, 4, 5, 6)	—
(4, -1, -2, 1)	(2, -1, -2, 5)	(2, 5, 7, 8)	—
(4, -1, -2, 1)	(1, -1, -2, -1)	(2, 3)	—
(4, -1, -2, 1)	(1, -2, -2, -2)	(3, 4, 5, 6, 7)	—
(4, -1, -2, 1)	(2, -3, -2, -1)	(2, 3, 4, 5, 6)	—
(4, -1, -2, 1)	(1, -2, -1, 4)	(0, 1, 2, 3)	—
(4, -1, -2, 1)	(1, -2, -2, -3)	(4, 5, 7, 8, 9, 10, 11)	—
(4, -1, -2, 1)	(2, -3, -2, -2)	(3, 5, 6, 7, 8, 9)	(4)
(4, -1, -2, 1)	(1, -2, 1, -1)	(5, 6)	—
(4, -1, -2, 1)	(2, -2, -1, -2)	(6, 7)	—
(4, -1, -2, 1)	(2, -2, -2, 7)	(1, 2, 6, 7, 10, 11, 13, 14, 15)	—

Table 3 — *Continued from previous page.*

$\mathbf{D_C}$	$\mathbf{D_L}$	Predicted spectrum	Missing values
(4, -1, -2, 1)	(3, -1, -2, 10)	(6, 14, 21, 27, 32)	—
(4, -1, -2, 1)	(1, -3, 1, -1)	(4, 5, 6, 7)	—
(6, -3, -2, -1)	(1, 1, -4, 1)	(0, 1, 2, 3, 4)	—
(6, -3, -2, -1)	(1, 0, -3, 1)	(0, 1)	—
(4, -1, -2, -1)	(3, -3, -1, -2)	(4, 5)	—
(4, -1, -2, -1)	(4, -7, -1, -3)	(3, 6, 8, 11, 12, 13, 15)	(4, 5, 7, 10, 17)
(4, -1, -2, -1)	(2, 3, -3, 1)	(5, 7, 8)	—
(4, -1, -2, -1)	(3, 1, -4, -1)	(3, 4)	—
(4, -1, -2, -1)	(2, 2, -4, 0)	(1, 2, 3, 4)	—
(4, -1, -2, -1)	(2, 1, -4, -3)	(0, 1, 2, 3, 4, 5, 6)	—
(4, -1, -2, -1)	(1, -1, -3, -2)	(0, 1, 2)	—
(4, -1, -2, -1)	(1, -2, -4, 2)	(0, 1, 2, 3, 4)	—
(4, -1, -2, -1)	(4, 3, -3, -8)	(4, 6, 7, 9, 10, 12, 13, 15, 16, 17, 18, 19)	(5, 8)
(4, -1, -2, -1)	(1, 3, -4, -5)	(0, 2, 4, 6, 7, 8, 9, 11)	(1)
(4, -1, -2, -1)	(3, 1, -4, -5)	(0, 2, 4, 5, 6, 7, 8, 9, 11)	(1)
(4, -1, -2, -1)	(3, 2, -3, -7)	(0, 1, 3, 4, 6, 7, 9, 10, 11, 12, 14, 15)	(2, 16)
(4, -1, -2, -1)	(3, 2, -3, -5)	(2, 3, 4, 5, 6, 7, 8, 9, 10, 11)	—
(4, -1, -2, -1)	(1, 1, -4, 2)	(0, 1, 2, 3, 4)	—
(4, -1, -2, -1)	(1, 0, -4, -1)	(0, 2, 3)	—

Table 3. Spectrum estimates from the *H0Approximator*.

References

[1] P. Berglund, T. Hubsch and L. Parkes, *Gauge neutral matter in three generation superstring compactifications*, *Mod. Phys. Lett. A* **5** (1990) 1485 [ɪɴSPIRE].

[2] P. Berglund and T. Hubsch, *Twisted three generation compactification*, *Phys. Lett. B* **260** (1991) 32 [ɪɴSPIRE].

[3] R. Donagi, Y.-H. He, B.A. Ovrut and R. Reinbacher, *Moduli dependent spectra of heterotic compactifications*, *Phys. Lett. B* **598** (2004) 279 [hep-th/0403291] [ɪɴSPIRE].

[4] R. Donagi, Y.-H. He, B.A. Ovrut and R. Reinbacher, *The particle spectrum of heterotic compactifications*, *JHEP* **12** (2004) 054 [hep-th/0405014] [ɪɴSPIRE].

[5] V. Bouchard and R. Donagi, *An* SU(5) *heterotic standard model*, *Phys. Lett. B* **633** (2006) 783 [hep-th/0512149] [ɪɴSPIRE].

