Hilbert Series, Machine Learning, and Applications to Physics

Jiakang Bao,1,∗ Yang-Hui He,1,2,3,† Edward Hirst,1,‡ Johannes Hofscheier,4,§ Alexander Kasprzyk,4,¶ and Suvajit Majumder1,∗∗

1Department of Mathematics, City, University of London, EC1V 0HB, UK
2Merton College, University of Oxford, OX1 4JD, UK
3School of Physics, NanKai University, Tianjin, 300071, P.R. China
4School of Mathematical Sciences, University of Nottingham, Nottingham, NG7 2RD, UK

We describe how simple machine learning methods successfully predict geometric properties from Hilbert series (HS). Regressors predict embedding weights in projective space to ∼1 mean absolute error, whilst classifiers predict dimension and Gorenstein index to > 90% accuracy with ∼0.5% standard error. Binary random forest classifiers managed to distinguish whether the underlying HS describes a complete intersection with high accuracies exceeding 95%. Neural networks (NNs) exhibited success identifying HS from a Gorenstein ring to the same order of accuracy, whilst generation of “fake” HS proved trivial for NNs to distinguish from those associated to the three-dimensional Fano varieties considered.

I. INTRODUCTION AND SUMMARY

The Hilbert series (HS) is an important invariant in the study of modern geometry. In physics, HS have recently become a powerful tool in high energy theory, appearing, for example, in the study of: Bogo

In parallel, a programme to use machine learning (ML) techniques to study mathematical structures has recently been proposed [19–21]. The initial studies were inspired by timely and independent works [19, 22–25]. In these, the effectiveness of ML regressor and classifier techniques in various branches of mathematics and mathematical physics has been investigated. Applications of ML include: finding bundle cohomology on varieties [24, 26, 27]; distinguishing elliptic fibrations [28] and invariants of Calabi–Yau threefolds [29]; the Donaldson algorithm for numerical Calabi–Yau metrics [30]; the algebraic structures of groups and rings [31]; arithmetic geometry and number theory [32–34]; quiver gauge theories and cluster algebras [35]; patterns in particle masses [36]; statistical predictions and model-building in string theory [37–39]; and classifying combinatorial properties of finite graphs [40]. Here we apply ML techniques to the plethystic programme of using Hilbert series to understand structures of quantum field theory. The physical motivation for this work has two primary applications. First, when considering a generic supersymmetric quantum field theory the number of BPS operators at each order is given by the initial terms in the Hilbert series. Computing these operator frequencies requires significant computational power, particularly for higher order terms (for the multi-trace case the growth is exponential). In this work the goal for the machine learning techniques implemented is to return information about the full series’ closed form, which can then directly provide the higher order information, hence bypassing the need for order-by-order computation. Second, from a string perspective the geometry of the moduli space has an array of physical applications and if these techniques can return the underlying variety’s geometric properties directly the vacuum can be analysed without need for complete information about the theory.

We examined databases of HS arising in geometry – see [41, 42] and the Graded Ring Database (GRDB) [43] – and “fake” HS generated to imitate the “real” geometric HS. Simple ML methods were able to successfully predict several geometric quantities associated to the HS, and were able to accurately distinguish real from fake HS.

Depending on the form of the HS, simple regression neural networks (NNs) managed to learn the embedding weights in projective space to mean absolute error (MAE) ∼1; whilst classification NNs predicted the dimension and Gorenstein index with both accuracy and Matthews correlation coefficient (MCC) in excess of 0.9.

Motivated by the question of whether ML can detect when a HS comes from a Gorenstein ring, we found that binary classifiers identified whether a fake HS had a palindromic numerator to accuracy and MCC greater than 0.9. Binary classifiers were easily able to distinguish the fake generated data from the dataset of HS associated to three-dimensional Fano varieties obtained from [43, 44].

A random forest classifier correctly predicted whether the HS described a complete intersection (CI); this was achieved with accuracy 0.9 and MCC 0.8 when the numerator (padded with 0’s) of the HS was used as input; and with accuracy 0.95 and MCC greater than 0.9 when the Taylor series (to order 100) of the HS was used.

Code scripts for these investigations, along with the datasets generated and analysed, are available from:
We write
\[ \text{dim}_C R_i = (\text{deg}_i \text{polynomials}) \text{ where } \text{dim}_C R_i \text{ is the dimension of } R_i \text{ in degree } i. \]

