ADV. THEOR. MATH. PHYS.
Volume 27, Number 5, 1405–1461, 2023

Mahler measuring the genetic code of amoebae

Siqi Chen, Yang-Hui He, Edward Hirst,
Andrew Nestor, and Ali Zahabi

Amoebae from tropical geometry and the Mahler measure from number theory play important roles in quiver gauge theories and dimer models. Their dependencies on the coefficients of the Newton polynomial closely resemble each other, and they are connected via the Ronkin function. Genetic symbolic regression methods are employed to extract the numerical relationships between the 2d and 3d amoebae components and the Mahler measure. We find that the volume of the bounded complement of a d-dimensional amoeba is related to the gas phase contribution to the Mahler measure by a degree-d polynomial, with d = 2 and 3. These methods are then further extended to numerical analyses of the non-reflexive Mahler measure. Furthermore, machine learning methods are used to directly learn the topology of 3d amoebae, with strong performance. Additionally, analytic expressions for boundaries of certain amoebae are given.

1. Introduction

The amoebae of affine algebraic varieties are interesting objects at the intersection of tropical geometry [18, 45, 53, 61] in mathematics and dimer models in physics [23, 24, 29, 41]. Amoebae are constructed out of logarithmic projections of complex varieties described by toric diagrams. These toric diagrams are lattice polytopes whose dimensions can be associated to complex coordinates and vertices to monomial terms in the varieties defining equation, via the Newton polynomial [11, 19, 25]. How the topology and geometry of the surface changes under the amoeba projection makes them particularly interesting objects of study.

The Mahler measure, which was first introduced by Kurt Mahler in 1962 [47], can be interpreted as the limit height function and the free energy in these dimer models [12, 60]. Further to this in crystal melting models [13, 49, 50, 59], the Mahler measure and the amoeba are closely related by the Ronkin function which is the limit shape of the molten crystal; with relation to quiver Yangians [3, 4]. In particular, the Mahler measure is the Ronkin function at the origin and the gradient of the Ronkin function is constant over each components of the amoeba complement. In [14], a number of observations were made which explicated how the dynamical aspects of the gauge theory are encoded in the Mahler measure by defining a concept called the Mahler flow. So far, the appearance of the Mahler measure in physics has only been studied in the context of 2-dimensional reflexive polygons, which have only a single free parameter in the Mahler flow. Thus, there is lots of room to dive deeper into the properties and relations related to the Mahler measure in broader context, such as in the case of non-reflexive polytopes.

In recent years, there has been an increasing effort in applying data science techniques to mathematical sciences based on the observation that mathematical data often take the form of labelled or unlabelled tensors that naturally resemble the data structure required in machine learning (ML). In mathematical physics, this was initiated in the exploration of string landscape [20, 30, 42, 54] and extended to a broader range of topics in [2, 6, 9, 10, 16, 17, 21, 22, 27, 28, 33, 34, 43, 44]. Interested readers can refer to [7, 31, 32] for comprehensive reviews on this application.

Specifically, [8] integrated these two directions by applying ML techniques to study the amoebae and tropical geometry, taking advantage of the classification and image-processing power of ML. In this paper, we extend the discussion in [8] to 3-dimensional amoebae, consider the Mahler flow proposed in [14] in greater details in the context of non-reflexive polytopes, and then apply standard ML techniques to make precise the qualitative relation between the Mahler measure and the bounded amoeba complements observed and conjectured in [14], as well as considering the implications in theories built from non-reflexive polytopes as discussed in [5]. Since computing exact properties of the amoeba has been a challenge with analytic results mostly concerning its approximations and special limits, for example in [40, 53], our numerical results obtained from ML could provide insights in its understanding in more general scenarios.

The paper is organised as follows. Section 2 reviews some preliminaries about amoebae and the Mahler measure and their relations in dimer models which motivate the effort of this paper to explore the links between these concepts in mathematics and physics. The following Sections 3 and 4 discuss some interesting physical properties related to amoebae and the Mahler measure respectively. More specifically, in Section 3, we apply feed-forward neural networks and convolutional neural networks to ML the second Betti number of the 3-dimensional amoebae associated with reflexive polytopes, based on the discussion of 2-dimensional amoebae in [8]. Section 4 extends the discussion of the Mahler flow of reflexive polytopes in [14] to the case of non-reflexive polytopes, where there are more than one amoeba holes, leading to interesting dynamics. In Section 5, we consider the more physically relevant quantities, namely the relations between the Mahler measure and the area (volume) of the bounded complementary region of the amoebae, implementing a genetic algorithm for symbolic regression. In doing so, we also obtain analytic expressions for the boundary of some amoebae. Finally, Section 6 discusses the main results in this paper and possible future directions.

2. Preliminaries

2.1. Amoeba

The amoeba, $\mathcal{A}_V \subset \mathbb{R}^r$, of an algebraic hypersurface, $V_{\mathbb{C}}(f) \subset \mathbb{C}^r$, is defined as the image of the logarithmic map,

$$\mathcal{A}_V \equiv \mathrm{Log}(V_{\mathbb{C}}(f)) \,, \tag{1}$$

where

$$\mathrm{Log}(z_1, \ldots, z_n) \equiv (\log|z_1|, \ldots, \log|z_n|) \,. \tag{2}$$

Since the hypersurface $V_{\mathbb{C}}(f) = \{z \in \mathbb{C}^n : f(z) = 0\}$ is the zero locus of the function f, the corresponding amoeba may also be denoted as $\mathcal{A}_f$.

The function of interest is the Newton polynomial defined with respect to a Newton polytope which is a convex lattice polytope, also known as a toric diagram. In the case of n complex dimensions, the Newton polynomial is of the form $P(\mathbf{z}) = \sum_{\mathbf{p}} c_{\mathbf{p}} \mathbf{z}^{\mathbf{p}}$, summing over the polytope vertices $\mathbf{p}$ each with coordinates p_i in the i-th lattice dimension. In particular, in three complex dimensional ($r = 3$) cases, Newton polynomial is of the form $P(u, v, w) = \sum_{\mathbf{p}} c_{(p_1,p_2,p_3)} u^{p_1} v^{p_2} w^{p_3}$; and now denoting $(z_1, z_2, z_3) \mapsto (u, v, w)$ to emphasise the restriction to $r = 3$.

An amoeba can be approximated using **lopsidedness** which is defined as follows. A list of positive numbers $\{c_1, \ldots, c_n\}$ is **lopsided** if one of the numbers is greater than the sum of the rest of numbers. If $\{c_1, \ldots, c_n\}$ is not lopsided, there exists a list of phases $\{\phi_i\}$ such that $\sum_i \phi_i c_n = 0$, via the triangle inequality [18]. Thus, the lopsided amoeba, $\mathcal{LA}_f$, is defined by

$$\mathcal{LA}_f \equiv \{\mathbf{a} \in \mathbb{R}^r | \, f\{\mathbf{a}\} \,\text{is not lopsided}\} \,, \tag{3}$$

so that $\mathcal{LA}_f \supseteq \mathcal{A}_f$.

Let n be a positive integer, $\mathbf{x} \in \mathbb{R}^r$, and $f(\mathbf{x})$ a (Newton) polynomial, define $\tilde{f}_n$ to be

$$\tilde{f}_n(\mathbf{x}) := \prod_{k_1=0}^{n-1} \cdots \prod_{k_r=0}^{n-1} f(e^{2\pi i k_1/n} x_1, \ldots, e^{2\pi i k_r/n} x_r) \,, \tag{4}$$

which is a cyclic resultant defined as

$$\tilde{f}_n = \mathrm{res}_{u_r}(\mathrm{res}_{u_{r-1}}(\ldots \mathrm{res}_{u_1}(f(u_1 x_1, \ldots, u_r x_r), u_1^n - 1)$$

$$(5) \qquad \ldots, u_{r-1}^n - 1), u_r^n - 1)$$

where $\mathrm{res}_u(f, g)$ is the resultant of f, g with respect to the variable u.

Theorem 2.1. *The lopsided amoeba $\mathcal{LA}_{\tilde{f}_n}$ converges uniformly to $\mathcal{A}_f$ as $n \to \infty$, where $\tilde{f}_n$ is the cyclic resultant of f .*

The Newton polytope of $\tilde{f}_n$ is $n^r \Delta(f)$, as a dilation of the original polytope [53].

An example of a 3-dimensional amoeba is given in Figure 1, generated through Monte Carlo sampling of points on the Riemann surface.

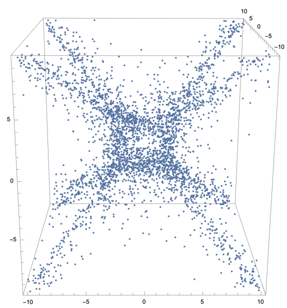


Figure 1. Amoeba of the hypersurface $P(z_1, z_2, z_3) = z_1 + z_1^{-1} + z_2 + z_2^{-1} + z_3 + z_3^{-1} + 10 = 0$.

Figure 2. The amoeba and its cross-section in **Figure 1** after transformation while preserving its topology.

Additionally, to improve the visualisation of the amoeba image, a $GL(3,\mathbb{Z})$ transformation can be performed, such that the Monte Carlo generated points occur as a more even sample across the full amoeba [8]. These transformations although changing the geometry preserve the amoebae topology, in particular the number of cavities (3-dimensional holes). The transformation matrix used here is $M = \begin{pmatrix} 5 & 1 & 2 \\ 1 & 2 & 5 \\ 1 & 5 & 1 \end{pmatrix}$, producing the amoeba shown in Figure 2. The complementary components of an amoeba may be bounded or unbounded. In Figure 2, there is a single bounded 3-dimensional hole (cavity), which is emphasised through plotting a cross-section of this amoeba. The number of such cavities depends on the choice of coefficients of the Newton polynomial, and is bounded above by the number of internal points in the respective toric diagram. The variation of the Newton polynomial coefficients considered changes the Riemann surface geometry but preserves the topology and existence of holes, however as the coefficients change the amoeba projection of this surface changes and coefficient values where the topology of the respective amoeba changes is the focus of interest in this study. It is worth noting here also there will be coefficient choices that make the Riemann surface singular, and change its topology, but we leave consideration of the respective amoeba transitions at these Riemann surface topological transitions to future work.

A lattice polytope Δ_n is reflexive if its dual polytope is also a lattice polytope in $\mathbb{Z}^n$. A necessary but not sufficient condition for reflexivity is for the polytope to have a single interior point, and this unique interior point is taken to be the origin.

Each lattice polytope can be associated with a compact toric variety with complex dimension equal to the polytope lattice dimension. For a reflexive polytopes, the corresponding compact toric variety is a Gorenstein toric Fano variety. Separately a non-compact toric Calabi-Yau $(n+1)$-fold can also be created from the polytope by embedding it in $\mathbb{Z}^{n+1}$, setting $p_{n+1} = 1 \; \forall \mathbf{p} \in \Delta_n$, and using the respective fan; effectively constructing the non-compact CY_4 as the affine cone over the comapct Fano variety.

From the physics perspective, the toric CY_4 singularities (from the non-compact construction with 3d lattice polyhedra) can be probed by $D1-$branes to give rise to the classical mesonic moduli space of the 2d $\mathcal{N} = (0, 2)$ gauge theory. These theories are encoded by the periodic quiver diagrams which specify their matter content involving two types of matter fields and gauge symmetry [26]. The graph dual to the periodic quivers on T^3 represents brane configurations of NS5-brane and D4-branes. The complex surface defined by the zero locus of the Newton polynomial of the toric CY_4 is the surface wrapped by the NS5-brane, which can be studied using the (co)amoeba/algae projection [23].

2.2. Mahler measure

The Mahler measure was first introduced in algebraic number theory in [47], and it is defined as such[1]. Given a non-zero Laurent polynomial in n complex variables, $P(z_1, \ldots, z_n) \in \mathbb{C}[z_1^{\pm 1}, \ldots, z_n^{\pm 1}]$, the Mahler measure $m(P)$ is given by

$$m(P) = \frac{1}{(2\pi i)^n} \int_{|z_1|=1} \cdots \int_{|z_n|=1} \log |P(z_1, \ldots, z_n)| \frac{\mathrm{d}z_1}{z_1} \cdots \frac{\mathrm{d}z_n}{z_n} . \tag{6}$$

In this paper, we focus on two- and three-variable Laurent polynomials. For simplicity, consider the two-variable Laurent polynomials of the form

$$P(z, w) = k - p(z, w) \ , \tag{7}$$

where $p(z, w)$ does not have a constant term. Then, for $|k| > \max_{|z|=|w|=1} |p(z, w)|$, Mahler measure (6) becomes

$$m(P) = \mathrm{Re} \left(\frac{1}{(2\pi i)^2} \int_{|z|=|w|=1} \log(k - p(z, w)) \frac{\mathrm{d}z}{z} \frac{\mathrm{d}w}{w} \right) . \tag{8}$$

The series expansion of $\log(k - p(z, w))$ converges uniformly on the support of the integration path and leads to

$$m(P) = \log k + \int_k^\infty (u_0(t) - 1) \frac{\mathrm{d}t}{t} \ , \tag{9}$$

where

$$u_0(k) = \frac{1}{(2\pi i)^2} \int_{|z|=|w|=1} \frac{1}{1 - k^{-1} p(z, w)} \frac{\mathrm{d}z}{z} \frac{\mathrm{d}w}{w} . \tag{10}$$

The Mahler measure (9) in the context of quiver gauge theories was discussed in [14], where the *Mahler flow equation* was introduced:

$$\frac{\mathrm{d}m(P)}{\mathrm{d}\log k} = k \frac{\mathrm{d}m(P)}{\mathrm{d}k} = u_0(k) \, . \tag{11}$$

Interestingly, this equation takes the similar form as the RG flow where the energy sale is replaced by $u_0(k)$.

[1]The Mahler measure is often referred to the exponential quantity, $\exp(m(P))$, in the literature.

2.3. Relation between amoebae and the Mahler measure

A close connection between amoebae and the Mahler measure is expected through their relations to the Ronkin function in the context of dimer models. In this section, we introduce the Ronkin function, dimer models, quivers and crystal melting models in order to explicate the significance of searching for their relations in both mathematics and physics and motivate our use of ML in the process.

To begin with, the n-dimensional Ronkin function $R(x_1, x_2, ..., x_n)$ for a Newton polynomial P is defined as

$$R(x_1, x_2, ..., x_n) \coloneqq \frac{1}{(2\pi i)^n} \int_{|z_1|=1} \cdots \int_{|z_n|=1} \log |P(e^{x_1} z_1, \ldots, e^{x_n} z_n)| \times \frac{\mathrm{d}z_1}{z_1} \cdots \frac{\mathrm{d}z_n}{z_n}. \tag{12}$$

The Ronkin function links to amoebae and the Mahler measure separately as follows: Different regions of the amoeba can be probed by the Ronkin function by considering the gradient of the Ronkin function. Specifically, the Ronkin function is strictly convex over the interior of amoeba and linear over each component of its complements [14]. Its gradient is given by the corresponding lattice point. This is important for the derivation of the expressions for the boundary of the amoeba, which is elaborated in Section 5.1. On the other hand, following their definitions, the Mahler measure is the Ronkin function defined at the origin $(0, 0)$. The convexity of the Ronkin function also implies that the Mahler measure is at the minimum of the Ronkin function.

Their relations are more of interest when considered in the context of quiver gauge theories and crystal melting models arising from dimer models. Given a bipartite graph $\mathcal{G}$ where each edge connects a black and a white node, the dimer model is the study of all perfect matchings of $\mathcal{G}$ where each edge is only incident on one node. For a reference matching with a unit flow, a height function is defined as the total flux with respect to the reference matching across a path from one face to another. The partition function of the set of all perfect matchings is given by the absolute value of the Kasteleyn matrix K which, after embedding on a torus, is given by the Newton polynomial P which then defines for us the corresponding amoeba and Mahler measure.