[6] V. Bouchard, M. Cvetič and R. Donagi, *Tri-linear couplings in an heterotic minimal supersymmetric standard model*, *Nucl. Phys. B* **745** (2006) 62 [hep-th/0602096] [INSPIRE].

[7] E.I. Buchbinder, A. Lukas, B.A. Ovrut and F. Ruehle, *Heterotic instantons for monad and extension bundles*, *JHEP* **02** (2020) 081 [arXiv:1912.07222] [INSPIRE].

[8] E.I. Buchbinder, A. Lukas, B.A. Ovrut and F. Ruehle, *Instantons and Hilbert functions*, *Phys. Rev. D* **102** (2020) 026019 [arXiv:1912.08358] [INSPIRE].

[9] J. Gray and J. Wang, *Jumping spectra and vanishing couplings in heterotic line bundle Standard Models*, *JHEP* **11** (2019) 073 [arXiv:1906.09373] [INSPIRE].

[10] C. Vafa, *Evidence for F-theory*, *Nucl. Phys. B* **469** (1996) 403 [hep-th/9602022] [INSPIRE].

[11] T. Watari, *Vector-like pairs and Brill-Noether theory*, *Phys. Lett. B* **762** (2016) 145 [arXiv:1608.00248] [INSPIRE].

[12] R. Donagi and M. Wijnholt, *Model building with F-theory*, *Adv. Theor. Math. Phys.* **15** (2011) 1237 [arXiv:0802.2969] [INSPIRE].

[13] A.P. Braun, A. Collinucci and R. Valandro, *G-flux in F-theory and algebraic cycles*, *Nucl. Phys. B* **856** (2012) 129 [arXiv:1107.5337] [INSPIRE].

[14] J. Marsano and S. Schäfer-Nameki, *Yukawas, G-flux, and spectral covers from resolved Calabi-Yau's*, *JHEP* **11** (2011) 098 [arXiv:1108.1794] [INSPIRE].

[15] T.W. Grimm and H. Hayashi, *F-theory fluxes, chirality and Chern-Simons theories*, *JHEP* **03** (2012) 027 [arXiv:1111.1232] [INSPIRE].

[16] S. Krause, C. Mayrhofer and T. Weigand, G_4 *flux, chiral matter and singularity resolution in F-theory compactifications*, *Nucl. Phys. B* **858** (2012) 1 [arXiv:1109.3454] [INSPIRE].

[17] S. Krause, C. Mayrhofer and T. Weigand, *Gauge fluxes in F-theory and type IIB orientifolds*, *JHEP* **08** (2012) 119 [arXiv:1202.3138] [INSPIRE].

[18] C. Mayrhofer, E. Palti and T. Weigand, *Hypercharge flux in IIB and F-theory: anomalies and gauge coupling unification*, *JHEP* **09** (2013) 082 [arXiv:1303.3589] [INSPIRE].

[19] V. Braun, T.W. Grimm and J. Keitel, *Geometric engineering in toric F-theory and GUTs with* U(1) *gauge factors*, *JHEP* **12** (2013) 069 [arXiv:1306.0577] [INSPIRE].

[20] M. Cvetič, A. Grassi, D. Klevers and H. Piragua, *Chiral four-dimensional F-theory compactifications with* SU(5) *and multiple* U(1)-*factors*, *JHEP* **04** (2014) 010 [arXiv:1306.3987] [INSPIRE].

[21] M. Cvetič, D. Klevers, D.K.M. Peña, P.-K. Oehlmann and J. Reuter, *Three-family particle physics models from global F-theory compactifications*, *JHEP* **08** (2015) 087 [arXiv:1503.02068] [INSPIRE].

[22] L. Lin, C. Mayrhofer, O. Till and T. Weigand, *Fluxes in F-theory compactifications on genus-one fibrations*, *JHEP* **01** (2016) 098 [arXiv:1508.00162] [INSPIRE].

[23] L. Lin and T. Weigand, G_4-*flux and standard model vacua in F-theory*, *Nucl. Phys. B* **913** (2016) 209 [arXiv:1604.04292] [INSPIRE].

[24] M. Cvetič, L. Lin, M. Liu and P.-K. Oehlmann, *An F-theory realization of the chiral MSSM with* Z_2-*parity*, *JHEP* **09** (2018) 089 [arXiv:1807.01320] [INSPIRE].

[25] M. Cvetič, J. Halverson, L. Lin, M. Liu and J. Tian, *Quadrillion F-theory compactifications with the exact chiral spectrum of the Standard Model*, *Phys. Rev. Lett.* **123** (2019) 101601 [arXiv:1903.00009] [INSPIRE].