In supersymmetric
\[ \text{P} \]Theorem 11.1]) there exists
\[ \text{where dim}_C R_i \text{ are turned on, the (vacuum) moduli spaces are} \]
gauge theories, when the vevs of scalars in different super-
\[ \text{are parameterised by a} \]
\[ \text{polynomial} \]
\[ \text{where the variables} \]
\[ \text{the w.p.s. induces a grading on} \]
\[ \text{the weight} \]
\[ \text{m homogeneous ideal generated by the polynomials defining} \]
\[ \text{geneous coordinate ring by} \]
\[ \text{d homogeneous space (w.p.s.)} \]
\[ \text{a topological invariant in that it depends on the embed-
\[ \text{cal properties of a projective algebraic variety. It is not} \]
\[ \text{Given a complex projective variety} \]
\[ \text{For} \]
\[ \text{parametrised by a} \]
\[ \text{us that} \]
\[ \text{the numerator is a palin-
\[ \text{Here dim is the dimension of} \]
\[ \text{the} \]
\[ \text{HS is the generating function for the dimensions} \]
\[ \text{of the multi-graded series} \]
\[ \text{obtained by considering multi-graded rings with pieces} X \]
\[ \text{for } \]
\[ \text{Duality and Moduli Spaces.} \text{ HS have been well-
\[ \text{studied in the context of quiver gauge theories.} \text{ For} \]
\[ \text{branches in low dimensions, HS obtained from the Molien–}
\[ \text{Weyl integral enable us to systematically study the geometry} \]
\[ \text{of SQCDs} \text{.} \text{ Such methods can also be used to study} \]
\[ \text{the instanton moduli spaces} \text{.} \text{ As the spaces of} \]
\[ \text{dressed monopole operators, i.e. the Coulomb branches,} \]
\[ \text{receive quantum corrections, monopole formula} \text{ and} \]
\[ \text{Hall–Littlewood formula} \text{ are used to obtain the HS.} \text{ This not only} \]
\[ \text{unveils the geometry of moduli spaces,} \text{ but also provides tools and evidences to study three-
\[ \text{The HS is an important quantity that encodes numeri-
\[ \text{cally important in that it depends on the embed-
\[ \text{ding under consideration} \text{.} \text{ We work} \]
\[ \text{throughout with varieties defined over} \}
\[ \text{widely studied for} \]
\[ \text{by feeding} \]
\[ \text{in coefficients of the corresponding HS.} \text{ Keras} \text{ with} \]
\[ \text{the} \]
\[ \text{Many methods can be used to determine whether fake HS of the form} \]
\[ \text{for} \]
\[ \text{in} \text{ as well as to the minimal super-
\[ \text{Non-trivial algebraic varieties} \text{ such as hyperkähler} \]
\[ \text{cones and (closures of) symplectic} \]
\[ \text{HS have been well-
\[ \text{studied for} \]
\[ \text{in} \text{ as well as to the minimal super-
\[ \text{Non-trivial algebraic varieties} \text{ such as hyperkähler} \]
\[ \text{cones and (closures of) symplectic} \]
\[ \text{HS have been well-
\[ \text{studied for} \]
\[ \text{in} \text{ as well as to the minimal super-
\[ \text{Non-trivial algebraic varieties} \text{ such as hyperkähler} \]
\[ \text{cones and (closures of) symplectic} \]
\[ \text{HS have been well-
\[ \text{studied for} \]
\[ \text{in} \text{ as well as to the minimal super-
\[ \text{Non-trivial algebraic varieties} \text{ such as hyperkähler} \]
\[ \text{cones and (closures of) symplectic} \]
\[ \text{HS have been well-
\[ \text{studied for} \]
\[ \text{in} \text{ as well as to the minimal super-
\[ \text{Non-trivial algebraic varieties} \text{ such as hyperkähler} \]
\[ \text{cones and (closures of) symplectic} \]
\[ \text{HS have been well-
\[ \text{studied for} \]
\[ \text{in} \text{ as well as to the minimal super-
\[ \text{Non-trivial algebraic varieties} \text{ such as hyperkähler} \]
\[ \text{cones and (closures of) symplectic} \]

\[ \text{The Hilbert Series and Physics} \]

III. MACHINE LEARNING

In this section we describe our approaches to ML prop-
\[ \text{erties of the rational representations} \text{ (1) and} \]
\[ \text{by feeding} \]
\[ \text{in coefficients of the corresponding HS.} \text{ Keras} \text{ with} \]
\[ \text{the} \]
\[ \text{Many methods can be used to determine whether fake HS of the form} \]
\[ \text{for} \]
\[ \text{in} \text{ as well as to the minimal super-
\[ \text{Non-trivial algebraic varieties} \text{ such as hyperkähler} \]
\[ \text{cones and (closures of) symplectic} \]
\[ \text{HS have been well-
\[ \text{studied for} \]
\[ \text{in} \text{ as well as to the minimal super-
\[ \text{Non-trivial algebraic varieties} \text{ such as hyperkähler} \]
\[ \text{cones and (closures of) symplectic} \]
\[ \text{HS have been well-
\[ \text{studied for} \]
\[ \text{in} \text{ as well as to the minimal super-
\[ \text{Non-trivial algebraic varieties} \text{ such as hyperkähler} \]
\[ \text{cones and (closures of) symplectic} \]
\[ \text{HS have been well-
\[ \text{studied for} \]
\[ \text{in} \text{ as well as to the minimal super-
\[ \text{Non-trivial algebraic varieties} \text{ such as hyperkähler} \]
\[ \text{cones and (closures of) symplectic} \]
\[ \text{HS have been well-
\[ \text{studied for} \]
\[ \text{in} \text{ as well as to the minimal super-
\[ \text{Non-trivial algebraic varieties} \text{ such as hyperkähler} \]
\[ \text{cones and (closures of) symplectic} \]
\[ \text{HS have been well-
\[ \text{studied for} \]
\[ \text{in} \text{ as well as to the minimal super-
\[ \text{Non-trivial algebraic varieties} \text{ such as hyperkähler} \]
\[ \text{cones and (closures of) symplectic} \]
\[ \text{HS have been well-
\[ \text{studied for} \]
\[ \text{in} \text{ as well as to the minimal super-
\[ \text{Non-trivial algebraic varieties} \text{ such as hyperkähler} \]
\[ \text{cones and (closures of) symplectic} \]
\[ \text{HS have been well-
\[ \text{studied for} \]
\[ \text{in} \text{ as well as to the minimal super-
\[ \text{Non-trivial algebraic varieties} \text{ such as hyperkähler} \]
\[ \text{cones and (closures of) symplectic} \]