The crystal melting models relate the counting of BPS bound states to melting crystals where different gauge groups in the quiver correspond to different types of atoms and matter contents correspond to chemical bonds in the crystals [50, 51]. Crystals are molten by removing atoms from them, and the thermodynamic limit is when a large number of atoms are removed. In this limit, the Ronkin function is the limit height function of the dimer model that can be interpreted as the limit shape of the molten crystal [41]. More importantly, the crystal melting models admit a statistical interpretation with respect to the height fluctuations where their phase structures are described by the corresponding amoebae [38, 39, 41]. This is illustrated in Figure 3. Specifically, the solid phase is where the height fluctuations are bounded almost certainly; liquid phase is where the covariance in the height function is unbounded as the distance between two distant points goes to infinity; and the gas phase is where the covariance of the average height difference is bounded but itself is unbounded (detailed discussion can be found in [38]). In the context of crystal melting models, the solid phase corresponds to the unmolten parts of the crystal which are the unbounded amoeba compliments, whereas the gas phase corresponds to the opening of the amoeba hole (oval).

Moreover, a particularly interesting boundary of the amoebae is the boundary of the bounded complement of the amoebae. It was observed in [14] that after recasting the Newton polynomial P into $P(\mathbf{z}) = k - p(\mathbf{z})$ where p contains no constant terms, the value of the parameter k controls geometrically the opening of the amoeba holes. The isoradial limit $k = k_c$ is the critical point where the amoeba hole would emerge. At $k = k_c$, the bounded complement of the amoebae is degenerated to a point, which can always be transformed to the origin. Beyond the critical point $k > k_c$, the area of the amoeba hole evolves as the value of k increases. This is consistent with the statistical interpretation of the crystal melting models as the gas phase grows when more atoms are removed. Correspondingly, the Mahler measure also changes continuously as k is varied as described by the Mahler flow equation (11) introduced in [14]. Importantly, the Mahler measure also grows monotonically along the Mahler flow as k increases above the isoradial limit. Thus, it is natural to expect that the Mahler flow is related to the evolution of the amoeba hole, and the Mahler measure is related to the bounded complement of the amoebae and perhaps its area.

The critical value of k at which the amoeba hole appears characterises the phase transition from the liquid phase to the gas phase. This motivates an associated definition of different phase contributions to the Mahler measure, proposed in [14]. In particular, the liquid and gas phase contributions to the Mahler measure, $m_{l,g}$, are defined as

$$m_l(P) = \begin{cases} m(P) & \text{for } k \leq k_c \\ m(P(k_c)) & \text{for } k > k_c \end{cases},$$

$$m_g(P) = m(P) - m(P(k_c)) \quad \text{for} \quad k \geq k_c\,. \tag{13}$$

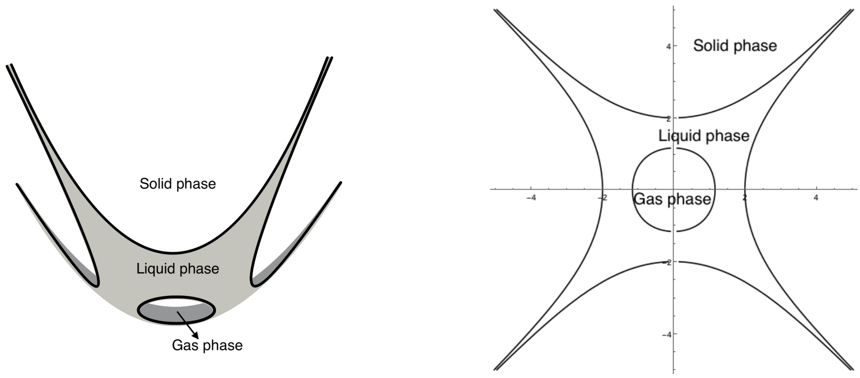


Figure 3. The Ronkin function (left) and the amoeba of $\mathbb{F}_0$ (right). Figures are adapted from [14].

From the physics perspective, the bounded complement of the amoebae is the gas phase of the dimer model and its entropy is related to the Mahler measure of the gas phase m_g, [14]. The relations between the area of the complement and m_g are numerically studied in concrete examples in Section 5. However, the precise analytic relation between the area of the gas phase and the Mahler measure is not yet understood and we hope to further study this problem in future.

Besides the interpretation of k as the parameter that controls the size of the Mahler measure and the area of the amoeba hole, it is also given physical interpretations in quiver gauge theories. As k is one of the coefficients of the Newton polynomial which are given by perfect matchings in the dimer models, they are understood to be related to Kähler moduli of the toric Gorenstein singularity [36], as discussed in more details in [14].

Given the difficulties in finding analytic expressions of areas of the amoeba and its compliments, discussions in this area have mainly focused on special limits or approximations of amoebae. In particular, the geometric interpretation of the Mahler flow in relation to amoebae was studied in [14] in the limit of $k \to \infty$ which is a tropical limit of the amoeba, and the topological features of two-dimensional amoebae were studied in [8] using the lopsided amoeba as a crude approximation. Therefore, in this paper, we adopt several ML techniques in the hope to provide some insights in obtaining explicit relations in general.

3. Machine learning 3-dimensional amoebae Betti numbers from coefficients

In this section, a variety of example complex 3d Riemann surfaces are considered, for each surface a set of polynomial coefficient vectors are generated for the respective Newton polynomial; each coefficient set giving a geometrically different surface. Each of these surfaces will have a different amoeba projection with potentially different topology under the projection.

The aim of this investigation is to establish how well ML architectures can learn to predict the second Betti number, b_2, dictating the number of 3-dimensional cavities, from the polynomial coefficients alone. For each of the example surfaces, across the set of generated amoebae the b_2 values are calculated using the topological data analysis technique of persistent homology on Monte Carlo sampled point clouds of the amoeba. These values are used as the outputs to be learnt from the coefficient vector inputs.

3.1. Estimating Betti numbers with persistent homology

The k-th homology group $\mathbf{H}_k(X)$ of a topological space X is a key concept in algebraic topology. It is defined as the quotient group of the cycle group $\mathbf{Z}_k$ by the boundary group $\mathbf{B}_k$,

$$\mathbf{H}_k \equiv \mathbf{Z}_k/\mathbf{B}_k \,, \tag{14}$$

where

$$\mathbf{Z}_k \equiv \mathrm{Ker}(\partial_k), \;\; \mathbf{B}_k \equiv \mathrm{Im}(\partial_{k+1}) \,, \tag{15}$$

under the boundary operator ∂_k, which in the simplicial complex context maps k-simplices to their boundaries made up of $(k-1)$-simplices. Thus, the dimension of the k-th homology group $\mathbf{H}_k(X)$, i.e., the k-th Betti number b_k, counts the number of k-dimensional holes in X (the number of cycles that are not boundaries of some simplicial complexes). The largest homology group one can consider is bounded by the dimension of X, such that in the case of 3-dimensional amoebae, the first homology group of interest is $\mathbf{H}_2(X)$ with dimension b_2, as the boundary of a 3-dimensional cavity is of dimension two.

Since the amoeba can be easily sampled to obtain the point cloud data for the space, its topological invariants can be obtained directly using a filtration starting from these points, via topological data analysis. This filtration of complexes is created by first considering the sampled points in $\mathbb{R}^3$, and a respective simplicial complex of as many points. Then imagining a 3d ball of radius δ centred on each point, the value of δ is continuously increased from 0 to ∞ and at each δ value where there is a new intersection of the balls the respective simplicial complex is updated to produce the next complex in the filtration. When $(k+1)$ balls intersect a k-simplex is drawn between their respective points in the simplicial complex (up to $k=3$ for these 3d data clouds).

The (p,q)-persistent k-homology $\mathbf{H}_k^{p,q}$ hence describes the birth (p) and death (q) of k-cycles created and subsequently filled as the complex changes through the filtration. There are many available algorithms and software tools for computing persistent homology, and we adopted the `python` package `ripser` due to its relative efficiency [56].

3.2. ML architecture

As in [8], we compared feed-forward neural networks and convolutional neural networks, coded in `Mathematica` [37], to ML the number of cavities present in the amoebae from the coefficients. The architectures are the same as in [8]:

MLP: one hidden layer of 100 perceptrons and ReLU activation function.

CNN: four 1d convolutional layers, each followed by a Leaky ReLU layer and a 1d MaxPooling layer.

For all neural networks, we used learning rates of 0.001 and Adam optimizer. We also used a 5-fold cross validation to compute the standard errors. The input data are the coefficients of a particular Newton polynomial and the output is the second Betti number of the corresponding amoeba,

$$\{c_1, \ldots, c_n\} \to b_2 \,. \tag{16}$$

3.3. Example: $\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1$

Consider the example surface of $P(z_1,z_2,z_3) = c_1z_1 + c_2z_1^{-1} + c_3z_2 + c_4z_2^{-1} + c_5z_3 + c_6z_3^{-1} + c_7 = 0$, where the corresponding toric diagram is shown in Figure 4, which is analogous to the toric diagram of $\mathbb{F}_0$ with an extra $\mathbb{P}^1$ fibration. An example of the associated amoeba is given in Figure 5 from

Monte Carlo sampling. Since its toric diagram has only one interior point, the maximum number of 3-dimensional cavities is one, i.e., $b_2 = 0$ or 1, such that this is a binary classification.

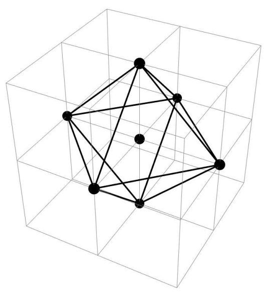


Figure 4. Toric diagram for $\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1$.

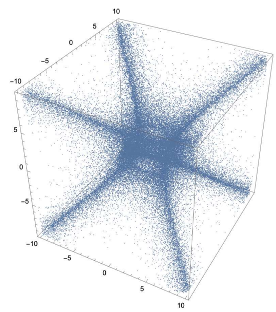


Figure 5. An example of the corresponding $\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1$ amoeba from Monte Carlo sampling.

3.3.1. Learning persistent homology b_2. A balanced dataset of 7200 random samples was generated of real coefficients with

$$c_{(1,0,0)}, c_{(-1,0,0)}, c_{(0,1,0)}, c_{(0,-1,0)}, c_{(0,0,1)}, c_{(0,0,-1)} \in [-5, 5]$$

and $c_{(0,0,0)} \in [-20, 20]$. For each set of coefficients, we used $\mathcal{LA}_{\tilde{f}_1}$ to approximate $\mathcal{A}_f$, and sampled $\mathcal{LA}_{\tilde{f}_1}$ with around 700 points to allow feasible computation. A matrix transformation is performed on the amoeba such that its boundary is clearer while preserving the value of b_2. Then, the point cloud data is passed into the `ripser` package to obtain the persistent pairs of $\mathbf{H}_2$.

After obtaining all the persistent pairs, a selection is required, as when the birth (p) and death (q) times are close to each other this may be the result of point sampling not being dense enough. Thus, persistent pairs (p, q) with $q - p \leq 1.45$ are discarded as noise. This value was selected as a heuristic optimum for the dataset considered, and negligible classification improvements were seen with values ~ 1 and for larger datasets ($\sim 10\times$). Since there is at most one cavity, the value of b_2 is determined by

$$b_2 = \begin{cases} 0 & \text{No persistent pairs with } q - p > 1.45\,; \\ 1 & \text{Otherwise}\,. \end{cases} \tag{17}$$

The identification of the b_2 Betti number from the topological data analysis is exemplified in Figure 6. Within this, the main source of error comes from the number of sampling points and the selection of the persistent pairs.

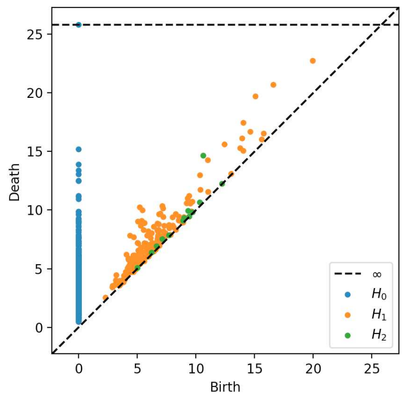


Figure 6. Example of a persistent diagram showing the homology groups H_0, H_1, H_2 for the point cloud data of the amoeba in Figure 1. The H_2 point (10.62652588, 14.6450882) (represented by the green point with coordinates (10.62652588, 14.6450882)) suggests the existence of a 2-dimensional cavity, i.e., $b_2 = 1$.

These b_2 values extracted from the persistent homology where used as the data labels for each amoeba. The subsequent ML hence performed the binary classification task of learning the b_2 value from the input vector of amoeba coefficients. Two NN architectures were used, and classification performance was measured with accuracy as the proportion of correctly predicted b_2 values. Across the 5-fold cross validation runs for both architectures, the performance measures of accuracy (ACC) and Matthews Correlation Coefficient (MCC) were:

$$\text{MLP:} \quad \text{ACC: } 0.771 \pm 0.014\,, \quad \text{MCC: } 0.543 \pm 0.029\,, \tag{18}$$

$$\text{CNN:} \quad \text{ACC: } 0.776 \pm 0.014\,, \quad \text{MCC: } 0.550 \pm 0.031\,. \tag{19}$$

Note also that performance could be marginally improved by increasing the number of sampling points at a cost of longer computation time for the persistent homology.

3.3.2. Learning analytic lopsidedness b_2. This example is simple enough that the condition for the number of cavities (b_2) can be derived in a similar way as for the example of $\mathbb{F}_0$ in $2d$, using lopsidedness. The condition obtained is

$$b_2 = \begin{cases} 0 & |c_7| \leq 2|c_1 c_2|^{1/2} + 2|c_3 c_4|^{1/2} + 2|c_5 c_6|^{1/2}\,; \\ 1 & \text{Otherwise}\,. \end{cases} \tag{20}$$

Now performing the ML using the analytic condition from the lopsided amoeba approximation to generate the b_2 output values for each input amoeba coefficient vector, the results improved. A balanced dataset of 5400 random samples was used to achieve learning measures for each architecture

$$\text{MLP:} \quad \text{ACC: } 0.937 \pm 0.008\,, \quad \text{MCC: } 0.874 \pm 0.016\,, \tag{21}$$

$$\text{CNN:} \quad \text{ACC: } 0.894 \pm 0.014\,, \quad \text{MCC: } 0.789 \pm 0.026\,. \tag{22}$$

The mismatch between two datasets is mostly due to the sampling points not being dense enough such that the separation of the points becomes comparable with the size of the cavity. Plotting the corresponding amoeba shows that it is difficult to tell the number of 3-dimensional cavities by eye in such cases.

3.3.3. MDS projection. Using the `yellowbrick` package [15], multidimensional scaling (MDS) projections (Figure 7) on the dataset obtained via persistent homology and the dataset obtained via analytic condition show similar separations. This MDS method performs non-linear dimensionality reduction of the $\mathbb{R}^7$ space of coefficient vectors into $\mathbb{R}^3$ for effective visualisation, and amoeba (as coefficient vector points) are coloured according to their computed b_2 value (via persistent homology, or analytically).

The difference in these plots may be attributed to poor sampling over the amoeba leading to false results for the persistent homology, or conversely may be due to the error caused by approximating the true amoeba by its lopsided counterpart for the analytic condition derivation. Both these features highlight the subtlety in determining amoebae topology.

Figure 7. MDS manifold projection on dataset obtained using persistent homology (left) and analytic condition (right).

Surface		$\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1$		$\mathbb{P}^3$		$\mathbb{P}^2 \times \mathbb{P}^1$	
		PH	Analytic	PH	Analytic	PH	Analytic
ACC	MLP	0.771 ± 0.014	0.937 ± 0.008	0.840 ± 0.015	0.939 ± 0.009	0.830 ± 0.016	0.947 ± 0.007
	CNN	0.776 ± 0.014	0.894 ± 0.014	0.727 ± 0.031	0.910 ± 0.010	0.825 ± 0.027	0.920 ± 0.011
MCC	MLP	0.543 ± 0.029	0.874 ± 0.016	0.699 ± 0.022	0.876 ± 0.017	0.652 ± 0.035	0.893 ± 0.014
	CNN	0.550 ± 0.031	0.789 ± 0.026	0.457 ± 0.073	0.819 ± 0.019	0.630 ± 0.063	0.841 ± 0.023

Table 1. Summary of the ML results, learning the homology of amoebae constructed from the stated Riemann surfaces with varying coefficients. Learning was performed by MLP and CNN architectures, predicting the b_2 values computed using persistent homology (PH) or lopsidedness (Analytic). Performance was measured with accuracy (ACC) and MCC over the 5-fold cross validation.