[26] G. Curio and R.Y. Donagi, *Moduli in N = 1 heterotic/F-theory duality*, *Nucl. Phys. B* **518** (1998) 603 [hep-th/9801057] [INSPIRE].

[27] R. Donagi and M. Wijnholt, *Gluing branes, I*, *JHEP* **05** (2013) 068 [arXiv:1104.2610] [INSPIRE].

[28] R. Donagi and M. Wijnholt, *Gluing branes II: flavour physics and string duality*, *JHEP* **05** (2013) 092 [arXiv:1112.4854] [INSPIRE].

[29] L.B. Anderson, J.J. Heckman and S. Katz, *T-branes and geometry*, *JHEP* **05** (2014) 080 [arXiv:1310.1931] [INSPIRE].

[30] M. Bies, C. Mayrhofer, C. Pehle and T. Weigand, *Chow groups, Deligne cohomology and massless matter in F-theory*, arXiv:1402.5144 [INSPIRE].

[31] M. Bies, C. Mayrhofer and T. Weigand, *Gauge backgrounds and zero-mode counting in F-theory*, *JHEP* **11** (2017) 081 [arXiv:1706.04616] [INSPIRE].

[32] M. Bies, *Cohomologies of coherent sheaves and massless spectra in F-theory*, Ph.D. thesis, Heidelberg U., Heidelberg, Germany (2018) [arXiv:1802.08860] [INSPIRE].

[33] Y.-H. He, *Deep-learning the landscape*, arXiv:1706.02714 [INSPIRE].

[34] D. Krefl and R.-K. Seong, *Machine learning of Calabi-Yau volumes*, *Phys. Rev. D* **96** (2017) 066014 [arXiv:1706.03346] [INSPIRE].

[35] F. Ruehle, *Evolving neural networks with genetic algorithms to study the string landscape*, *JHEP* **08** (2017) 038 [arXiv:1706.07024] [INSPIRE].

[36] J. Carifio, J. Halverson, D. Krioukov and B.D. Nelson, *Machine learning in the string landscape*, *JHEP* **09** (2017) 157 [arXiv:1707.00655] [INSPIRE].

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

[38] D. Klaewer and L. Schlechter, *Machine learning line bundle cohomologies of hypersurfaces in toric varieties*, *Phys. Lett. B* **789** (2019) 438 [arXiv:1809.02547] [INSPIRE].

[39] C.R. Brodie, A. Constantin, R. Deen and A. Lukas, *Machine learning line bundle cohomology*, *Fortsch. Phys.* **68** (2020) 1900087 [arXiv:1906.08730] [INSPIRE].

[40] C.R. Brodie, A. Constantin, R. Deen and A. Lukas, *Topological formulae for the zeroth cohomology of line bundles on surfaces*, arXiv:1906.08363 [INSPIRE].

[41] C.R. Brodie, A. Constantin, R. Deen and A. Lukas, *Index formulae for line bundle cohomology on complex surfaces*, *Fortsch. Phys.* **68** (2020) 1900086 [arXiv:1906.08769] [INSPIRE].

[42] M. Bies, *SheafCohomologyOnToricVarieties*, https://github.com/homalg-project/ToricVarieties_project/tree/master/SheafCohomologyOnToricVarieties, (2020).

[43] M. Bies, M. Cvetič, R. Donagi, L. Lin, M. Liu and F. Rühle, *Database*, https://github.com/Learning-line-bundle-cohomology/Database, (2020).

[44] M. Bies, *H0Approximator*, https://github.com/homalg-project/ToricVarieties_project/tree/master/H0Approximator, (2020).

[45] T. Weigand, *F-theory*, PoS(TASI2017)016 (2018) [arXiv:1806.01854] [INSPIRE].

[46] M. Cvetič and L. Lin, *TASI lectures on Abelian and discrete symmetries in F-theory*, `PoS(TASI2017)020` (2018) [`arXiv:1809.00012`] [ɪɴSPIRE].

[47] A. Brill and M. Nöther, *Ueber die algebraischen Functionen und ihre Anwendung in der Geometrie* (in German), *Math. Annalen* **7** (1874) 269.

[48] D. Mumford, *Curves and their Jacobians*, in *The Ziwet lectures*, University of Michigan Press, Ann Arbor, MI, U.S.A. (1975).

[49] P.A. Griffiths and J. Harris, *Principles of algebraic geometry*, Wiley classics library, Wiley, New York, NY, U.S.A. (1994).

[50] P. Griffiths and J. Harris, *On the variety of special linear systems on a general algebraic curve*, *Duke Math. J.* **47** (1980) 233.