\[ \text{The Plethystic Programme.} \text{ In supersymmetric} \]
\[ \text{gauge theories, when the vevs of scalars in different super-
\[ \text{multiplets are turned on, the (vacuum) moduli spaces are} \]


\[ \text{https://github.com/edhirst/HilbertSeriesML.git} \]
didate HS conjecturally associated to three-dimensional Q-Fano varieties with Fano index one, as constructed in [41, 42]. Such varieties come with a natural choice of ample divisor $D = -K$, the anti-canonical divisor. We call these HS “real”. See Appendix B for the distributions of the parameters $d, \{a_i\}, \{s\}, \{p_i\}, \{q_i\}$ for this set of data. Here we are using notation as in (1), and write $P(t) = 1 + \sum_{i=1}^{d} a_i t^i$ for the numerator polynomial.

**Example 1** Consider the three-dimensional Q-Fano variety $X \subset \mathbb{P}(1^3, 2^2, 3^2)$ (number 11122 in the GRDB). This is of codimension 3, with $B = \{\frac{1}{2}(1, 1, 1), 2 \times \frac{1}{3}(1, 1, 2)\}$ isolated orbifold points, and hence has Gorenstein index $j = 6$. Writing the HS in the form (1) gives:

$$H(t; X) = \frac{P(t)}{(1-t)^3(1-t^2)^2(1-t^3)^2}$$

where $P(t) = 1 - 2t^4 - 2t^5 + 2t^7 + 2t^8 - t^{12}$.

Rewriting this in the form (2) gives:

$$H(t; X) = \frac{\hat{P}(t)}{(1-t^6)^4}$$

where $\hat{P}(t) = 1 + 3t + 8t^2 + \ldots + 8t^{21} + 3t^{22} + t^{23}$.

For the HS of this dataset, there are two competing phenomena that contribute to its coefficients: the initial part $P_{ini}$ that coincides with the HS in small degrees and the “correction terms” $P_{orb}(Q)$ for each isolated orbifold point $Q = \frac{1}{r}(b_1, \ldots, b_{\dim})$ of $X$. More precisely, we have [58]

$$H(t; X) = P_{ini} + \sum_{Q \in B} P_{orb}(Q)$$

where the sum is taken over the set $B$ of isolated orbifold points of $X$. $P_{ini}$ and $P_{orb}(Q)$ $(Q = \frac{1}{r}(b_1, \ldots, b_{\dim}))$ satisfy

$$P_{ini} = \frac{A(t)}{(1-t)^{\dim+1}}, \quad P_{orb}(Q) = \frac{B_Q(t)}{(1-t)^{\dim(1-t^r)}}$$

where $A(t), B_Q(t)$ are integral palindromic polynomials with degrees related via $\deg B_Q(t) - \deg A(t) = r - 1$. The coefficients (called plurigenera) of the HS of $H$ coincide with $P_{ini}$ in degrees $\leq \left\lfloor \frac{\deg A(t)}{2} \right\rfloor$, whilst in higher degrees the orbifold points start to contribute to the plurigenera. Because of this phenomenon, extra care must be taken when computing parameters for the representations (1) and (2) from a finite set of coefficients of the HS. Our investigations show that ML can cope with this behaviour.

### B. Generating and ML Fake HS

The “fake” HS generated take the forms (1) and (2), with numerators of the form $1 + \sum_{i=1}^{d} a_i t^i$. The numerator $\hat{P}(t)$ of (2) is required to be palindromic (and, as a consequence, $a_d = 1$). Coefficient sets consisting of the parameters $d, \{a_i\}, \{s\}, \{p_i\}, \{q_i\}$, where $1 \leq i \leq d$ and $1 \leq \ell \leq s$, were randomly generated and the Taylor expansions of the resulting fake HS were computed to order $\sim 1000$. If the parameters did not satisfy $\sum t p_i q_i > d$, if there were negative coefficients in the resulting Taylor expansion, or if they matched a real Hilbert series then the parameters were discarded.

The resulting data were fed into a NN to learn the desired properties of the fake HS. The input was a vector of Taylor expansion coefficients: either a vector of coefficients for low-order terms 0 to 100; or for high-orders terms 1000 to 1009. Although coefficients of low-order terms are easier to calculate, predictions based on those inputs are more error-prone as contributions from orbifold points take effect only for high-order terms (see §III A).