3.4. Summary of the ML results

Across the three 3d examples that we considered (details are given in Appendix A), the ML architectures perform similarly learning the b_2 Betti numbers computed from either persistent homology or lopsidedness. For ease of comparison the ML results are repeated for all 3 examples in Table 1, and the MDS projections computed for each in Table 2.

A consistently poorer ML performance for the data obtained using persistent homology is observed across these examples, in comparison to the results using analytic lopsidedness. It is worth noting that this is expected due to the two sources of errors in data generation using persistent homology mentioned in Section 3.3.3 instead of the limitation of ML techniques. Nonetheless, this could, in principle, be improved by using a larger number of sampling points and could be useful in more complex examples where the analytic condition using lopsidedness is absent.

4. Non-Reflexive Mahler measure and Mahler flow

As we mention in Section 2, a reflexive polytope Δ on $\mathbb{Z}^n$ is one whose dual polytope

$$\Delta^\circ = \{\mathbf{v} \in \mathbb{Z}^n | \mathbf{v} \cdot \mathbf{u} \leq -1, \forall \mathbf{u} \in \Delta\} \tag{23}$$

is also reflexive. For $n = 2$, we can show that Δ is reflexive iff the polytope has one interior point. In this section, we focus on polytopes with two interior points, which are therefore non-reflexive.

We can deal with the Mahler measure of non-reflexive polytopes in a similar way to the reflexive case introduced in Section 2. We first consider polynomials of the form $P(z, w) = k_1 - p(z, w)$, where all coefficients of $p(z, w)$

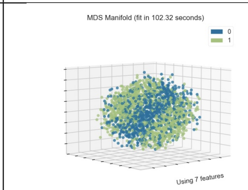
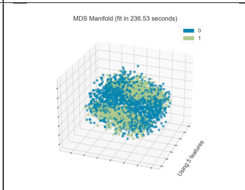
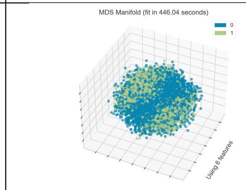
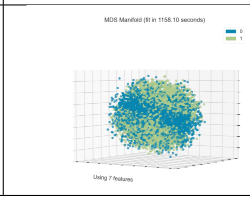
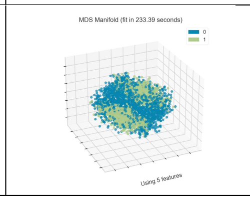
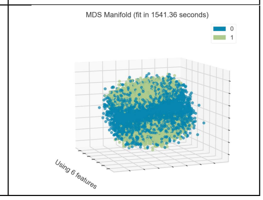
Surface	$\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1$	$\mathbb{P}^3$	$\mathbb{P}^2 \times \mathbb{P}^1$
MDS Projection (PH)			
MDS Projection (Analytic)			

Table 2. MDS projections for each of the three Riemann surface examples considered, colouring according to the b_2 values $\{0, 1\}$ computed via persistent homology (PH) or lopsidedness (Analytic) respectively.

are positive. For polytopes with two interior points, we can write this as $P(z,w) = k_1 - k_2 z^n w^m - p'(z,w)$, where $p'(z,w) = p(z,w) - k_2 z^n w^m$, and the position of the second interior point is (n,m). For cases where $k_2 > \max|k_1 - p'(z,w)|$, we can calculate the Mahler measure $m(P)$ using Cauchy's residue theorem. We factor out $\log(k_2 z^n w^m)$ and are left with:

$$m(P) = \mathrm{Re}\Big(\log k_2 + \frac{1}{(2\pi i)^2}\int_{|w|,|z|=1} \log\left(1 - \frac{1}{k_2 z^n w^m}(k_1 - p'(z,w))\right)\frac{dz}{z}\frac{dw}{w}\Big)\,, \tag{24}$$

where the $\log k_2$ term contributes to the residue, and therefore to the Mahler measure. The $\log(z^n w^m)$ term also contributes to the residue, but since it is purely imaginary, it does not contribute to the measure. To get the full value of the Mahler measure, we expand the $\log\big(1 - (k_2(z^n w^m)^{-1}(k_1 - p'(z,w)))\big)$ in powers of the second argument. A full example of this can be seen in Appendix C.

Similar to Eq. (9), we can also write the above equation as:

$$m(P) = \log k_2 + \int_{k_2}^{\infty}(u_2(t) - 1)\frac{dt}{t}\,, \tag{25}$$

where

$$u_2(k_2) = \frac{1}{(2\pi i)^2}\int_{|w|,|z|=1}\frac{1}{1 - (k_2 z^n w^m)^{-1}(k_1 - p'(z,w))}\frac{dz}{z}\frac{dw}{w}\,. \tag{26}$$

As in the single variable case, $u_2(k_2)$ is the period of a holomorphic 1-form ω_Y on the curve Y defined by $1-(k_2(z^n w^m)^{-1}(k_1 - p'(z,w)))$, and it therefore satisfies the Picard-Fuchs equation [57]:

$$A(k_2)\frac{d^2 u_2(k_2)}{dk_2^2} + B(k_2)\frac{du_2(k_2)}{dk_2} + C(k_2)u_2(k_2) = 0\,. \tag{27}$$

Combined with the similar equation for computing the Mahler measure when the polynomial is lopsided in favour of k_1 (i.e. $k_1 > \max|p(z,w)|$), we now have a means of calculating the measure for the whole k_1k_2-plane except for a strip of width $\sqrt{2}\max|p'(z,w)|$ centered along $k_1 = k_2$. Within these two disconnected parts of the k_1, k_2-plane, the Mahler measure behaves as we expect.

We can redefine the Mahler flow, first introduced in [14], using two equations, one for each disconnected section:

$$\begin{aligned} \frac{\partial m(P)}{\partial \log k_1} &= u_0(k_1) \\ \frac{\partial m(P)}{\partial \log k_2} &= u_2(k_2)\,. \end{aligned} \tag{28}$$

As u_0 and u_2 both represent periods, they are always positive. Therefore, the Mahler measure is always increasing as we move along each flow. When travelling perpendicular to the respective flows, however, this is not necessarily true, as we will see in the next subsection.

As in the reflexive case, many polytopes also have lattice points lying on the edges. We often like to vary the coefficients of these edge points as well as the interior points. If the coefficient of this point is such that the polynomial becomes lopsided in its favour, we deal with it analogously to how we did above. In general, for a polynomial of the form $P(\mathbf{z}) = \sum_{\mathbf{n}} c_{\mathbf{n}} \mathbf{z}^{\mathbf{n}}$, where $\mathbf{z} = (z_1, \ldots, z_i)$ and $\mathbf{n} = (n_1, \ldots, n_i)$ are lattice points, as any $c_{\mathbf{n}}$ tends to ∞, the Mahler measure $m(P)$ tends to $\max_{\mathbf{n}} \log c_{\mathbf{n}}$.

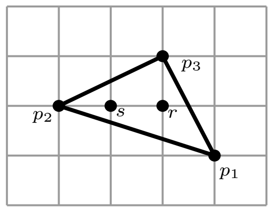
In the limit of large (k_1, k_2), the Mahler measure tends to

$$\max(\log k_1, \log k_2)\,,$$

and we therefore get an infinite measure at the tropical limit at infinity.

4.1. Example: $\mathbb{C}^3/\mathbb{Z}_5$

For a concrete example, we will analyse the surface of $\mathbb{C}^3/\mathbb{Z}_5$ toric Calabi-Yau threefold whose associated toric diagram is pictured below. As mentioned above, for cases where $k_1 > \max|p(z,w)|$ or $k_2 > \max|k_1 - p'(z,w)|$, we can expand the Newton polynomial, taking only the constant term. In this section, we have primarily used `Mathematica` [37] for computations, and the infinite sum is truncated by taking the first 200 terms for a reasonable approximation.



There are two expansions of the Mahler measure, depending on the values of k_1 and k_2. In both cases we take the origin to be the left interior point (labelled s in the above figure). This corresponds to a Newton polynomial given by $P(z,w) = k_1 - k_2 z - zw - z^2 w^{-1} - z^{-1}$. First, we expand for cases where the Newton polynomial is lopsided in favor of k_1. We get a Mahler measure given by:

$$m_1\left(P_s(z,w)\right) = \log k_1 - \sum_{n=1}^{\infty}\sum_{i=0}^{n} \binom{n}{i}\binom{n-i}{\frac{n-i}{2}}\binom{i}{\frac{5i-3n}{4}} \frac{k_2^{\frac{5i-3n}{4}}}{k_1^n n}. \tag{29}$$

Similarly, when the polynomial is lopsided in favor of k_2, we get an expression given by:

$$m_2\left(P_s(z,w)\right) = \log k_2 - \sum_{n=1}^{\infty}\sum_{i=0}^{n} \binom{n}{i}\binom{i}{\frac{i}{2}}\binom{n-i}{\frac{3i-2n}{2}} \frac{k_1^{\frac{4n-5i}{2}}(-1)^{\frac{5i-2n}{2}}}{k_2^n n}. \tag{30}$$

In both cases, we have constraints on allowed combinations of n and i, such that for every binomial coefficient $\binom{n}{r}$ all coefficients are positive integers, and $n \geq r$ (if not, the contributing summand is zero). This greatly reduces the number of terms we need to calculate, reducing the computing time.

From Figure 8, we see that in each component, the Mahler measure increases monotonically along the respective Mahler flows. As we increase the value of k_1 and/or k_2, the plot tends to $\max(\log k_1, \log k_2)$. At large

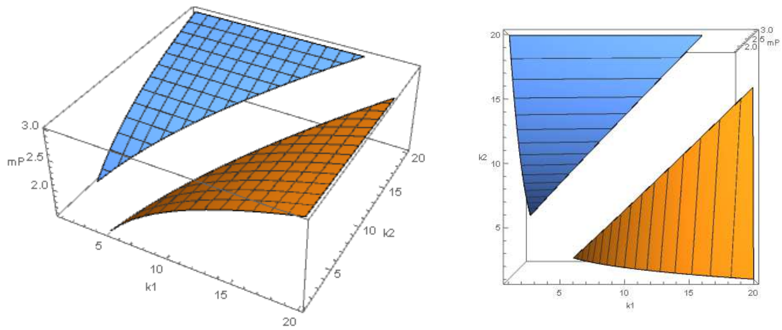


Figure 8. The Mahler measure of the $\mathbb{C}^3/\mathbb{Z}_5$ polynomial. As expected, we see two disconnected components.

values of k_1, the plot therefore looks like $\log k_1$ as we move parallel to the k_1-axis and likewise for k_2. We can explicitly check this monotonic increase by using the definition of the Mahler flow, differentiating the above equations:

$$\frac{\partial m_1(P)}{\partial \log k_1} = 1 + \sum_{n=1}^{\infty}\sum_{i=0}^{n}\binom{n}{i}\binom{n-i}{\frac{n-i}{2}}\binom{i}{\frac{5i-3n}{4}}\frac{k_2^{\frac{5i-3n}{4}}}{k_1^n}. \tag{31}$$

As k_1 and k_2 are always positive, the right hand side of Eq. (31) is clearly always positive and the Mahler measure always increases. This is not necessarily true while travelling along the perpendicular direction (on the same component of the surface). In this case, we obtain

$$\frac{\partial m_1(P)}{\partial \log k_2} = -\sum_{n=1}^{\infty}\sum_{i=0}^{n}\binom{n}{i}\binom{n-i}{\frac{n-i}{2}}\binom{i}{\frac{5i-3n}{4}}\frac{5i-3n}{4}\frac{k_2^{\frac{5i-3n}{4}}}{k_1^n}. \tag{32}$$

One of the conditions for the third binomial coefficient in Eq. (32) to be defined is that $(5i-3n) \geq 0$. Since all other terms are also necessarily positive, this derivative is negative. As we travel along a path of constant k_1 within the k_1 component, the Mahler measure is therefore always decreasing. We can see this behaviour in the orange surface in Figure 8.

Although we observe the same behaviour for the k_2 section of the plot in Figure 8, it is not as immediately obvious from the derivatives. First examining the behaviour along the Mahler flow as defined above, we get:

$$\frac{\partial m_2(P)}{\partial \log k_2} = 1 + \sum_{n=1}^{\infty}\sum_{i=0}^{n}\binom{n}{i}\binom{i}{\frac{i}{2}}\binom{n-i}{\frac{3i-2n}{2}}\frac{k_1^{\frac{4n-5i}{2}}(-1)^{\frac{5i-2n}{2}}}{k_2^n}. \tag{33}$$

The factor of -1 means that we will have some negative terms in the expansion in Eq. (33). In order to have a monotonically increasing Mahler measure, the second term on the right hand side of Eq. (33) must be greater than -1 for all values of (k_1, k_2) within the blue region in Figure 8, i.e., for all values of (k_1, k_2) which satisfy $k_1 < k_2 - 4$. Specifically, we plot this second term for these values of (k_1, k_2) as in Figure 9. We see a decrease in the size of the term as we move along k_1, but it never goes below zero. For the values mentioned above, the sum over n will always converge. This corresponds to the value of each consecutive term decreasing. As we move along k_2, we decrease the size of each term, causing the sum to converge to a smaller number. As $k_2 \to \infty$, this term tends to zero, and the derivative tends to 1, as expected. Moving along k_1 also decreases the Mahler measure, though the gradient is much less than its equivalent in the k_1 section. This is again expected, as all terms in the k_1 section are negative, while the sign of the terms in the k_2 section alternate. This gradient is given by:

$$\frac{\partial m_2(P)}{\partial \log k_1} = -\sum_{n=1}^{\infty}\sum_{i=0}^{n}\binom{n}{i}\binom{i}{\frac{i}{2}}\binom{n-i}{\frac{3i-2n}{2}}\frac{4n-5i}{2}\frac{k_1^{\frac{4n-5i}{2}}(-1)^{\frac{5i-2n}{2}}}{k_2^n n}. \tag{34}$$

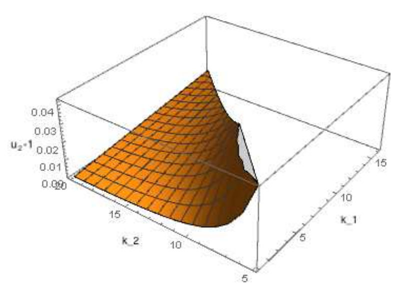


Figure 9. Derivative of the second term in the k_2 section of the $\mathbb{C}^3/\mathbb{Z}_5$ expansion.

4.2. Numerical analysis

Although we cannot obtain a similar expression for the Mahler measure when $|k_1 - k_2| \leq \max|p'(z,w)|$ using the expansion method, we can resort to direct numerical integration to obtain values. Specifically, results from numerical integration in the case of $\mathbb{C}^3/\mathbb{Z}_5$ are plotted in Figure 10. Polynomials whose measure can not be calculated using the expansion method will have poles for certain values of (z, w), which means we will have to integrate over singularities. Nevertheless, results are still accurate to at least 5 decimal places when tested against known exact results, such as those found in [55] and results found using the expansion method above. These singularities correspond to instances when the origin lies within the interior of the related amoeba. In general, shorter computation time is required for numerical integration for polynomials with many terms than using the expansion method above.

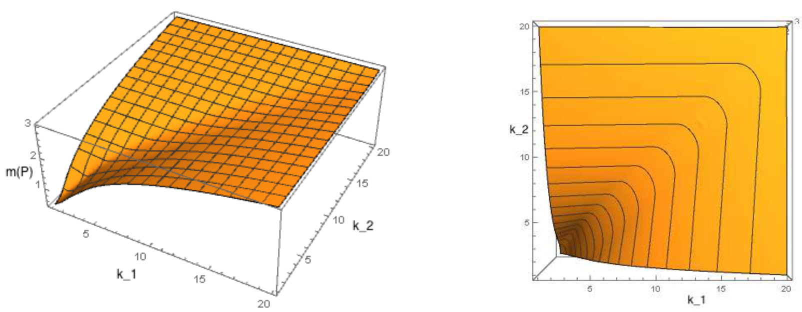
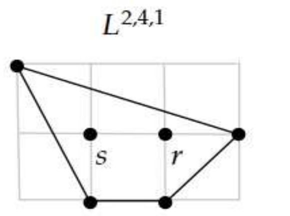
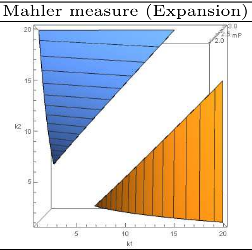
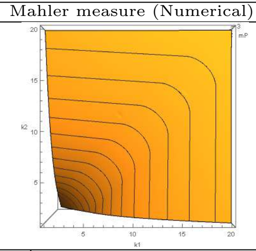


Figure 10. The Mahler measure of the $\mathbb{C}^3/\mathbb{Z}_5$ polynomial calculated numerically.