[51] I. Bacharach, *Ueber den Cayley'schen Schnittpunktsatz* (in German), *Math. Annalen* **26** (1886) 275.

[52] A. Cayley, *On the intersection of curves*, Cambridge University Press, Cambridge, U.K. (1889).

[53] D. Eisenbud, M. Green and J. Harris, *Cayley-Bacharach theorems and conjectures*, *Bull. Amer. Math. Soc.* **33** (1996) 295.

[54] D.R. Grayson and M.E. Stillman, *Macaulay2, a software system for research in algebraic geometry*, http://www.math.uiuc.edu/Macaulay2/.

[55] W. Bosma, J. Cannon and C. Playoust, *The Magma algebra system. I: the user language*, *J. Symbol. Comput.* **24** (1997) 235.

[56] E. Witten, *On flux quantization in M-theory and the effective action*, *J. Geom. Phys.* **22** (1997) 1 [`hep-th/9609122`] [ɪɴSPIRE].

[57] B.R. Greene, D.R. Morrison and M.R. Plesser, *Mirror manifolds in higher dimension*, *Commun. Math. Phys.* **173** (1995) 559 [`hep-th/9402119`] [ɪɴSPIRE].

[58] A.P. Braun and T. Watari, *The vertical, the horizontal and the rest: anatomy of the middle cohomology of Calabi-Yau fourfolds and F-theory applications*, *JHEP* **01** (2015) 047 [`arXiv:1408.6167`] [ɪɴSPIRE].

[59] E. Freitag, *Complex analysis* 2*: Riemann surfaces, several complex variables, Abelian functions, higher modular functions*, Springer, Berlin, Heidelberg, Germany (2011).

[60] D. Cox, J. Little and H. Schenck, *Toric varieties*, American Mathematical Society, U.S.A. (2011).

[61] R. Blumenhagen, B. Jurke, T. Rahn and H. Roschy, *Cohomology of line bundles: a computational algorithm*, *J. Math. Phys.* **51** (2010) 103525 [`arXiv:1003.5217`] [ɪɴSPIRE].

[62] S.-Y. Jow, *Cohomology of toric line bundles via simplicial Alexander duality*, *J. Math. Phys.* **52** (2011) 033506 [`arXiv:1006.0780`].

[63] R. Blumenhagen, B. Jurke, T. Rahn and H. Roschy, *Cohomology of line bundles: applications*, *J. Math. Phys.* **53** (2012) 012302 [`arXiv:1010.3717`] [ɪɴSPIRE].

[64] R. Blumenhagen, B. Jurke and T. Rahn, *Computational tools for cohomology of toric varieties*, *Adv. High Energy Phys.* **2011** (2011) 152749 [`arXiv:1104.1187`] [ɪɴSPIRE].

[65] B. Jurke, T. Rahn, R. Blumenhagen and H. Roschy, *cohomCalg ++ Koszul extension*, manual v0.31, May 2011.

[66] *cohomCalg package*, https://github.com/BenjaminJurke/cohomCalg, (2010).

[67] T. Rahn and H. Roschy, *Cohomology of line bundles: proof of the algorithm*, *J. Math. Phys.* **51** (2010) 103520 [arXiv:1006.2392] [INSPIRE].

[68] L.B. Anderson, Y.-H. He and A. Lukas, *Heterotic compactification, an algorithmic approach*, *JHEP* **07** (2007) 049 [hep-th/0702210] [INSPIRE].

[69] L.B. Anderson, *Heterotic and M-theory compactifications for string phenomenology*, arXiv:0808.3621 [INSPIRE].

[70] L.B. Anderson, J. Gray, Y.-H. He and A. Lukas, *Exploring positive monad bundles and a new heterotic Standard Model*, *JHEP* **02** (2010) 054 [arXiv:0911.1569] [INSPIRE].

[71] L.B. Anderson, J. Gray, A. Lukas and E. Palti, *Two hundred heterotic standard models on smooth Calabi-Yau threefolds*, *Phys. Rev. D* **84** (2011) 106005 [arXiv:1106.4804] [INSPIRE].

[72] L.B. Anderson, J. Gray, A. Lukas and E. Palti, *Heterotic line bundle standard models*, *JHEP* **06** (2012) 113 [arXiv:1202.1757] [INSPIRE].

[73] The Sage developers, *SageMath, the Sage Mathematics Software System, version* 8.5, (2018).

[74] R. Hartshorne, *Algebraic geometry*, Springer, New York, NY, U.S.A. (1977).

[75] The Toric varieties project authors, *The* Toric Varieties *project*, https://github.com/homalg-project/ToricVarieties_project, (2019)–(2020).