Fewer coefficients were required when learning from coefficients deeper in the Taylor expansion; geometric data are more readily extracted from larger plurigenera. We found the following analogy from toric geometry insightful. When counting the number of lattice points $c_m = |m\Delta \cap \mathbb{Z}^{\dim}|$ in the $m$-th dilation of a polytope $\Delta$ then, for $m \gg 0$, $c_m \sim \text{Vol}(m\Delta) = n_m^{\dim}\text{Vol}(\Delta)$. (This is a toric rephrasing of the HS, with $\Delta$ the polytope associated with an ample divisor $D$ and $c_m = h^0(mD)$.) The first investigation used supervised regressor NNs to learn $\{p_i\}$ for fake HS in the form (1). Supervised classifier NNs were trained to predict the Gorenstein index $j$ and the dimension $\dim$ of fake HS in the form (2). Classifiers were used since the NN outputs were single numbers and hence associated well to classifier data structures.

We conclude with a comparison of the collected fake HS data with the real HS data from the GRDB. We use the unsupervised method of principal component analysis (PCA) to project the classes onto the highest variance linear component (see Figure 1). The PCA was performed on the vectors of the first 100 coefficients, with prior scalar transformation. The explained variance ratios give the normalised eigenvalues for the covariance matrix, sorted into a decreasing order. For the fake to real HS comparison the first eigenvalue (0.78) was significantly larger than the second (0.16) and subsequent 98 eigenvalues ($< 0.04$). This indicates that one principal component is sufficient for description of the data distribution, and this principal component pays linearly progressively more attention to coefficients throughout the input HS vector up to the 24th where it then considers equal contributions from the remaining coefficients.

The projection shows a separation between the classes, indicating that there is linear structure in the data. Despite great efforts we were unable to break this separation. This raises the following question, to which we do not currently know the answer: what additional properties do fake HS need to satisfy to better approximate the GRDB HS data?

**HS Regressor Investigations.** For this investigation $\sim 10,000$ fake HS of the form (1) were uniformly drawn from a sample space given by $d = 3$, $s = 3$, $|a_i| \leq$
FIG. 1. PCA for HS Taylor expansion coefficients coming from the GRDB, ‘Real’, or those randomly generated, ‘Fake’.

| Orders of Input | MAE       |
|-----------------|-----------|
| 0 to 100        | 1.94 ± 0.11 |
| 1000 to 1009    | 1.04 ± 0.12 |

TABLE I. Averaged MAE, with standard error, of the 5-fold cross-validation of the NN learning the weights $p_\ell$ (with multiplicity) of the form (1) of the HS from input vectors of HS coefficients to the specified orders.

$10, p_\ell \leq 10$. This space was chosen to provide a sufficiently large range of fake HS whilst ensuring that its size was still feasible for ML training. The goal was to predict the values $\{p_\ell\}$ and $\{q_\ell\}$ of the form (1) from a given (finite) range of HS coefficients. This information was encoded into a single vector where each $p_\ell$ was repeated $q_\ell$ times, and the entries were given in increasing order.

A 5-fold cross-validation (in the sense of [59]) was performed for a feed-forward regressor NN with 4 hidden dense layers of 1024 neurons each, using LeakyReLU activation (with $\alpha = 0.01$), in batches of 32 for 20 epochs over the full dataset. The NN had a final dense layer with as many neurons as $p_\ell$’s (counting multiplicities). Dropout layers between the dense layers reduced the risk of overfitting (dropout factor 0.05). The NN was trained with the Adam optimiser [60] using a $\text{log}(\cosh)$ loss function and the training performance was measured via MAE. $\text{log}(\cosh)$ is a continuous version of MAE used as the loss function such that training performance would be improved for gradient descent near the MAE discontinuity, however MAE provides a more interpretable metric of learning performance so is used as the metric on the independent test data.

Table I summarises the averaged MAE, with standard error, over the 5-fold cross-validation for two ranges of HS coefficients: the first 101 coefficients; and the coefficients of order 1000 to 1009. In both cases the MAE is below 2, i.e. the true denominator of the form (1) of the underlying HS could be extracted with reasonably good accuracy from the HS coefficients alone.

**HS Classification Investigations.** In this investigation a 5-fold cross-validation for a feed-forward classifier NN with the same layer structure as before was trained. We again used an Adam optimiser, but now with *sparse categorical cross entropy* loss to reflect the classification question. Training performance was measured with accuracy and MCC. The final dense layer now had as many neurons as classes in the investigation (5 in both cases), with softmax activation, and neurons representing the values the learnt parameters could take.

This time $\sim 10\,000$ HS of the form (2) were uniformly drawn from a sample space given by $d = 5$, $|a_\ell| \leq 50$, $j \leq 5$, $\dim \leq 5$. The goal this time was to train an NN to predict the Gorenstein index $j$, the dimension $\dim$, and the form (2) from the HS coefficients in the same orders of degrees.

Note if coefficients in larger degrees were used as input, the larger values caused problems with the loss function. This issue was mitigated by log-normalising the HS coefficients, i.e. by taking the natural logarithm input values were scaled down to ranges the loss function and optimiser could handle. However some fake HS contained 0 coefficients and were therefore omitted, hence resulting in a full dataset of 8711 HS for the training with HS coefficients of larger degree. Note also that log-normalisation was only used in this case and in no other investigations.

Table II summarises the averaged accuracies and MCCs, with standard error, over the 5-fold cross-validation of the NN. These results show almost perfect classification of both the Gorenstein index, $j$, and the dimension, $\dim$, from HS coefficients in low degrees. Interestingly the performance is worse when using terms deeper in the HS, presumably due to the required log-normalisation of the coefficients removing the finer structure of the coefficients required to determine the exact parameter value being learnt.