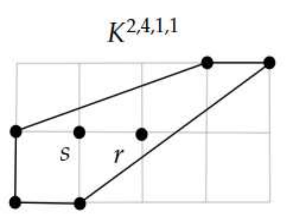
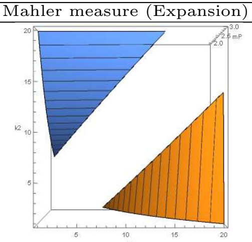
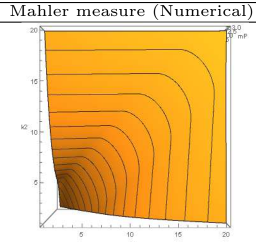
4.3. Summary of results for non-reflexive Mahler measure

We repeated this analysis for more non-reflexive polytopes and obtained expressions for their expansions for large k_1 and k_2, which are summarised in Table 3 for clarity. We also plotted the Mahler measure numerically in each case. Although we only performed these expansions around interior points, similar expressions can be obtained when the polynomial is lopsided in favour of points lying on the polytope edges. As we can see from the plots in Table 3, the numerical and expansion methods give the same results wherever the expansion is defined. Within each section for the expansion plots, the Mahler measure increases monotonically along the Mahler flow, but may decrease or increase when moving perpendicular to it. This variability is particularly visible in the k_2 section, where the expansion series alternates its sign. In the k_1 section, we do not see this as there is no factor of -1, and all terms in the expansion are negative. This results in a decreasing measure.

Toric diagram	Mahler measure (Expansion)	Mahler measure (Numerical)
		

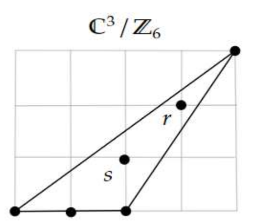
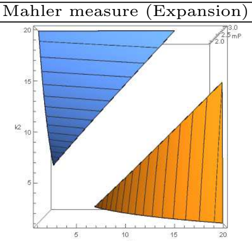
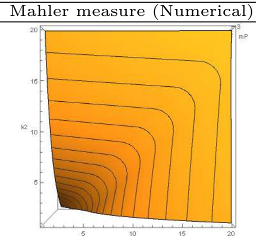
$$m_1\left(P_s(z,w)\right) = \log k_1 - \sum_{n=1}^{\infty}\sum_{i=0}^{n}\sum_{l=0}^{n-i}\binom{n}{i}\binom{i}{\frac{i}{2}}\binom{n-i}{l}\binom{\frac{i}{2}}{\frac{5i+2l-4n}{2}}\frac{k_2^l}{k_1^n n}$$

$$m_2\left(P_s(z,w)\right) = \log k_2 - \sum_{n=1}^{\infty}\sum_{i=0}^{n}\sum_{l=0}^{n-i}\binom{n}{i}\binom{i}{\frac{i}{2}}\binom{n-i}{l}\binom{\frac{i}{2}}{n-2l-2i}\frac{k_1^l(-1)^{n-2i-l}}{k_2^n n}$$

Toric diagram	Mahler measure (Expansion)	Mahler measure (Numerical)
		

$$m_1\left(P_s(z,w)\right) = \log k_1 - \sum_{n=1}^{\infty}\sum_{i=0}^{n}\sum_{l=0}^{n-i}\sum_{h=0}^{\frac{i}{2}}\binom{n}{i}\binom{n-i}{l}\binom{i}{\frac{i}{2}}\binom{\frac{i}{2}}{n+h-2i-2l}\frac{k_2^l}{k_1^n n}$$

$$m_2\left(P_s(z,w)\right) = \log k_2 - \sum_{n=1}^{\infty}\sum_{i=0}^{n}\sum_{l=0}^{n-i}\sum_{h=0}^{\frac{i}{2}}\binom{n}{i}\binom{i}{\frac{i}{2}}\binom{n-i}{l}\binom{\frac{i}{2}}{h}\binom{\frac{i}{2}}{h+2n-l-2i}\frac{k_1^l(-1)^{n-l}}{k_2^n n}$$

Toric diagram	Mahler measure (Expansion)	Mahler measure (Numerical)
		

$$m_1\left(P_s(z,w)\right) = \log k_1 - \sum_{n=1}^{\infty}\sum_{i=0}^{n}\sum_{l=0}^{i}\binom{n}{i}\binom{n-i}{2n-3i}\binom{i}{l}\binom{l}{2l-i}\frac{k_2^{2n-3i}}{k_1^n n}$$

$$m_2\left(P_s(z,w)\right) = \log k_2 - \sum_{n=1}^{\infty}\sum_{i=0}^{n}\sum_{l=0}^{i}\binom{n}{i}\binom{n-i}{\frac{n-3i}{2}}\binom{i}{l}\binom{l}{i-l}\frac{k_1^{\frac{n-3i}{2}}(-1)^{\frac{n+3i}{2}}}{k_2^n n}$$

Table 3. Summary of results for some non-reflexive polytopes. Plots generated by the expansion method and the numerical method are consistent with each other. As we travel along the Mahler flow, the Measure increases monotonically.

5. ML, amoeba, and the Mahler measure

It is noticed in [14] that the changes in the liquid and gas phase contributions to the Mahler measure along the Mahler flow are similar to the changes in the area of the amoeba and the area of the bounded amoeba complement. The conjecture is that given a Newton polynomial $P(z,w) = k - p(z,w)$, the liquid phase contribution to the Mahler measure $m_l(P)$ is solely determined by the area of the amoeba and the gas phase contribution $m_g(P)$ is solely determined by the area of the bounded amoeba complement, i.e., its hole.

As introduced in Section 2.3, the relation between Mahler measure and amoeba is evident via the Ronkin function. The amoeba is the region where the gradient of the Ronkin function is non-linear, whereas the Mahler measure is the Ronkin function evaluated at $(0,0)$. It is possible to use ML to make this relation more precise.

5.1. Area of the bounded amoeba complement

Only reflexive polytopes as toric diagrams are considered such that the definition of the gas phase contributions to the Mahler measure is most obvious. Thus, we are only considering a single bounded region for the amoeba. The area of this bounded amoeba complement (the amoeba hole), A_h, is obtained using both sampling and analytic solutions as a crosscheck for each other.

It is possible to sample only the bounded complement of the amoeba using lopsidedness and restricting the sampled region to the bounded region formed by its spines. This bounded region formed by its spines can be determined from Theorem 3.7 in [14].

The analytic boundary of the amoeba is derived by considering the boundary conditions where the gradient of the Ronkin function changes from being linear to being non-linear. This is when the pole in the gradient of the Ronkin function, i.e., $P(z,w) = 0$, within the integration path is independent of the phase angle of $w = |w|e^{i\theta}$ at constant $y = \ln|w|$, following the considerations in [46].

The areas obtained from both methods agree with each other rather well, so we choose to use the analytic solutions in this section for the ease of computation.

5.2. Symbolic regression and genetic algorithm

Symbolic regression is a machine learning technique which allows us to determine the mathematical relationship between the independent variables and the dependent variable targets. Genetic Programming refers to the technique of automated evolution of programs, usually starting from random programs which are progressively evolved using operations analogous to naturally occurring genetic operations. The `gplearn` package is an implementation of Genetic Programming to perform symbolic regression. It first generates a population of random formulas and then each subsequent population is obtained by performing genetic operations on the fittest individuals from the preceding population. With the help of `gplearn`, we were able to obtain numerical relations between the area(volume) of the amoeba hole and the coefficients of the Newton polynomial and the numerical relations between the area(volume) of the amoeba hole and gas phase contribution to the Mahler measure.

Specifically, in this section, the genetic algorithm has the following structure in which equations are represented as trees with selected operations from {addition, subtraction, multiplication, division, negation, square root, logarithm, inverse, absolute value} applied to variables and constants in the range $(-10, 10)$. It begins by initialising with a random population of size 5000. The raw fitness metric, the mean absolute error (MAE) in this case, of the true output values for all input values is calculated for each equation in the population to give a performance loss which is weighted by the complexity of the equation with weight 0.02. Then, the fittest 0.4 percent of the population are selected to evolve to successive generation of equations via the genetic operations including performing crossover with probability of 0.85, subtree mutation with probability of 0.02, leave mutation with probability 0.01, and hoist mutation with probability of 0.015. This process is iterated for 100 generations, and equations are selected early if the metric score reaches 0.001.

5.3. 2d Example: $\mathbb{F}_0 = \mathbb{P}^1 \times \mathbb{P}^1$

The Newton Polynomial in this case is $P(z, w) = k - z - z^{-1} - w - w^{-1}$. The analytic boundary of the amoeba is found to be

$$x = \ln \left(\left| \frac{k}{2} \pm \cosh y \pm \sqrt{\left(\frac{k}{2} \pm \cosh y\right)^2 - 1} \right| \right), \tag{35}$$

where $x = \ln|z|, y = \ln|w|$. The boundaries of amoebae with $k = -0.5, 4, 10$ are plotted in Figure 11, where $k = 4$ is the critical value of k at which the amoeba hole starts to appear. The boundary of the hole agrees well with the sampled boundary of the amoeba. The areas of the hole obtained by sampling and by analytic boundaries agree with each other to at least 2 decimal places depending on the density of points.

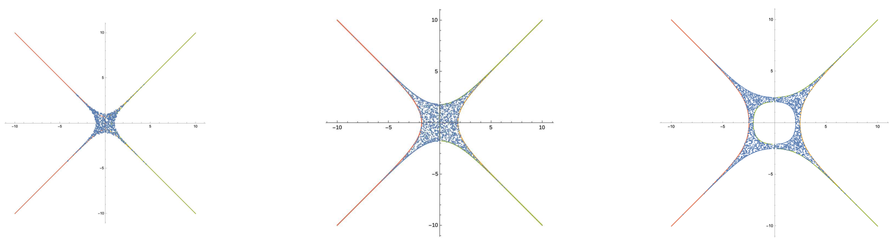


Figure 11. The analytic boundaries of amoebae with $k = -0.5$ (left), $k = k_c = 4$ (middle) and $k = 10$ (right).

The relation between the amoeba hole area and the value of k is fitted with 7500 pairs of (k, A_h) and is found to be

$$A_h = 4 \ln^2 k - 6.601\,, \tag{36}$$

with an R^2 score of 1.0000 and a mean absolute error of 0.0318. In the limit of $k \to \infty$, the leading term scales as $4 \ln^2 k$ (as plotted in Figure 12). This agrees with Conjecture 3.9 in [14], and the power of $\ln k$ is given by the dimension of the amoeba.

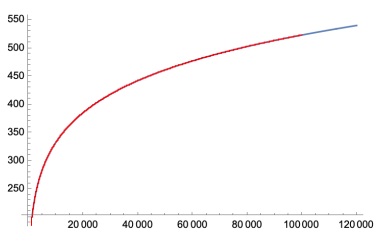


Figure 12. The relation between amoeba hole area A_h and k. Original data points are shown red, and the relation found is plotted blue.

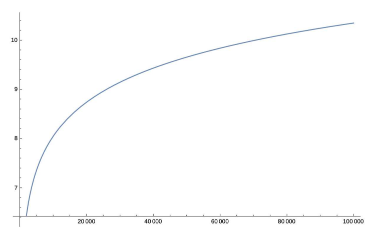


Figure 13. Plot of gas phase contribution to Mahler measure m_g against k.

5.3.1. Mahler measure and hole area for $k \geq 4$. The Mahler measure for the Newton polynomial $P(z, w) = k - z - z^{-1} - w - w^{-1}$ is

$$m(P) = \ln k - 2k^{-2}{}_4F_3\left(1, 1, \frac{3}{2}, \frac{3}{2}; 2, 2, 2; 16k^{-2}\right), \quad (37)$$

for $k \geq 4$ [14]. The gas phase contribution to the Mahler measure, $m_g(P) = m(P) - m(P(k = 4))$, is plotted in Figure 13 which shows similar trend as in Figure 12. This suggests a possible direct relationship between A_h and $m_g(P)$, as motivated in Section 2.

The relation between the gas phase contribution $m_g(P)$ and the amoeba hole area is fitted with about 50000 data pairs and found to be

$$A_h = 3.9804m_g^2 + 9.888m_g - 1.5243\sqrt{m_g}, \quad (38)$$

with an R^2 score of 1.0000 and a mean absolute error of 0.0249 (plotted in Figure 14). In the limit of large k and hence large m_g, $A_h \sim 4m_g^2$ and the leading coefficient in Eq. (38) is close to the leading coefficient in Eq. (36). This is expected as the large k behaviour of the Mahler measure is of $\ln k$.

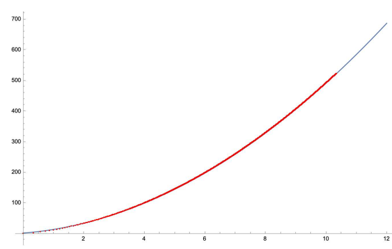


Figure 14. Data points (red) and the fitted relation (blue) between A_h and m_g.

5.4. 2d Example with more than one parameter: $Y^{2,2}$

We also considered an example which has more than one coefficient of the Newton polynomial that is not constant. Specifically, we looked at the surface $Y^{2,2}$ with the associated Newton polynomial $P(z, w) = z^2 + bz + k + r(w + w^{-1})$ and coefficients (b, k, r). Its toric diagram is given in Figure 15. The physical interpretation of the coefficients can be found in [46].

The analytic boundaries of the amoeba can be determined using the same method as before, and they are given by

$$x = \ln \left| -\frac{b}{2} \pm \sqrt{\frac{b^2}{4} - (k \pm 2r \cosh y)} \right| , \tag{39}$$

where $x = \ln |z|, y = \ln |w|$. The boundary of the amoeba hole agrees with the sampled boundary of the $n = 1$ lopsided amoeba, as shown in Figure 16.

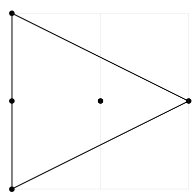


Figure 15. The toric diagram associated with $Y^{2,2}$.

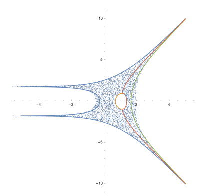


Figure 16. The boundary curves of the amoeba.

The relation between the coefficients b, k, r and the amoeba hole area is fitted with 40328 pairs of $(\{b, k, r\}, A_h)$ where $0 < b, k, r \leq 40$ and $A_h \neq 0$, and it is found to be

$$A_h = \sqrt{\left(b + 2.4142\sqrt{r} - r\right)\left(\frac{b}{\sqrt{r}} - \ln k\right)} \tag{40}$$

with an R^2 score of 0.9519 and a mean absolute error of 1.0478. The area of the amoeba hole does not scale as $\ln^2 k$ in the large k limit at constant b and r, and it is most significantly affected by the value of b instead of k.

Moreover, the gas phase contribution to the Mahler measure cannot be analogously defined here because the Mahler measure takes different values for coefficients that give the same hole area. An example is given in Table 4.

(b, k, r)	$(4, 1, 1)$	$(8, 4, 4)$	$(12, 9, 9)$	$(16, 16, 16)$	$(20, 25, 25)$	$(24, 36, 36)$
A_h	1.63644	1.63644	1.63644	1.63644	1.63644	1.63644
$m(P)$	1.43518	2.17779	2.7313	3.15535	3.51961	3.8322

Table 4. Example of numerical values of the Mahler measure for different sets of coefficients (b, k, r) with the same amoeba hole area.

5.5. Summary of 2d results

The 2d compact Fano varieties considered in this work, along with the respective Newton polynomials used and toric diagrams, are given in Table 5. The amoebae and Mahler measure information for each are respectively collected in Table 6 for ease of comparison; along with the symbolic regression results in Table 7. Another detailed example is given in Appendix B.