**C. Identifying the Gorenstein Property**

In this section we investigate the effectiveness of binary classifiers to detect if the numerator of form (2) of an HS is palindromic. Recall from Section II that the numerator is palindromic if the ring $R$ is Gorenstein (by Serre duality). Then the numerator of form (1) is palindromic too (possibly up to a sign); see Example 1 for an illustration. The goal was to use a NN to distinguish whether a HS is
coming from a Gorenstein ring, i.e. the numerator polynomial of form (2) is palindromic. As before the NN’s input were HS coefficients from the same ranges of degrees.

For the investigation two equally sized sets of fake HS, one with and the other without palindromic numerators, were uniformly drawn from a sample space given by $d = 9$, $|a_i| \leq 50$, $j = 5$, dim $+1 = 6$. The same reasons as before apply for this choice of space. The HS in each of the two sets were then labelled and together comprised the full dataset for a 5-fold cross-validation to be performed using a feed-forward classifier NN with the same layer structure as in the previous investigation. Also the same Adam optimiser was used for training, but now with binary cross-entropy loss to reflect the classification question. Training performance was measured with accuracy and MCC. The final dense layer of the NN now had 2 neurons corresponding to whether the HS comes from a Gorenstein ring or not.

Table III summarises the averaged accuracies and MCCs, with standard error, over the 5-fold cross-validation of the NN. The results show good success in detecting if a HS comes from a Gorenstein ring using HS coefficients alone. The classifier performed better on coefficients in larger degrees indicating that the palindromicity property is more readily evident from plurigenera deeper in the HS (possibly because of the bigger variation).

In addition, PCA was also applied to the data in this binary classification problem, as seen in Figure 2, with similar behaviour for both low and high orders of input. This figure (for the low order inputs) highlights a lack of linear structure which the architecture could take advantage of. The PCA explained variance ratios for the 101 low order inputs show equal importance of the first two principal components (0.29, 0.27), lower importance for the next three components (0.19, 0.11, 0.10), minimal importance of the next four components ($\sim 0.01$), then negligible contribution from the remaining 92 ($\lesssim 10^{-30}$). Equivalently for the high order inputs the first two components are dominant (0.30, 0.26), with the next three less important (0.20, 0.12, 0.10), and the remaining five negligible ($\lesssim 10^{-10}$). In both cases the two dominant principal components have a mix of contributions from components with no discernible pattern across the HS vector of coefficients. The full outputs can be observed in this paper’s respective GitHub scripts.

| Orders of Input | Performance Measures | Accuracy | MCC       |
|-----------------|-----------------------|----------|-----------|
| 0 to 100        |                       | 0.844 ± 0.087 | 0.717 ± 0.155 |
| 1000 to 1009    |                       | 0.954 ± 0.043 | 0.919 ± 0.073 |

TABLE III. Averaged accuracy and MCC, with standard error, of the 5-fold cross-validation of a NN learning whether the HS has palindromic numerator in form (2) from HS coefficients to the specified orders as input.

FIG. 2. The PCA for HS Taylor expansion coefficients corresponding to HS defined over Gorenstein rings or non-Gorenstein rings.

D. Differentiating Real and Fake HS

This investigation examined the success of a binary classifier in distinguishing whether a HS, represented by a finite set of HS coefficients, corresponds to a real HS from the GRDB, or a randomly generated fake HS. The dataset consisted of HS candidates conjecturally associated to 3-dimensional Fano polytopes from the GRDB, amounting to $\sim 29\,000$ HS, along with as many fake HS with the same structure which were randomly generated.

A 5-fold cross-validation for a feed-forward classifier NN with the same layer structure as in the previous investigations was performed. For training an Adam optimiser with a binary cross-entropy loss with the same parameters as before was used. Training performance was measured with accuracy and MCC. The final dense layer had 2 neurons corresponding to whether the inputted HS coefficients were associated to a real or fake HS.

The $\sim 29\,000$ fake HS were generated randomly using form (1) parameters drawn from probability distributions reflecting the real HS data as given in Appendix B. An equal number of real HS were taken from the GRDB to produce the full dataset, and as before HS coefficients to the same order of degrees were used as NN inputs.

In this investigation the averaged accuracies and MCCs exceeded 0.99 for both ranges of degrees of HS coefficients. Further analysis of the data showed that coefficients of fake HS were orders of magnitudes different to the real case which possibly made this classification far easier. Resampling such that the coefficients were more comparable, although improving this investigations complexity, would make the fake data less representative with respect to the underlying variety’s properties. Hence we chose to use the same data throughout all investigations despite this binary classification becoming more trivial; as corroborated by the 1d PCA separation in Figure 1. This also highlights the uniqueness of real HS which come with a wealth of further impactful structure, e.g. on the parameters of the corresponding forms (1) and (2).
E. Detecting Complete Intersection

An important application of the plethystic logarithm (see Appendix A for details and references) is that it detects whether the underlying variety is a complete intersection (CI), i.e. the defining ideal (the ideal of polynomials vanishing on the variety) is generated by exactly codimension many polynomials. Such optimal intersection has been widely used in the physics literature, e.g. in string model-building [61, 62]. As can be seen from the definition, the PE$^{-1}$ involves the number-theoretic $\mu$-function, making the computation non-trivial. A natural question arises as to whether a trained classifier can identify whether $X$ is CI, i.e. when PE$^{-1}$ terminates as a Taylor series, by only “looking” at the the shape of the HS.

Suppose $X = \{f_1 = 0, \ldots, f_c = 0\}$ defines a complete intersection in $\mathbb{P}^{\dim X}_C$, where each $f_i$ is a homogeneous polynomial of degree $m_i$ so that the first map becomes a morphism of graded rings. Notice $X$ is a projective variety of codimension $c$ in $\mathbb{P}^\dim X_C$, i.e. has dimension $\dim = k - c$.

This time 10 000 fake HS of the form (3) represent HS with degrees shifted by $m_i$ so that the first map becomes a morphism of graded rings. Notice $X$ is a projective variety $X$ of codimension $c$ in $\mathbb{P}^\dim X_C$, i.e. has dimension $\dim = k - c$.

If we truncate the Taylor series at order 100 and train on 10% of the data, the accuracy is $\sim 0.80$ with MCC $\sim 0.61$. However, including higher and higher orders of coefficients results into more and more improved results (where the increase in improvement stagnates for sufficiently high orders). For example, if we use Taylor expansions to order 300 and train on 10% of the data, the PCA+random forest model could give over 0.95 accuracy and over 0.9 MCC. More precisely, a 10-fold cross validation (with training performed on the 10% chunks) would give 0.965$\pm$0.002 accuracy (with 95% confidence interval). We can reproduce these results by using PCA and a feed-forward NN with 4 hidden dense layers of 32 neurons each, dropout layers between the dense layers (dropout factor 0.05), LeakyReLU activation (with $\alpha = 0.01$), binary cross-entropy loss function and Adam optimiser. A 10-fold cross validation with the same input (training performed on the 10% chunks) yields 0.951$\pm$0.004 accuracy (with 95% confidence interval).

PCA shows a clear separation of CIs and non-CIs (see Figure 3). The explained variance ratios show one dominant component with eigenvalue 0.98, where this component has roughly equal contributions from all the series coefficients. This raises the question if this implies that PCA can efficiently separate CI from non-CI (real) HS or if this is an artefact of our data generation. With 20 000 samples of CIs and real non-CIs, we find that a random forest could give $\sim 0.8$ accuracy and $\sim 0.6$ MCC for a 10-fold cross validation. Although this is a decent result, it would be natural to investigate in future whether there could be better techniques/algorithms to improve such performance. Further study is also necessary to confirm that PCA is an effective discriminator between CI and non-CI in this case.

Acknowledgements. JB is supported by a CSC scholarship. YHH would like to thank STFC for grant ST/J00037X/1. EH would like to thank STFC for a PhD studentship. JH is supported by a Nottingham
Appendix A: The Plethystic Programme

For a function \( f(t) = \sum_{n=0}^{\infty} a_n t^n \), we can define the plethystic exponential (sometimes known as the Euler transform) as

\[
    \text{PE}[f(t)] := \exp \left( \frac{\sum_{n=1}^{\infty} f(t^n) - f(0)}{n} \right) = \prod_{n=1}^{\infty} (1 - t^n)^{-a_n}.
\]

For instance, the mesonic BPS operators fall into two categories: single- and multi-trace. Then the HS is the generating function for counting the basic single-trace invariants. Moreover, the HS of the \( N \)-th symmetric product is given by \( g_N(t; M) = f(t; \text{sym}^N(X)) \), \( \text{sym}^N(X) := M^N/S_N \), where the “grand-canonical” partition function is given by the fugacity-inserted plethystic exponential of the Hilbert series: \( \text{PE}_f[f(t)] := \prod_{n=0}^{\infty} (1 - \nu t^n)^{-a_n} = \sum_{N=0}^{\infty} g_N(t) \nu^N \). In gauge theory, this is considered to be at finite \( N \) and the expansion \( g_N(t) = \sum_{n=0}^{\infty} b_n t^n \) gives the number \( b_n \) of operators of charge \( n \).

There is also an analytic inverse function to PE, which is the plethystic logarithm, given by

\[
    \text{PE}^{-1}[g(t)] = \sum_{k=1}^{\infty} \frac{\mu(k)}{k} \log(g(t^k)),
\]

where \( \mu(k) \) is the Möbius function. The first positive terms in the Taylor expansion of \( \text{PE}^{-1} \) encodes generators at different degrees, and the first negative terms give the relations among them. Higher order terms are known as the syzygies. In particular, if \( X \) is a complete intersection, then \( \text{PE}^{-1}[H(t)] \) is a polynomial of \( t \) (i.e. terminates at a finite order).

Appendix B: Real HS parameter distributions

The dataset of real HS associated to 3-dimensional Fano varieties considered in this paper [43] that was analysed to produce distributions of the HS function form parameters \( d, \{a_i\}, s, \{p_i\}, \{q_i\} \) as shown in Figures 4-8. These distributions, and their respective fittings were used to make fake HS generation more representative of the real HS data.