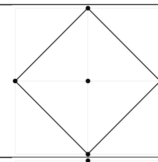
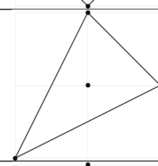
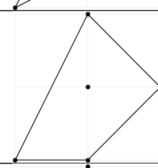
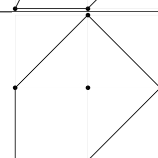
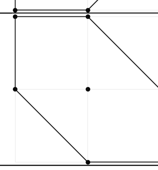
Surface	Newton Polynomial	Toric Diagram
$\mathbb{F}_0$	$P(z,w) = k - (z + z^{-1} + w + w^{-1})$	
$\mathbb{P}^2(\mathrm{dP}_0)$	$P(z,w) = k - (z + w + z^{-1}w^{-1})$	
dP_1	$P(z,w) = k - (z + w + w^{-1} + z^{-1}w^{-1})$	
dP_2	$P(z,w) = k - (z + z^{-1} + w + w^{-1} + z^{-1}w^{-1})$	
dP_3	$P(z,w) = k - (z + z^{-1} + w + w^{-1} + zw^{-1} + z^{-1}w)$	

Table 5. Examples of toric surfaces, each with an associated specific Newton polynomial and the respective toric diagram.

5.6. 3d Example: $\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1$

Similar methods are applied to the 3-dimensional example of $\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1$. The analytic expressions for the boundary surfaces of the associated amoeba

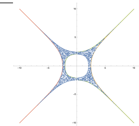
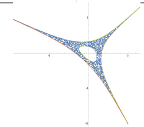
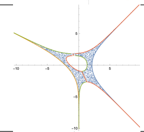
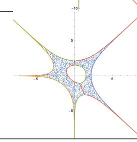
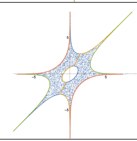
$\mathcal{V}$	Amoeba Boundary	Example	Mahler measure
$\mathbb{F}_0$	$x = \ln\left(\left\lvert\frac{k}{2} \pm \cosh y \pm \sqrt{\left(\frac{k}{2} \pm \cosh y\right)^2 - 1}\right\rvert\right)$		$m(P) = \ln k - 2k^{-2}{}_4F_3\left(1,1,\frac{3}{2},\frac{3}{2};2,2,2;16k^{-2}\right)$
$\mathbb{P}^2$ (dP_0)	$x = \ln\left(\left\lvert\frac{k}{2} \pm \frac{e^y}{2} \pm \sqrt{\frac{1}{4}\left(k \pm e^y\right)^2 \pm e^{-y}}\right\rvert\right)$		$m(P) = \ln k - 2k^{-3}{}_4F_3\left(1,1,\frac{4}{3},\frac{5}{3};2,2,2;27k^{-3}\right)$
dP_1	$x = \ln\left(\left\lvert\frac{k}{2} \pm \cosh y \pm \sqrt{\left(\frac{k}{2} \pm \cosh y\right)^2 \pm e^{-y}}\right\rvert\right)$		$m(P) = \ln k - \sum_{n=1}^{\infty}\sum_{i=0}^{n}\frac{1}{nk^n}\binom{n}{i}\binom{i}{\frac{i}{2}}\binom{n-i}{\frac{2n-i}{4}}$, $2n - i \mod 4 = 0$ and $i \mod 2 = 0$.
dP_2	$x = \ln\left(\left\lvert\frac{k}{2} \pm \cosh y \pm \sqrt{\left(\frac{k}{2} \pm \cosh y\right)^2 - (1 \pm e^{-y})}\right\rvert\right)$		$m(P) = \ln k - \sum_{n=1}^{\infty}\sum_{i=0}^{n}\sum_{j=0}^{i}\frac{1}{nk^n}\binom{n}{i}\binom{n-i}{\frac{n-i}{2}}\binom{i}{j}\binom{j}{\frac{i}{2}}$, $n - j \mod 2 = 0$, $i \mod 2 = 0$, and $j \geq \frac{i}{2}$.
dP_3	$x = \ln\left(\left\lvert(1 \pm e^{-y})^{-1}\left(\frac{k}{2} \pm \cosh y \pm \sqrt{\left(\frac{k}{2} \pm \cosh y\right)^2 - (2 \pm 2\cosh y)}\right)\right\rvert\right)$		$m(P) = \ln k - \sum_{n=1}^{\infty}\sum_{i=0}^{n}\sum_{l=0}^{i}\sum_{j=0}^{n-i}\frac{1}{nk^n}\binom{n}{i}\binom{i}{l}\binom{l}{\frac{l+n-i-2j}{2}}\binom{i-l}{\frac{2i+2j-l-n}{2}}\binom{n-i}{j}$, $l + n \mod 2 = 0$, $i \mod 2 = 0$, and $n - i - 2j \leq l \leq n - 2j$.

Table 6. Summary of 2d results for surfaces $\mathcal{V}$.

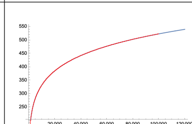
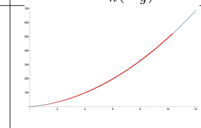
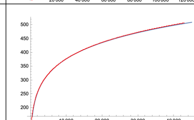
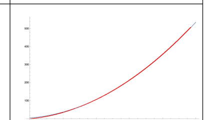
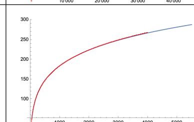
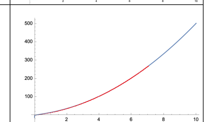
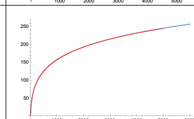
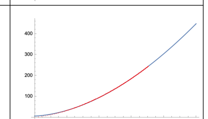
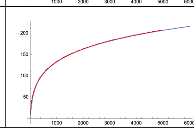
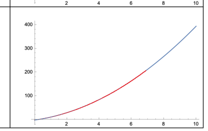
$\mathcal{V}$	Fitted $A_h(k)$	Plot $A_h(k)$	Fitted $A_h(m_g)$	Plot $A_h(m_g)$
$\mathbb{F}_0$	$A_h = 4\ln^2 k - 6.601$ with an R^2 score of 1.0000 and a mean absolute error of 0.0318.		$A_h = 3.9804m_g^2 + 9.888m_g - 1.5243\sqrt{m_g}$ with an R^2 score of 1.0000 and a mean absolute error of 0.0249.	
$\mathbb{P}^2$ (dP_0)	$A_h = 2\ln^2 k + \ln(k\ln k) - 5.49 + \ln k^2 \ln\big((k - \ln k^2)(\ln(k\ln k) - 5.49)\big)$ with an R^2 score of 0.9998 and a mean absolute error of 1.0983.		$A_h = 4.5904m_g^2 + 7.2486m_g + 5.0980$ with an R^2 score of 1.0000 and a mean absolute error of 0.1340.	
dP_1	$A_h = 3.891\ln(k - 9.457) \times \ln(k - 3.891\ln(x - 9.457))$ with an R^2 score of 0.9994 and a mean absolute error of 0.8515.		$A_h = 4.146m_g^2 + 8.291m_g + \ln m_g + 2.958$ with an R^2 score of 1.0000 and a mean absolute error of 0.1514.	
dP_2	$A_h = 3.891\ln k \ln\Big(0.246k + \frac{k}{\ln(k\ln k)}\Big)$ with an R^2 score of 0.9998 and a mean absolute error of 0.4549.		$A_h = 5.471m_g^2\sqrt{1 - 0.194\ln m_g} + 3.407m_g + 7.294$ with an R^2 score of 1.0000 and a mean absolute error of 0.1442.	
dP_3	$A_h = 2.783(\ln k + 1) \times (\ln k - \frac{1}{\ln(2.510(0.025k+1)^{1/4})}) - 5.809$ with an R^2 score of 1.000 and a mean absolute error of 0.2014.		$A_h = 3m_g^2 + 9.623m_g - 1.573$ with an R^2 score of 1.0000 and a mean absolute error of 0.0618.	

Table 7. Summary of 2d results for surfaces $\mathcal{V}$.

are given by (Figure 17)

$$x = \ln\left(\left|\frac{|k|}{2} \pm \cosh y \pm \cosh z \pm \sqrt{\left(\frac{|k|}{2} \pm \cosh y \pm \cosh z\right)^2 - 1}\right|\right), \tag{41}$$

where $x = \ln|u|, y = \ln|v|, z = \ln|w|$ and the $\pm$ sign in front of two cosh's are the same for both y and z.

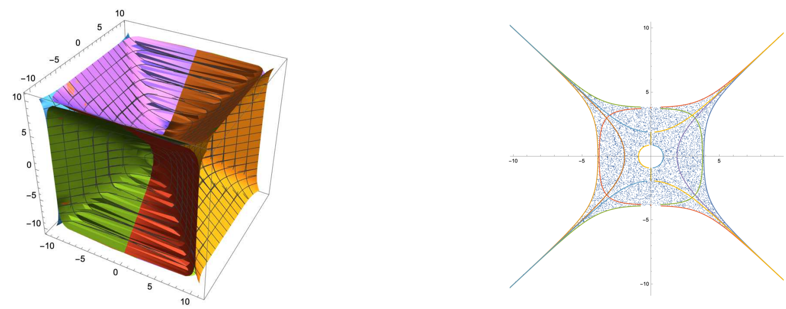


Figure 17. Boundary surfaces of the amoeba corresponding to $P(u,v,w) = k - u - u^{-1} - v - v^{-1} - w - w^{-1}$ and its cross-section at $z = 0$.

In particular, the boundary surfaces of the amoeba hole are formed by

$$\begin{aligned} x = \ln\Bigg(\Bigg|\frac{|k|}{2} &- \cosh y - \cosh z \\ &\pm \sqrt{\left(\frac{|k|}{2} - - \cosh y - \cosh z\right)^2 - 1}\Bigg|\Bigg). \end{aligned} \tag{42}$$

The numerical relation between the volume of the bounded complement of the amoeba, V_h, and the value of k is found by fitting 20000 pairs of (k, V_h) values. It is found to be

$$\begin{aligned} V_h(k) &= \ln(k + \ln(0.4854k)) \\ &\times \left(8.374 \ln\left(0.4854k - \ln^2(0.4854k)\right) + \ln\left(\frac{7.19}{\ln k}\right)\right) \ln(k + \ln k)\,, \end{aligned} \tag{43}$$

with an R^2 score of 1.0000 and a mean absolute error of 3.2689 (plotted in Figure 18). In the limit of $k \to \infty$, the leading term scales as $\ln^3 k$, which again agrees with Conjecture 3.9 in [14].

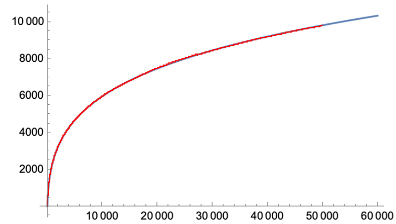


Figure 18. Data points (red) and the fitted relation (blue) for V_h against k.

Using Taylor expansion and Cauchy residue theorem, Mahler measure for $P(u, v, w) = k - u - u^{-1} - v - v^{-1} - w - w^{-1}$ as a function of k for $k > 6$ is found to be

$$m(P) = \ln k - \sum_{n=1}^{\infty} \frac{1}{2nk^{2n}} \binom{2n}{n} \sum_{l=0}^{n} \binom{2l}{l} \binom{n}{l}^2. \tag{44}$$

If there exists an associated 3-dimensional dimer model that allows a similar interpretation of the Mahler measure in different phases, the gas phase contribution to the Mahler measure $m_g(P)$ may be analogously defined as $m_g(P) = m(P) - m(P(k_c))$. For the expansion method used in obtaining Eq. (44) to be valid, k must be greater than $\max_{|u|,|v|,|w|=1} |u + u^{-1} + v + v^{-1} + w + w^{-1}| = 6$. Thus, the critical value of k is $k_c = 6$, which is also the value at which the bounded amoeba complement begins to form. The relation between the volume of the amoeba hole and the analogously defined m_g is fitted with 5000 pairs of values, and is found to be

$$\begin{aligned} V_h = 3.835 \Big| m_g \big(7.968 \\ - \sqrt{m_g}(-9.743 m_g + \ln(m_g - 9.664) + \ln(0.170 m_g))\big) \Big|, \end{aligned} \tag{45}$$

with an R^2 score of 1.0000 and a mean absolute error of 6.0229 (plotted in Figure 19). In the large k limit, V_h is found to scale as $m_g^{5/2}$ which deviates from the expected power of 3. This may be due to the erratic nature of genetic algorithm, so we explicitly tested this conjectured relation again with specific ansatz which will be elaborated in the following subsection.

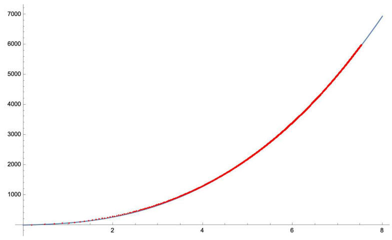


Figure 19. Data points (red) and the fitted relation (blue) between gas phase Mahler measure m_g and amoeba hole volume V_h.

5.7. Summary of 3d results

In Tables 8, 9, and 10, we summarise our results of ML the relationship between the coefficient k, the volume of bounded complementary region of amoeba, and the Mahler measure for clarity.

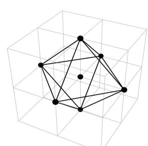
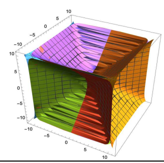
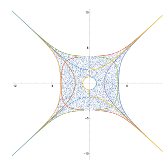
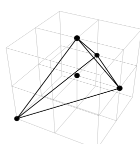
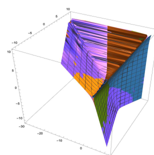
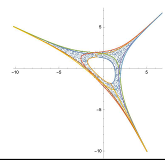
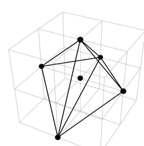
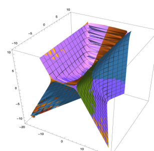
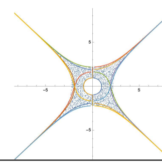
Surface	Newton Polynomial	Toric Diagram	Boundary	Cross section at z=0
$\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1$	$P(z,u,w) = k - (z + z^{-1} + u + u^{-1} + w + w^{-1})$			
$\mathbb{P}^3$	$P(z,u,w) = k - (z + u + w + z^{-1}u^{-1}w^{-1})$			
$\mathbb{P}^2 \times \mathbb{P}^1$	$P(z,u,w) = k - (z + u + w + z^{-1} + u^{-1}w^{-1})$			

Table 8. Examples of 3d Fano varieties and their associated Newton polynomials, toric diagrams, plots of the boundary, and cross-sections of their amoebae.

Moreover, since the results obtained using symbolic regression in the 3-dimensional case are too complicated to be useful, we also included the results obtained using NonlinearModelFit in `Mathematica` [37] in Table 11 to

Surface	Amoeba Boundary	Mahler measure
$\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1$	$x = \ln\left(\left\lvert \frac{\lvert k\rvert}{2} \pm \cosh y \pm \cosh z \pm \sqrt{\left(\frac{\lvert k\rvert}{2} \pm \cosh y \pm \cosh z\right)^2 - 1}\right\rvert\right)$	$m(P) = \ln k - \sum_{n=1}^{\infty} \sum_{l=0}^{n} \frac{1}{2nk^{2n}} \binom{2n}{n}\binom{2l}{l}\binom{n}{l}^2$
$\mathbb{P}^3$	$x = \ln\left(\left\lvert \frac{\lvert k\rvert - (\pm e^y \pm e^z)}{2} \pm \sqrt{\left(\frac{\lvert k\rvert - (\pm e^y \pm e^z)}{2}\right)^2 - (\pm e^{-y})(\pm e^{-z})}\right\rvert\right)$	$m(P) = \ln k - \sum_{n=1}^{\infty} \frac{1}{4nk^{4n}} \binom{4n}{2n}\binom{2n}{n}^2$
$\mathbb{P}^2 \times \mathbb{P}^1$	$x = \ln\left(\left\lvert \frac{\lvert k\rvert - (\pm e^y \pm e^z + (\pm e^{-y})(\pm e^{-z}))}{2} \pm \sqrt{\left(\frac{\lvert k\rvert - (\pm e^y \pm e^z + (\pm e^{-y})(\pm e^{-z}))}{2}\right)^2 - 1}\right\rvert\right)$	$m(P) = \ln k - \sum_{n=1}^{\infty} \sum_{i \geq \frac{n}{2}}^{n} \frac{\binom{n}{i}\binom{i}{2n-3i}\binom{n-i}{2n-3i}\binom{4i-2n}{2i-n}}{nk^n}$

Table 9. Summary of 3d results.

Surface	Fitted $V_h(k)$	Fitted $V_h(m_g)$
$\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1$	$V_h(k) = \ln(k + \ln(0.4854k))\Big(8.374 \ln\left(0.4854k - \ln^2(0.4854k)\right) + \ln\left(\frac{7.19}{\ln k}\right)\Big) \ln(k + \ln k)$ with an R^2 score of 1.0000 and a mean absolute error of 3.2689.	$V_h = 3.835\lvert m_g\big(7.968 - \sqrt{m_g}(-9.743m_g + \ln(m_g - 9.664) + \ln(0.170m_g))\big)\rvert$ with an R^2 score of 1.0000 and a mean absolute error of 6.0229. $k_c = 6$ and $m_g(k) = m(P(k)) - m(P(6))$.
$\mathbb{P}^3$	$V_h = \Big\lvert -0.380\Big(-3.732 + \pi i + 2.397\sqrt{0.174k - 1}\Big(3.724k - 73.933 \times \Big\lvert\ln\lvert\sqrt{k} - 314.956\rvert(k - 5.744)(2.489k - k^{-0.5})\rvert(0.174k - 1)^2 k^{9.5}\rvert\Big\rvert\Big) \times (2.397\sqrt{0.174k - 1} - 0.489k)^{-1} - 73.933\sqrt{k} + 1.489k\Big) \times (\ln\lvert 7.888k - 731.045\rvert)^{-0.5} + 244.509\sqrt{\lvert 0.174k - 1\rvert} - 0.186k - 1.074\Big\rvert$ with an R^2 score of 1.0000 and a mean absolute error of 4.1666.	$V_h = \left(\sqrt{m_g} + 7.95\right)\left(1.523m_g + 12.4201\right)\Big(0.4794m_g^2 + \Big(0.0264\big(3.385 + 5.424/m_g + 0.0028/\sqrt{m_g} + 0.8802m_g - 0.4794m_g^2\big)\Big)/\sqrt{m_g}\Big)$ with an R^2 score of 0.9998 and a mean absolute error of 1.5644. $k_c = 4$ and $m_g(k) = m(P(k)) - m(P(4))$.
$\mathbb{P}^2 \times \mathbb{P}^1$	$V_h = \Big\lvert\frac{1}{0.0068\ln k - \frac{0.0068}{0.215 - 0.0068\sqrt{k}}} + (256.761 + 5.049\pi i)\sqrt{0.118k - 1} - \sqrt{k} + \sqrt{\frac{1}{0.0068\ln k - 0.0068\sqrt{k}} + 0.550k} + 4.174\sqrt{\frac{k}{\ln k - \sqrt{k}} + \frac{k}{\ln k}} + 4.174\sqrt{-\frac{k}{\ln\left(2k - 72.719k\sqrt{\frac{k}{\ln k - \sqrt{k}} + \frac{k}{\ln k}} + 0.0061\right) - \sqrt{k}}} + \frac{1}{0.0045\log(k - 8.338) - 0.0045\sqrt{k}} - 0.214k + \ln k\Big\rvert$ with an R^2 score of 0.9999 and a mean absolute error of 8.0802.	$V_h = \Big\lvert 53.0889 + (m_g^2 - 0.054m_g - 1.09316)\Big(53.0889 + \Big(\Big(-2m_g + (\frac{0.146092}{m_g - 8.175} - m_g - 3.60814) \times (\frac{9.966}{m_g - 8.175} - 7.324m_g(m_g + 2.04377i)))\ln^{-1} m_g\Big)\Big)\Big\rvert$ with an R^2 score of 1.0000 and a mean absolute error of 3.1879. $k_c = 5$ and $m_g(k) = m(P(k)) - m(P(5))$.

Table 10. Summary of 3d results using symbolic regression.

specifically test Conjecture 3.9 in [14]. Notably, the use of NonlinearModelFit or other fitting functions in `Mathematica` requires an input assumption of the structure of the fit model. In this case, a cubic equation in $\ln k$ and a cubic equation in m_g were assumed based on the results of symbolic regression in two dimensions. The results are relatively accurate based on their R^2 values and the mean prediction errors. Using the functional form obtained from ML directly in `Mathematica` provides us with a faster way to get better fitting results in comparison to speculating possible functional forms as inputs to try in `Mathematica`. For example, if we make a guess of a second-order equation in m_g for the relation between V_h and m_g based on the plot of the data in the case of $\mathbb{P}^2 \times \mathbb{P}^1$, the fitted relation obtained is $V_h = 138.74m_g^2 - 321.537m_g + 346.213$. It has a mean prediction error of 0.5920 which is much greater than the error of 0.0131 using the cubic relation based on previous ML results.

Results in Table 11 agree with Conjecture 3.9 in [14] in the $n = 3$ case: In the large k limit, the volume of a bounded complementary region of the amoeba, V_h, is cubic in $\ln k$. However, we also notice that it is not possible to generalise an analytic expression for V_h analogous to the expression in Conjecture 3.8 in [14], because the volume of the 3-dimensional amoeba is almost always infinite whereas the area of the 2-dimensional amoeba is bounded from above [52].

5.8. Non-reflexive polytopes

Following our consideration of the non-reflexive case in Section 4, it is interesting to also look at the relation between amoeba holes and the Mahler measure in this case. Amoebae for non-reflexive polytopes can have a geometric genus ranging from 1 to n, where n is the number of interior points in the corresponding Newton polytope. It is noted in [40] that for amoebae with all holes open, a decrease in the size of one hole corresponds to an increase in the size of all others.

Specifically, we are going to consider here the polytopes with two interior points, which corresponds to amoebae with a maximum of two bounded complementary regions. As we mention in Section 4, the Mahler measure can now be represented as a function of two variables, k_1 and k_2, which correspond the two interior points. We can make a choice of which interior point we use as the origin. Where in the one dimensional case, we can generally find a critical value for k at which the gas phase emerges in the amoeba, in two or more dimensions, we instead get a set of values for (k_1, k_2) where gas phases emerge. We get another set of (k_1, k_2) points where the genus of the

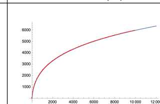
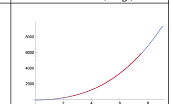
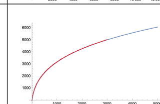
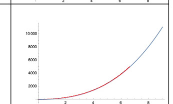
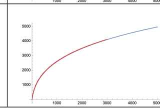
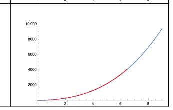
Surface	Fitted $V_h(k)$	Plot $V_h(k)$	Fitted $V_h(m_g)$	Plot $V_h(m_g)$
$\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1$	$V_h = 7.981 \ln^3 k + 0.292 \ln^2 k - 26.439 \ln k - 17.822$ with an R^2 score of 1.0000 and a mean prediction error of 0.0175.		$V_h = 7.912 m_g^3 + 41.381 m_g^2 + 36.856 m_g - 17.055$ with an R^2 score of 1.0000 and a mean prediction error of 0.0357.	
$\mathbb{P}^3$	$V_h = 10.299 \ln^3 k - 0.941 \ln^2 k - 25.871 \ln k + 4.027$ with an R^2 score of 1.0000 and a mean prediction error of 0.0051.		$V_h = 10.294 m_g^3 + 40.936 m_g^2 + 27.867 m_g - 6.787$ with an R^2 score of 1.0000 and a mean prediction error of 0.0055.	
$\mathbb{P}^2 \times \mathbb{P}^1$	$V_h = 8.196 \ln^3 k + 1.986 \ln^2 k - 26.103 \ln k - 9.320$ with an R^2 score of 1.0000 and a mean prediction error of 0.0123.		$V_h = 8.169 m_g^3 + 40.011 m_g^2 + 35.901 m_g - 12.671$ with an R^2 score of 1.0000 and a mean prediction error of 0.0131.	

Table 11. Summary of 3d results using NonlinearModelFit in `Mathematica` [37].

amoeba changes from 1 to 2. Each of these gas phases appear and disappear individually depending on both the value of k_1 and of k_2.

Based on our observation, for polytopes with two interior points, there are three ways that the bounded complementary regions of amoeba can evolve as we fix k_1 and move along k_2, or vice versa:

1) There are initially no holes. At some critical value of k_2 a hole opens up and continues to grow as we increase $k_2 \to \infty$. This is similar to the reflexive case and only occurs when k_1 is also small.
2) There is initially one hole. As we move along k_2, the area of this hole decreases, until it closes. Another hole subsequently opens at the same or larger value of k_2. This hole increases as we increase $k_2 \to \infty$, like the reflexive case.
3) There is initially one hole. As we move along k_2, the area of this hole decreases. At some value of k_2, a second hole opens. The area of this second hole continues to increase as the area of the first hole decreases. At some finite value of k_2, the first hole closes, and the area of the second hole increases, as in the reflexive case.

We get the same three cases if we instead fix k_2 and move along k_1. The values of k_1 and k_2 for which holes open up are not symmetrical, however. This is illustrated in Figure 20.

With respect to the relation between the amoeba holes and the Mahler measure, in general we expect a monotonic increase in the Mahler measure if we start at a point where no holes are open and move along either k_1 or k_2. This is very similar to the reflexive case, with there only ever being at most one hole open. This hole opens at some critical value of k_1, and its area continues to increase as $k_1 \to \infty$. However, there are also instances where as the Mahler measure decreases, the area of the holes increase, or vice versa. An example is given in Figures 21 and 22 where the value of the Mahler measure decreases as k_1 increases, but when compared with the evolution of the holes of the amoeba for coefficients in the same range, we see their area increases with increasing k_1.

We have however observed that as we move along the Mahler flow, as defined for the two disconnected regions of the (k_1, k_2) plane in Section 4, we do seem to get a monotonic increase in the area of the holes. This matches the monotonic behaviour we see in the Mahler measure in these regions.

It is evident that in the case of non-reflexive polytopes, the relations between the coefficients, the amoeba holes, and the Mahler measure are

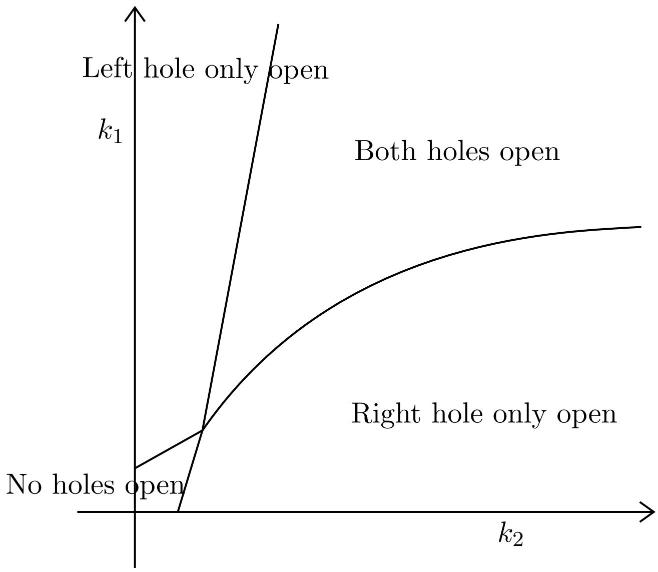


Figure 20. The number of bounded amoeba complements present with respect to the value of k_1, k_2 in the case of 2-dimensional non-reflexive polytopes with two interior points.

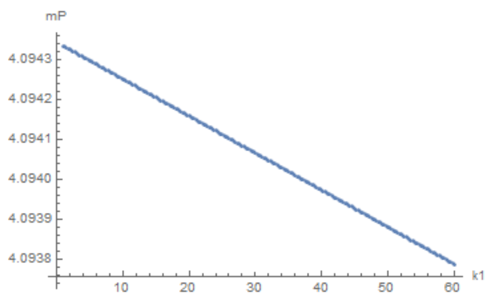


Figure 21. Mahler measure of polynomial associated with $\mathbb{C}^3/\mathbb{Z}_5$, with the origin being the left interior point. We set $k_2 = 60$ and varied k_1. We can see a clear decrease in the Mahler measure as we move along k_1

much more complicated. Nonetheless, we can still employ ML techniques, especially generic algorithm, to make their relations more precise.

5.8.1. 2d Example: $\mathbb{C}^3/\mathbb{Z}_5$. As a concrete example, we considered again the surface $\mathbb{C}^3/\mathbb{Z}_5$ whose associated toric diagram is given in Figure 23, which has two interior points and is thus non-reflexive. Taking the left interior point

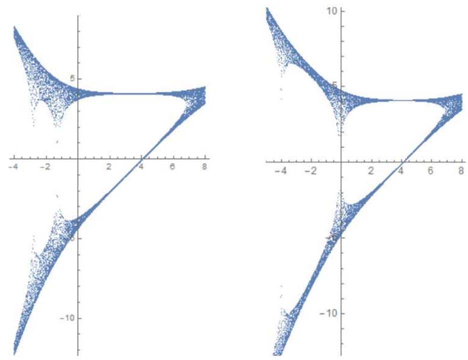


Figure 22. Amoebae of polynomial associated with $\mathbb{C}^3/\mathbb{Z}_5$. In both cases we take the left hand interior point to be the origin and set $k_2 = 60$. In the left amoeba, we set $k_1 = 10$ and in the right we set $k_1 = 55$. There is a clear increase in hole size for larger k_2.

as the origin, the Newton polynomial is

$$P(z, w) = k_1 + k_2 z + z^{-1} + zw + z^2 w^{-1}.$$

Following the same method in Section 5.1, the analytic boundary of the amoeba (Figure 24) is given by

$$(46)\quad y = \ln\left(\left|\frac{\pm k_1 e^{-x} - k_2 - e^{-2x}}{2} \pm \sqrt{\left(\frac{\pm k_1 e^{-x} - k_2 - e^{-2x}}{2}\right)^2 \pm e^x}\right|\right).$$

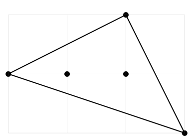


Figure 23. The toric diagram associated with $\mathbb{C}^3/\mathbb{Z}_5$.

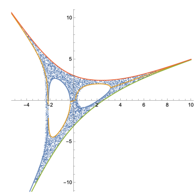


Figure 24. The boundary curves of the amoeba where $k_1 = 10$ and $k_2 = 10$.

We then explored the numerical relations between the areas of the bounded amoeba complements, the values of k_1 and k_2, and the Mahler measure using symbolic regression and NonLinearModelFit with an assumed form. Specifically, we restricted ourselves to the range of values where the two amoeba holes are both present.

We first machine-learned the relation between $A_{1,2}$ and $k_{1,2}$: From our discussion in the reflexive case and observation of the gplearn results, we expect the leading order dependence of the area on $k_{1,2}$ should be second order in the logarithm of $k_{1,2}$. Thus, we assumed the form of

$$a\ln^2 k_1 + b\ln k_1 \ln k_2 + c\ln^2 k_2 + d\ln k_1 + e\ln k_2 + h$$

to use the NonLinearModelFit function in Mathematica. The ML and fitting results are presented in Table 12.

gplearn result	Mathematica result
$A_1(k_1,k_2) = -k_1^{0.25} + 9.62(k_1/\ln^{0.5} k_2)^{0.5} - 2k_2/k_1$ with an R^2 score of 0.98054 and mean absolute error of 4.22914.	$A_1(k_1,k_2) = -20.0273 + 8.9449\ln k_1 + 2.81341\ln^2 k_1 - 4.3546\ln k_2 + 1.9363\ln k_1 \ln k_2 - 1.6381\ln^2 k_2$ with an R^2 score of 0.99997 and mean prediction error of 0.00378.
$A_2(k_1,k_2) = 2.4692\big(-(0.1086k_1\ln(k_2^{0.5}/k_1) + 7.527k_2 - 69.3161)/\ln(7.527/k_1)\big)^{0.5}$ with an R^2 score of 0.99431 and mean absolute error of 1.19120.	$A_2(k_1,k_2) = -0.7777 - 3.4303\ln k_1 + 0.9557\ln^2 k_1 + 1.9742\ln k_2 - 3.3533\ln k_1 \ln k_2 + 4.5230\ln^2 k_2$ with an R^2 score of 0.99998 and mean prediction error of 0.00170.