Fittings used sums of Gaussian distributions, reflecting a Central Limit Theorem motivation in analysis of this large dataset of \( \sim 54 000 \) HS. In all cases the sum of 2 independent Gaussian distributions sufficed in making a visually accurate fit. Thus, using these distribution in fake HS generation would ideally produce HS of the same form. Interestingly, the fake HS still had quite different coefficient growth rates to the real HS, stabilising deeper in the series. This phenomena is further discussed in §III D.
FIG. 6. Histogram of distribution of real HS number of denominator factors $s$, with Gaussian fitting.

FIG. 7. Histogram of distribution of real HS denominator internal powers (i.e. denominator weights) $p_\ell$, with Gaussian fitting.

FIG. 8. Histogram of distribution of real HS denominator external powers (i.e. number of repetitions of each denominator weight) $q_\ell$, with Gaussian fitting.

[1] S. Benvenuti, B. Feng, A. Hanany, and Y.-H. He, “Counting BPS operators in gauge theories: quivers, syzygies and plethystics,” J. High Energy Phys. no. 11, (2007) 050, 48.

[2] B. Feng, A. Hanany, and Y.-H. He, “Counting gauge invariants: the plethystic program,” J. High Energy Phys. no. 3, (2007) 090, 42.

[3] J. Gray, Y.-H. He, A. Hanany, N. Mekareeya, and V. Jejjala, “SQCD: a geometric aperçu,” Journal of High Energy Physics 2008 no. 05, (May, 2008) 099–099.

[4] A. Hanany, N. Mekareeya, and G. Torri, “The Hilbert series of adjoint SQCD,” Nuclear Phys. B 825 no. 1-2, (2010) 52–97.

[5] Y. Chen and N. Mekareeya, “The Hilbert series of U/SU SQCD and Toeplitz determinants,” Nuclear Phys. B 850 no. 3, (2011) 553–593.

[6] N. Jokela, M. Järvinen, and E. Keski-Vakkuri, “New results for the SQCD Hilbert series,” J. High Energy Phys. no. 3, (2012) 048, front matter+30.

[7] S. Benvenuti, A. Hanany, and N. Mekareeya, “The Hilbert series of the one instanton moduli space,” J. High Energy Phys. no. 6, (2010) 100, 40.

[8] A. Hanany, N. Mekareeya, and S. S. Razamat, “Hilbert series for moduli spaces of two instantons,” J. High Energy Phys. no. 1, (2013) 070, front matter + 48.

[9] E. I. Buchbinder, A. Lukas, B. A. Ovrut, and F. Ruehle, “Instantons and Hilbert Functions,” Phys. Rev. D 102 no. 2, (2020) 026019, arXiv:1912.08358 [hep-th].

[10] A. Hanany, E. E. Jenkins, A. V. Manohar, and G. Torri, “Hilbert series for flavor invariants of the Standard Model,” J. High Energy Phys. no. 3, (2011) 096, 7.

[11] L. Lehman and A. Martin, “Low-derivative operators of the Standard Model effective field theory via Hilbert series methods,” Journal of High Energy Physics 02 no. 81. (2016).