Table 12. Fits obtained from symbolic regression and NonLinearModelFit function

To learn the relation between $m(P)$ and $A_{1,2}$, we computed the Mahler measure associated with the amoeba with two holes present in the range of $0 \leq k_{1,2} \leq 800$. The data points are plotted in Figure 25. There seems to be a discontinuous transition in the Mahler measure as we vary the sizes of the amoeba holes, and meaningful ML results using genetic symbolic regression can only be obtained if we fit two regions (left and right) separately. The numerical relations obtained are presented in Table 13.

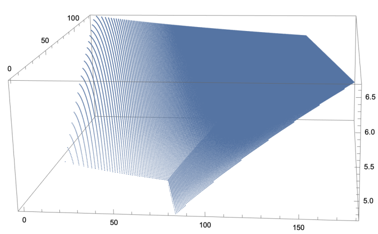


Figure 25. Plot of the Mahler measure against the areas of two amoeba holes

gplearn result	Mathematica result
$m_L(A_1, A_2) = \left((0.115314A_2 - 0.0832856A_1^{0.5})(A_1 + A_2) \right)^{0.25}$ with an R^2 score of 0.98803 and mean absolute error of 0.03493.	$m_L(A_1, A_2) = 1.4397 - 0.1465A_1^{0.25} + 0.2205A_1^{0.5} - 0.5654A_2^{0.25} - 0.1208A_1^{0.25}A_2^{0.25} + 0.6588A_2^{0.5}$ with an R^2 score of 1.00000 and mean prediction error of 0.00002.
$m_R(A_1, A_2) = 1.1745(-0.1761A_1 - 0.7250A_2^{0.5} + 1)^{0.5}$ with an R^2 score of 0.99545 and mean absolute error of 0.01709.	$m_R(A_1, A_2) = 1.0878 - 0.4489A_1^{0.25} + 0.5654A_1^{0.5} - 0.0019A_2^{0.25} - 0.1194A_1^{0.25}A_2^{0.25} + 0.1639A_2^{0.5}$ with an R^2 score of 0.99999 and mean prediction error of 0.00005.

Table 13. Fits obtained from symbolic regression and NonLinearModelFit function

Moreover, we changed the parsimony coefficient in `gplearn` which controls the complexity of the equations from 0.02 to 0.002 in order for better learning result. The ansatz used in the NonLinearModelFit function is $a + bA_1^{0.25} + cA_1^{0.5} + dA_2^{0.25} + eA_1^{0.25}A_2^{0.25} + hA_2^{0.5}$ in both regions, by inverting the conjectured relation in 2d. Specifically, the line of intersection of two surfaces is found to be

$$\begin{aligned} A_2 = &-0.4975 - 0.7157A_1^{1/4} + 1.1806A_1^{1/2} - 0.8423A_1^{3/4} + 0.4857A_1 \\ &+ (-0.0079 - 0.1660A_1^{1/4} - 0.7928A_1^{1/2} + 1.0819A_1^{3/4} + 2.7237A_1 \\ &- 4.5800A_1^{5/4} + 1.7311A_1^{3/2} + 0.0090A_1^{7/4} + 0.0000116A_1^2)^{1/2}. \end{aligned} \tag{47}$$

The presence of this line of special values resemble the plots of Mahler measure in Section 4. The fitting results are plotted together with the data points in Figure 26.

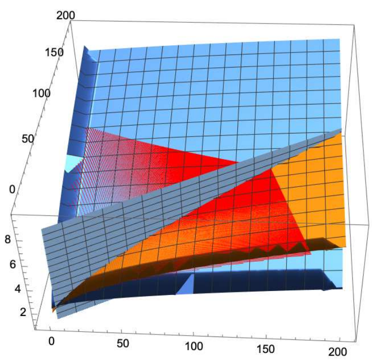


Figure 26. Plots of the two fitted surfaces (blue and orange), the plane that passes through the line of intersection (grey), and the data points (red).

The fitting results using NonLinearModelFit in Tables 12 and 13 have a rather high R^2 value close to 1. This provides further support for the adopted assumed forms based on previous `gplearn` results and conjectures, i.e. the degree of the polynomial relation equals to the dimension of the amoeba.

Given the extraordinary performance using NonLinearModelFit in `Mathematica`, one is tempted to conjecture an exact formula. Converting the numerical coefficients to potential closed form [58], an example for m_L in Table 13 is:

$$m_L(A_1, A_2) = \frac{11\pi}{24} - \frac{2\pi}{43}A_1^{1/4} + \frac{4\pi}{57}A_1^{1/2}$$

$$-\frac{9\pi}{50}A_2^{1/4} - \frac{\pi}{26}A_1^{1/4}A_2^{1/4} + \frac{56}{85}A_2^{1/2}\,. \tag{48}$$

It would be interesting to prove results such as the above.

Additionally, we would like to note that the choice of origin would affect the areas of the corresponding amoeba holes given the same toric diagram. Specifically, if we set the coefficients to be of the form $P(z, w) = k_1 - p(z, w)$, with all coefficients of $p(z, w)$ positive, the values of (k_1, k_2) for which holes open up do not seem to be related. If we however set all coefficients in $P(z, w)$ to be the same sign, amoebae are equivalent to each other i.e. $A_l(k_1, k_2) = A_r(k_2, k_1)$, where A_l is the amoeba when the left interior point is taken to be the origin. This is expected, as it is the same as multiplying the related polynomial by a factor of $z^a w^b$, while keeping all coefficients the same.

6. Discussions and outlook

In this paper, we brought together amoebae in tropical geometry and the Mahler measure in number theory, in the context of brane configurations and dimer models.

First, we continued the study of applying machine learning techniques to the analysis of amoebae topology, initiated in [8]. We applied both MLP and CNN to examples of 3-dimensional reflexive amoebae and compared the results using data obtained from persistent homology and analytic conditions using lopsidedness. Although the analytic conditions always give clearer data separation shown with the MDS projection, it may not be available for complicated examples and persistent homology can be helpful in those cases. The ML performance on data from lopsidedness only improves marginally if the size of the ML data is increased, whilst the ML performance on data from persistent homology can be improved by increasing the data size at the cost of longer computation time. Similar to the 2-dimensional results in [8], a simple MLP or CNN can predict the number of 2-dimensional cavities characterised by the second Betti number to a high accuracy.

Second, we extended the definition of the Mahler flow in [14] to incorporate the extra degrees of freedom present in non-reflexive polytopes. We investigate the properties of the flow using a combination of analytical and numerical techniques, and discuss its relation to amoebae and dimers.

Finally and most importantly, we obtained a more precise relation between the amoeba and the Mahler measure which are closely but mysteriously related through dimer models and crystal melting models. To do so, we performed symbolic regression using genetic algorithms to machine learn the numerical relations between the volume of the bounded amoeba complement, coefficient k in the Newton polynomial, and the Mahler measure, which are conjectured in [14]. We obtained the analytic expressions of the amoeba boundary by considering the poles of the gradient of the Ronkin function, which allowed computation of the volume of the bounded amoeba complement. Although the mean absolute error may be high in complicated examples such as the 3-dimensional amoebae or non-reflexive amoebae, the ML results are useful in making ansätzen required in the NonLinearModelFit function in `Mathematica` to obtain a better fit. At the end, we also considered an example 2-dimensional non-reflexive polytopes where the dynamics between the coefficients, the Mahler measure, and the areas of the amoeba holes is much more complicated. That said, we were able to find a numeric relation between these non-reflexive amoebae and the Mahler flow.

Our results from genetic symbolic regression in Section 5 provide numerical evidence for Conjecture 3.8 in [14] in both two and three dimensions. Specifically, we found that the volume of the bounded complement of the amoeba is related to the gas phase contribution to the Mahler measure by a polynomial of degree of the dimension of the amoeba.

In our discussion of the relation between Mahler measure and amoeba hole, we used the notion of gas phase contribution to the Mahler measure, but we also found that this notion needs to be refined in the case involving multiple coefficients or non-reflexive polytopes in 2-dimensional cases. In 3 dimensions, we defined an analogous notion of $m_g(k) = m(k) - m(k_c)$, where k_c is the critical value at which the 2-dimensional cavity first appears. The interpretation of this m_g would require a 3-dimensional dimer model and can be a subject of future studies. We will leave the physical interpretation of the Mahler measure in these more complicated scenarios to future work.

Our analysis also implies the power of numerical analysis in this context, and we can continue in this direction to study concepts such as the closely related Ronkin functions and its Legendre dual.

Acknowledgments. The authors would like to thank Jiakang Bao and Elli Heyes for insightful discussion. YHH would like to thank STFC for grant ST/J00037X/2. EH would like to thank STFC for a PhD studentship. The research of AZ has been supported by the French "Investissements d'Avenir" program, project ISITE-BFC (No. ANR-15-IDEX-0003), and EIPHI Graduate School (No. ANR17-EURE-0002).

Appendix A. Additional examples of ML the Betti number of 3d amoebae

A.1. $\mathbb{P}^3$

The Newton polynomial corresponding to $\mathbb{P}^3$ is $P(z_1, z_2, z_3) = c_1 z_1 + c_2 z_2 + c_3 z_3 + c_4 z_1^{-1} z_2^{-1} z_3^{-1} + c_5$, whose toric diagram is given in Figure A1, with an example Monte Carlo sampled amoebae in Figure A2.

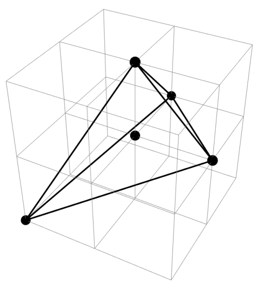


Figure A1. Toric diagram for $\mathbb{P}^3$.

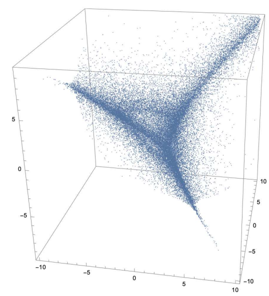


Figure A2. An example of the corresponding $\mathbb{P}^3$ amoeba from Monte Carlo sampling.

A.1.1. Learning persistent homology b_2. Using persistent homology to obtain the values of b_2 for a set of 3000 coefficient lists. The values of b_2 is determined as follows

$$b_2 = \begin{cases} 0 & \text{No persistent pairs with } q - p > 0.24\,; \\ 1 & \text{Otherwise}\,. \end{cases} \tag{A.1}$$

Then performing ML on this dataset achieves performance measures

$$\text{MLP:} \quad \text{ACC: } 0.840 \pm 0.015\,, \quad \text{MCC: } 0.699 \pm 0.022\,, \tag{A.2}$$

$$\text{CNN:} \quad \text{ACC: } 0.727 \pm 0.031\,, \quad \text{MCC: } 0.457 \pm 0.073\,. \tag{A.3}$$

A.1.2. Learning analytic lopsidedness b_2. The analytic condition for b_2 using lopsidedness is

$$b_2 = \begin{cases} 0 & |c_5| \leq |c_1 c_4|^{1/4} + |c_2 c_4|^{1/4} + |c_3 c_4|^{1/4} + |c_1 c_2 c_3|^{3/4} |c_4|^{1/4}\,; \\ 1 & \text{Otherwise}\,. \end{cases} \tag{A.4}$$

For a balanced dataset of 7000 random samples with $c_{1,2,3,4} \in [-5, 5]$ and $c_5 \in [-10, 10]$ using this analytic condition, the ML performance measures achieved over the 5-fold cross-validation were

(A.5) MLP: ACC: 0.939 ± 0.009, MCC: 0.876 ± 0.017,

(A.6) CNN: ACC: 0.910 ± 0.010, MCC: 0.819 ± 0.019.

A.2. $\mathbb{P}^2 \times \mathbb{P}^1$

The Newton polynomial associated with $\mathbb{P}^2 \times \mathbb{P}^1$ is $P(z_1, z_2, z_3) = c_1 z_1 + c_2 z_2 + c_3 z_3 + c_4 z_1^{-1} + c_5 z_2^{-1} z_3^{-1} + c_6$, and the toric diagram and example Monte Carlo amoeba are given in Figures A3 and A4. This is also a reflexive polytope with only one interior point. Thus, $b_2 = 0$ or 1.

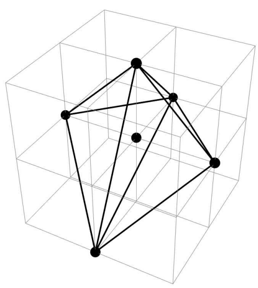


Figure A3. Toric diagram for $\mathbb{P}^2 \times \mathbb{P}^1$.

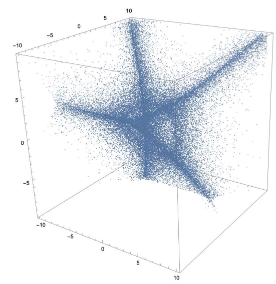


Figure A4. An example of the corresponding $\mathbb{P}^2 \times \mathbb{P}^1$ amoeba from Monte Carlo sampling.

A.2.1. Learning persistent homology b_2.

A balanced data set of 4000 random samples is used with with $c_{1,2,3,4,5} \in [-5, 5]$ and $c_6 \in [-15, 15]$. The values of b_2 were determined as follows using persistent homology

$$b_2 = \begin{cases} 0 & \text{No persistent pairs with } q - p > 0.28\,; \\ 1 & \text{Otherwise}\,, \end{cases} \tag{A.7}$$

leading to ML results

(A.8) MLP: ACC: 0.830 ± 0.016, MCC: 0.652 ± 0.035,

(A.9) CNN: ACC: 0.825 ± 0.027, MCC: 0.630 ± 0.063.

A.2.2. Learning analytic lopsidedness b_2. Following the same derivation methods, the b_2 values determined using lopsidedness used

$$b_2 = \begin{cases} 0 & |c_6| \leq 2|c_1c_4|^{1/2} + 3|c_2c_3c_5|^{1/3}\,; \\ 1 & \text{Otherwise}\,, \end{cases} \tag{A.10}$$

equivalently leading to ML results

$$\text{MLP:} \quad \text{ACC: } 0.947 \pm 0.007\,, \quad \text{MCC: } 0.893 \pm 0.014\,, \tag{A.11}$$

$$\text{CNN:} \quad \text{ACC: } 0.920 \pm 0.011\,, \quad \text{MCC: } 0.841 \pm 0.023\,. \tag{A.12}$$

Appendix B. Additional example of ML 2d amoebae and Mahler measure

B.1. $\mathbb{P}^2$

The Newton Polynomial in this case is $P(z, w) = k - z - w - z^{-1}w^{-1}$. The analytic boundary of the amoeba is

$$x = \ln\left(\left|\frac{k}{2} \pm \frac{e^y}{2} \pm \sqrt{\frac{1}{4}\left(k \pm e^y\right)^2 \pm e^{-y}}\right|\right)\,, \tag{B.13}$$

for $k \geq 3$ (Figure B1).

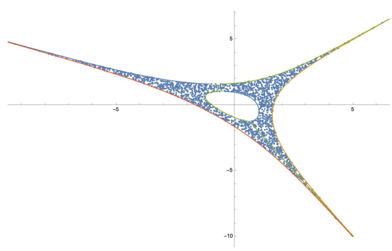


Figure B1. The analytic boundary of amoeba with $k = 4$.

The Mahler measure for $P(z, w) = k - z - w - z^{-1}w^{-1}$ as a function of k obtained using Taylor expansion and Cauchy residue theorem is

$$m(P) = \ln k - 2k^{-3}{}_4F_3\left(1, 1, \frac{4}{3}, \frac{5}{3}; 2, 2, 2; 27k^{-3}\right)\,, \tag{B.14}$$

The relation between the gas phase contribution $m_g(P)$ and the amoeba hole area is fitted with 20000 data pairs and found to be

$$A_h = 4.59038 m_g^2 + 7.24861 m_g + 5.09803\,, \tag{B.15}$$

with an R^2 score of 1.0000 and mean absolute error of 0.1340 (plotted in Figure B2).

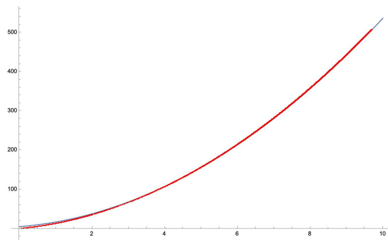


Figure B2. Data points (red) and the fitted relation (blue) between A_h and m_g.

The relation between the amoeba hole area and the value of k for $k \geq 3$ is fitted with 5000 data pairs, and is found to be

$$A_h = 2\ln^2 k + \ln(k \ln k) - 5.49 + \ln k^2 \ln\big((k - \ln k^2)(\ln(k \ln k) - 5.49)\big)\,,$$

with an R^2 score of 0.9998 and mean absolute error of 1.0983 (plotted in Figure B3).

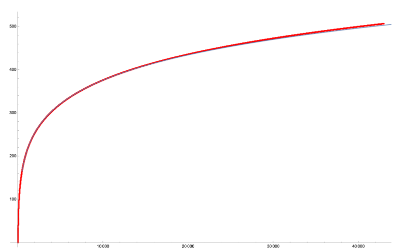


Figure B3. Data points (red) and the fitted relation (blue) between A_h and k.

Appendix C. Explicit example of the expansion method

In this section we outline the method used to calculate expressions for the Mahler measure, and present an explicit example. This method can be used to calculate the Mahler measure of any polynomial. In this example, we will derive Eq.(29), which corresponds to the $\mathbb{C}^3/\mathbb{Z}_5$ polygon with s as the origin:

$$m(P_s(z,w)) = m\left(k_1 - k_2 z - \frac{1}{z} - zw - \frac{z^2}{w}\right). \tag{C.16}$$

In order to expand, we write $P_s(z,w) = k_1 - p_s(z,w)$, where $p_s(z,w) = k_2 z + z^{-1} + zw + z^2 w^{-1}$. From Eq.(25), we expand as:

$$m(P_s(z,w)) = \log k_1 - \sum_{n=1}^{\infty} \frac{[p_s^n(z,w)]_0}{n k_1^n}, \tag{C.17}$$

where $[p_s^n(z,w)]_0$ is the constant term of the n^{th} power of $p_s(z,w)$. To calculate these constant terms, we use a binomial expansion as follows:

$$p_s^n(z,w) = \left(k_2 z + z^{-1} + zw + z^2 w^{-1}\right)^n \tag{C.18}$$

$$= \sum_{i=0}^{n} \binom{n}{i} (k_2 z + z^{-1})^i (zw + z^2 w^{-1})^{n-i} \tag{C.19}$$

$$= \sum_{i=0}^{n} \binom{n}{i} \left[\sum_{l=0}^{i} \binom{i}{l} k_2^l z^l z^{l-i}\right] \times \left[\sum_{j=0}^{n-i} \binom{n-i}{j} z^{2(n-i-j)} w^{-(n-i-j)} z^j w^j\right]. \tag{C.20}$$

We are looking for constant terms only, so the sum of the powers of both z and w should be equal to zero. Grouping w and z terms individually, we get:

$$2j + i - n = 0 \Rightarrow j = \frac{n-i}{2}, \tag{C.21}$$

$$2l + 2n - 3i - j = 0 \Rightarrow l = \frac{5i-3n}{4}. \tag{C.22}$$

Subbing this into Eq.(C.20), we arrive at

$$[p_s^n(z,w)]_0 = \sum_{i=0}^{n} \binom{n}{i} \binom{n-i}{\frac{n-i}{2}} \binom{i}{\frac{5i-3n}{4}} k_2^{\frac{5i-3n}{4}} . \tag{C.23}$$

Finally, inserting this into Eq.(C.17), we arrive at our final result:

$$m\left(P_s(z,w)\right) = \log k_1 - \sum_{n=1}^{\infty} \sum_{i=0}^{n} \binom{n}{i} \binom{n-i}{\frac{n-i}{2}} \binom{i}{\frac{5i-3n}{4}} \frac{k_2^{\left(\frac{5i-3n}{4}\right)}}{k_1^n n}, \tag{C.24}$$

which is valid for all $k_1 \geq \max_{|z|,|w|=1} |p_s(z,w)|$. This expression comes with some constraints, which ensures all entries in the binomials are positive integers, and for $\binom{n}{r}$, we always have $n \geq r$. We require that $i \geq 3n/5$ and that $(3i - 5n) \bmod 4 = 0$. This can greatly decrease the number of terms in the series.

This method can be used to calculate the Mahler measure of any polynomial. In cases where the number of terms in the polynomial becomes large, we may have to sum over a large number of indices. In general, the number of indices we sum over is equal to (excluding n): Number of indices summed over = ((Number of non-constant terms in the polynomial)−1)−(Number of variables). Because of this, for polynomials with a large number of variables, this expansion method is often much more efficient than numerical integration method, where we would have to integrate over each variable.

References

[1] M. Forsberg, M. Passare, A. Tsikh, *Laurent determinants and arrangements of hyperplane amoebas*, Advances in Mathematics **151** (2000), 45–70.

[2] G. Arias-Tamargo, Y. He, E. Heyes, E. Hirst, and D. Rodriguez-Gomez, *Brain webs for brane webs*, *Phys. Lett. B* **833** 2022 137376.

[3] Jiakang Bao, *A note on quiver Yangians and R-matrices*, JHEP **08** (2022) 219.

[4] Jiakang Bao, *Quiver Yangians and $\mathcal{W}$-Algebras for Generalized Conifolds*, J. Phys. A **56** (2023) 225203.

[5] Jiakang Bao, Grace Beaney Colverd, and Yang-Hui He, *Quiver Gauge Theories: Beyond Reflexivity*, JHEP **06** (2020) 161.

[6] Jiakang Bao, Sebastián Franco, Yang-Hui He, Edward Hirst, Gregg Musiker, and Yan Xiao, *Quiver Mutations, Seiberg Duality and Machine Learning*, Phys. Rev. D **102** (2020) 086013.

[7] Jiakang Bao, Yang-Hui He, Elli Heyes, and Edward Hirst, *Machine Learning Algebraic Geometry for Physics*, `arXiv:2204.10334`.

[8] Jiakang Bao, Yang-Hui He, and Edward Hirst, *Neurons on Amoebae*, J. Symb. Comput. **116** (2022), 1–38.

[9] Jiakang Bao, Yang-Hui He, Edward Hirst, Johannes Hofscheier, Alexander Kasprzyk, and Suvajit Majumder, *Hilbert Series, Machine Learning, and Applications to Physics*, Phys. Lett. B **827** (2022) 136966.

[10] Jiakang Bao, Yang-Hui He, Edward Hirst, Johannes Hofscheier, Alexander Kasprzyk, and Suvajit Majumder, *Polytopes and Machine Learning*, `arXiv:2109.09602`.

[11] Jiakang Bao, Yang-Hui He, Edward Hirst, and Stephen Pietromonaco. Lectures on the Calabi-Yau Landscape, `arXiv.2001.01212`.

[12] Jiakang Bao, Yang-Hui He, and Ali Zahabi, *Reflexions on Mahler: Dessins, Modularity and Gauge Theories*, `arXiv:2111.03655`.

[13] Jiakang Bao, Yang-Hui He, and Ali Zahabi, *Crystal melting, BPS quivers and plethystics*, JHEP **06** (2022) 016.

[14] Jiakang Bao, Yang-Hui He, and Ali Zahabi, *Mahler Measure for a Quiver Symphony*, Commun. Math. Phys. **394** (2022), 573–624.

[15] Benjamin Bengfort and Rebecca Bilbro, *Yellowbrick: Visualizing the Scikit-Learn Model Selection Process*, J. Open Source Softw. **4** (2019) 1075.

[16] Per Berglund, Giorgi Butbaia, Tristan Hübsch, Vishnu Jejjala, Damián Mayorga Peña, Challenger Mishra, and Justin Tan, *Machine Learned Calabi–Yau Metrics and Curvature*, `arXiv:2211.09801`.

[17] David S. Berman, Yang-Hui He, and Edward Hirst, *Machine learning Calabi-Yau hypersurfaces*, Phys. Rev. D **105** (2022) 066002.

[18] Milo Bogaard, *Introduction to amoebas and tropical geometry*, Masters thesis, U. Amsterdam 2015.

[19] Agostino Butti and Alberto Zaffaroni, *R-charges from toric diagrams and the equivalence of a-maximization and Z-minimization*, JHEP **11** (2005) 019.

[20] Jonathan Carifio, James Halverson, Dmitri Krioukov, and Brent D. Nelson, *Machine Learning in the String Landscape*, JHEP **09** (2017) 157.

[21] Heng-Yu Chen, Yang-Hui He, Shailesh Lal, and Suvajit Majumder, *Machine learning Lie structures & applications to physics*, Phys. Lett. B **817** (2021) 136297.

[22] Heng-Yu Chen, Yang-Hui He, Shailesh Lal, and M. Zaid Zaz, *Machine Learning Etudes in Conformal Field Theories*, arXiv:2006.16114.

[23] Bo Feng, Yang-Hui He, Kristian D. Kennaway, and Cumrun Vafa, *Dimer models from mirror symmetry and quivering amoebae*, Adv. Theor. Math. Phys. **12** (2008), 489–545.

[24] Sebastian Franco, Amihay Hanany, Kristian D. Kennaway, David Vegh, and Brian Wecht, *Brane dimers and quiver gauge theories*, JHEP **01** (2006) 096.

[25] Sebastian Franco, Amihay Hanany, Dario Martelli, James Sparks, David Vegh, and Brian Wecht, *Gauge theories from toric geometry and brane tilings*, JHEP **01** (2006) 128.

[26] Sebastian Franco, Sangmin Lee, and Rak-Kyeong Seong, *Brane Brick Models, Toric Calabi-Yau 4-Folds and 2d (0,2) Quivers*, JHEP **02** (2016) 047.

[27] Yarin Gal, Vishnu Jejjala, Damián Kaloni Mayorga Peña, and Challenger Mishra, *Baryons from Mesons: A Machine Learning Perspective*, Int. J. Mod. Phys. A **37** (2022) 2250031.

[28] Xin Gao and Hao Zou, *Applying machine learning to the Calabi-Yau orientifolds with string vacua*, Phys. Rev. D **105** (2022) 046017.

[29] Amihay Hanany and Kristian D. Kennaway, *Dimer models and toric diagrams*, arXiv:0503149.

[30] Yang-Hui He, *Deep-Learning the Landscape*, Phys. Lett. B **774** (2017), 564–568.

[31] Yang-Hui He, The Calabi-Yau Landscape: From Geometry, to Physics, to Machine Learning, Lecture Notes in Mathematics, 2021.

[32] Yang-Hui He, *From the String Landscape to the Mathematical Landscape: a Machine-Learning Outlook*, In 14th International Workshop on Lie Theory and Its Applications in Physics, 2022.

[33] Yang-Hui He, Edward Hirst, and Toby Peterken, *Machine-learning dessins d'enfants: explorations via modular and Seiberg-Witten curves*, J. Phys. A **54** (2021) 075401.

[34] Yang-Hui He and Juan Manuel Pérez Ipiña, *Machine-learning the classification of spacetimes*, Phys. Lett. B **832** (2022) 137213.

[35] Yang-Hui He and Shing-Tung Yau, *Graph Laplacians, Riemannian Manifolds and their Machine-Learning*, `arXiv:2006.16619v1`.

[36] Kentaro Hori, Amer Iqbal, and Cumrun Vafa, *D-branes and mirror symmetry*, `arXiv:hep-th/0005247`.

[37] Wolfram Research, Inc. Mathematica, version 13.1. Champaign, IL, 2022.

[38] Richard Kenyon, *Dimer problems*, Encyclopedia of Mathematical Physics, ed. J.-P. Françoise, GL Naber, TS Tsun, Academic Press, Amsterdam, 2006.

[39] Richard Kenyon, *Lectures on dimers*, `arXiv:0910.3129`.

[40] Richard Kenyon and Andrei Okounkov, *Planar dimers and harnack curves*, `arXiv:math/0311062`.

[41] Richard Kenyon, Scott Sheffield, and Andrei Okounkov, *Dimers and amoebae*, Annals of mathematics **163** (2003), 1019–1056.

[42] Daniel Krefl and Rak-Kyeong Seong, *Machine Learning of Calabi-Yau Volumes*, Phys. Rev. D **96** (2017) 066014.

[43] Sven Krippendorf, Dieter Lust, and Marc Syvaeri, *Integrability Ex Machina*, Fortsch. Phys. **69** (2021) 2100057.

[44] Magdalena Larfors, Andre Lukas, Fabian Ruehle, and Robin Schneider, *Learning Size and Shape of Calabi-Yau Spaces*, `arXiv:2111.01436`.

[45] D. Maclagan and B. Sturmfels, Introduction to Tropical Geometry, Graduate Studies in Mathematics, American Mathematical Society, 2015.

[46] Takashi Maeda and Toshio Nakatsu, *Amoebas and instantons*, Int. J. Mod. Phys. A **22** (2007) 937–984.

[47] Kurt Mahler, *On some inequalities for polynomials in several variables*, Journal of The London Mathematical Society-second Series (1962), 341–344.

[48] Grigory Mikhalkin, *Amoebas of algebraic varieties and tropical geometry*, `arXiv:0403015`.

[49] Andrei Okounkov, Nikolai Reshetikhin, and Cumrun Vafa, *Quantum Calabi-Yau and classical crystals*, Prog. Math. (2006), 244–597.

[50] Hirosi Ooguri and Masahito Yamazaki, *Crystal Melting and Toric Calabi-Yau Manifolds*, Commun. Math. Phys. **292** (2009), 179–199.

[51] Hirosi Ooguri and Masahito Yamazaki, *Emergent Calabi-Yau Geometry*, Phys. Rev. Lett. **102** (2009) 161601.

[52] M. Passare and H. Rullgård, *Amoebas, monge-ampère measures, and triangulations of the newton polytope*, Duke Mathematical Journal **121** (2004), 481–507.

[53] Kevin Purbhoo, *A nullstellensatz for amoebas*, `arXiv:math/0603201`.

[54] Fabian Ruehle, *Evolving neural networks with genetic algorithms to study the String Landscape*, JHEP **08** (2017) 038.

[55] Chris Smyth, The Mahler measure of algebraic numbers: a survey, London Mathematical Society Lecture Note Series, Cambridge University Press, 2008.

[56] Christopher Tralie, Nathaniel Saul, and Rann Bar-On, *Ripser.py: A lean persistent homology library for python*, The Journal of Open Source Software **3** (2018) 9.

[57] F. Rodriguez Villegas, Modular Mahler Measures I, Springer US, Boston, MA, 1999.

[58] Wolfram—Alpha, `https://www.wolframalpha.com`.

[59] Masahito Yamazaki, *Geometry and Combinatorics of Crystal Melting*, RIMS Kokyuroku Bessatsu B **28** (2011) 193.

[60] Ali Zahabi, *Toric Quiver Asymptotics and Mahler Measure:* $\mathcal{N} = 2$ *BPS States*, JHEP **07** (2019) 121.

[61] Ali Zahabi, *Quiver asymptotics and amoeba: Instantons on toric Calabi-Yau divisors*, Phys. Rev. D **103** (2021) 086024.

Balliol College, University of Oxford
OX1 3BJ, UK
London Institute for Mathematical Sciences, Royal Institution
London W1S 4BS, UK
E-mail address: chensiqi1024@outlook.com

London Institute for Mathematical Sciences, Royal Institution
London W1S 4BS, UK
Department of Mathematics, City, University of London
EC1V 0HB, UK
Merton College, University of Oxford
OX1 4JD, UK
School of Physics, NanKai University
Tianjin, 300071, P.R. China
E-mail address: hey@maths.ox.ac.uk

London Institute for Mathematical Sciences, Royal Institution
London W1S 4BS, UK
Department of Mathematics, City, University of London
EC1V 0HB, UK
E-mail address: edward.hirst@city.ac.uk

London Institute for Mathematical Sciences, Royal Institution
London W1S 4BS, UK
Hertford College, University of Oxford
OX1 3BW, UK
E-mail address: andrewnestor123@gmail.com

London Institute for Mathematical Sciences, Royal Institution
London W1S 4BS, UK
Institut de Mathématiques de Bourgogne, Université Bourgogne Franche-Comté
BP 47870, France
E-mail address: zahabi.ali@gmail.com