[12] V. Braun, “Counting points and Hilbert series in string physics,” J. High Energy Phys. 02 no. 3, (2012) 048, front matter+30.
theory,” in Strings, gauge fields, and the geometry behind, pp. 225–235. World Sci. Publ., Hackensack, NJ, 2013.
[13] L. Lehman and A. Martin, “Hilbert series and plethysm: paving the path towards 2HDM- and MLRSM-EFT,” J. High Energy Phys. no. 9, (2016) 035, 113.
[14] A. Kobach and S. Pal, “Hilbert series and operator bases with derivatives in effective field theories,” Comm. Math. Phys. 347 no. 2, (2016) 363–388.
[15] A. Davies, P. Veličković, L. Buesing, S. Blackwell, Y.-H. He, “The Calabi–Yau landscape: from geometry, commutative algebra and machine learning,” Theor. Math. Phys. 200 no. 1, (2015) 105014.
[16] B. Henning, X. Lu, T. Melia, and H. Murayama, “Hilbert series and operator bases with derivatives in effective field theories,” Comm. Math. Phys. 347 no. 2, (2016) 363–388.
[17] L. Graf, B. Henning, X. Lu, T. Melia, and H. Murayama, “Hilbert series and operator bases with derivatives in effective field theories,” Phys. Lett. B 772 (2017) 225–231.
[18] Anisha, S. Das Bakshi, J. Chakrabortty, and S. Prakash, “Hilbert series and plethystics: paving the path towards 2HDM- and MLRSM-EFT,” J. High Energy Phys. no. 9, (2019) 035, 113.
[19] C. B. Marinissen, R. Rahn, and W. J. Waalewijn, “... 83106786, 114382724, 1509048322, 2343463290, 27410087742, ... efficient Hilbert series for effective theories,” Phys. Lett. B 808 (2020) 135632, 7.
[20] L. Graf, B. Henning, X. Lu, T. Melia, and H. Murayama, “... 12, 117, 1959, 45171, 1170686, ... a Hilbert series for the QCD chiral Lagrangian,” JHEP 01 (2021) 142, arXiv:2009.01239 [hep-ph].
[21] Y.-H. He, “Deep-learning the landscape.”
[22] Y.-H. He, “The Calabi–Yau landscape: from geometry, to physics, to machine-learning,” arXiv:1812.02893 [hep-th], 2018.
[23] A. Davies, P. Veličković, L. Buesing, S. Blackwell, D. Zheng, N. Tomašev, R. Tanburn, P. Battaglia, C. Blundell, A. Juhász, M. Lackenby, G. Williamson, D. Hassabis, and P. Kohli, “Advancing mathematics by guiding human intuition with AI,” Nature 600 (2021) 70–74.
[24] Y.-H. He, “Machine-learning the string landscape,” Phys. Lett. B 774 (2017) 564–568.
[25] D. Krefl and R.-K. Seong, “Machine learning Calabi–Yau volumes,” Phys. Rev. D 96 no. 6, (2017) 066014, 8.
[26] F. Ruehle, “Evolving neural networks with genetic algorithms to study the string landscape,” J. High Energy Phys. no. 8, (2017) 038, front matter+19.
[27] J. Carifio, J. Halverson, D. Krioukov, and B. D. Nelson, “Machine learning in the string landscape,” J. High Energy Phys. no. 9, (2017) 057, front matter+35.
[28] C. R. Brodie, A. Constantin, R. Deen, and A. Lukas, “Machine learning line bundle cohomology,” Fortschr. Phys. 68 no. 5, (2020) 2000005, 13.
[29] Y.-H. He and S.-T. Yau, “Graph Laplacians, Riemannian manifolds and their machine-learning,” arXiv:2006.16619 [math.CO], 2020.
[30] S. Altmok, G. Brown, and M. Reid, “Fano 3-folds, K3 surfaces and graded rings,” in Topology and geometry: commemorating SISTAG, vol. 314 of Contemp. Math., pp. 25–53. Amer. Math. Soc., Providence, RI, 2002.
[31] Y.-H. He and A. M. Kasprzyk, “Kawamata boundedness for Fano threefolds and the Graded Ring Database.”
[32] Y.-H. He and A. M. Kasprzyk, “The Graded Ring Database.” Online. http://www.grdb.co.uk/.
[33] G. Brown and A. M. Kasprzyk, “The Fano 3-fold database,” Zenodo https://doi.org/10.5281/zenodo.5820338, 2022.
[34] J. Harris, Algebraic geometry, vol. 133 of Graduate Texts in Mathematics. Springer-Verlag, New York, 1995. A first course, Corrected reprint of the 1992 original.
[35] I. Dolgachev, “Weighted projective varieties,” in Group actions and vector fields (Vancouver, B.C., 1981), vol. 956 of Lecture Notes in Math., pp. 34–71. Springer, Berlin, 1982.
[36] M. F. Atiyah and I. G. Macdonald, Introduction to commutative algebra. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1969.
[37] R. P. Stanley, “Hilbert functions of graded algebras,” Advances in Mathematics 28 (1978) 57–83.
[38] F. Buccella, J. P. Derendinger, S. Ferrara, and C. A. Savoy, “Patterns of symmetry breaking in supersymmetric gauge theories,” Phys. Lett. B 115 no. 5, (1982) 375–379.
[39] M. A. Luty and W. Taylor, IV, “Varieties of vacua in supersymmetric gauge theories,” J. High Energy Phys. no. 7, (1996) 3399–3405.
[40] Y.-H. He, Y.-H. He, and J. D. Hauenstein, “Numerical algebraic geometry: a new perspective on gauge and string theories,” J. High Energy Phys. no. 7, (2012) 018,
[52] A. Hanany, N. Mekareeya, and S. S. Razamat, “Hilbert series for moduli spaces of two instantons,” *J. High Energy Phys.* no. 1, (2013) 070, front matter + 48.

[53] S. Cremonesi, A. Hanany, and A. Zaffaroni, “Monopole operators and Hilbert series of Coulomb branches of 3d $\mathcal{N}=4$ gauge theories,” *Journal of High Energy Physics* 5 (2014).

[54] S. Cremonesi, A. Hanany, N. Mekareeya, and A. Zaffaroni, “Coulomb branch Hilbert series and Hall–Littlewood polynomials,” *Journal of High Energy Physics* 178 (2014).

[55] Y.-H. He, V. Jejjala, C. Matti, and B. D. Nelson, “Veronese geometry and the electroweak vacuum moduli space,” *Phys. Lett. B* 736 (2014) 20–25.

[56] Y. Xiao, Y.-H. He, and C. Matti, “Standard model plethystics,” *Phys. Rev. D* 100 (Oct, 2019) 076001.

[57] e. a. Abadi, M., “TensorFlow: Large-scale machine learning on heterogeneous systems.” Online, 2015.

[58] A. Buckley, M. Reid, and S. Zhou, “Ice cream and orbifold Riemann-Roch,” *Izv. Ross. Akad. Nauk Ser. Mat.* 77 no. 3, (2013) 29–54.

[59] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning.* Springer Series in Statistics. Springer, New York, second ed., 2009. Data mining, inference, and prediction.

[60] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization.” *arXiv:1412.6980 [cs.LG]*, 2014.

[61] P. Candelas, A. M. Dale, C. A. Lütken, and R. Schimmrigk, “Complete intersection Calabi-Yau manifolds,” *Nuclear Phys. B* 298 no. 3, (1988) 493–525.

[62] L. B. Anderson, Y.-H. He, and A. Lukas, “Heterotic compactification, an algorithmic approach,” *J. High Energy Phys.* no. 7, (2007) 049, 34.