Testing Swampland Conjectures with Machine Learning

Nana Cabo Bizet

Physics Department,
Division of Sciences and Engineering,
University of Guanajuato, Mexico.

string_data 2020, CERN.

December 15th 2020

Collaborators: Cesar Damian, Oscar Loaiza-Brito, Damián K. Mayorga Peña, J. A. Montañez-Barrera,

The European Physical Journal C Vol. 80 766 (2020) [arXiv:2006.07290].

Collaborators

Cesar Damian
University of Guanajuato

Oscar Loaiza-Brito
University of Guanajuato

Damián Kaloni Mayorga Peña
University of the Witwatersrand

J.A. Montañez-Barrera
University of Guanajuato

1. Introduction
2. Non geometric flux compactification
3. Genetic algorithms
4. Artificial Neural Network
5. Non-hierarchical fluxes
6. Hierarchical fluxes
7. Results
8. Conclusions

String theory constitutes a fertile candidate for a unified description of fundamental interactions. The compactification of the extra dimensions in CY manifolds with fluxes leads to a large landscape of vacua: $\mathcal{N} = 1$ 4D theories with potentially all moduli stabilized.

An estimate of the flux vacua is given by L^K, where L is the number of fluxes, and K is the number of cycles of the CY compactification. [Bousso, Polchinski,00].

The vacua of type IIB string theory flux compactifications on CY are finite [Douglas,02][Ashok, Douglas,03]. The finiteness comes from tadpole cancellation which bounds the fluxes, and the constrains from SUSY and dualities. The number of vacua is estimated to be of order $\mathcal{O}(10^{500})$ [Denef, Douglas,04].

In F-Theory (non perturbative type IIB strings) there is an estimate of the number of vacua to be $\mathcal{O}(10^{272000})$ [Taylor, Wang,15].

Landscape in the swampland

There are characteristics of vacua that can be studied via an statistical analysis: the cosmological constant, the SUSY breaking scale, the Standard Model like effective theories, etc.

How big is this Landscape in the bigger space of semi-classical descriptions of gravity which do not have a UV completion? [Vafa,05]

The Swampland is the set of all the theories apparently consistent semi-classically, not coming from a quantum gravity theory.

The Landscape constitutes islands in the Swampland. It is the set of all effective theories consistent with an UV completion of gravity.

The latter yields restrictions on consistent effective field theories, some of them are conjectural, and summarized as the Swampland conjectures. [Hamed, Motl, Nicolis, Vafa,06][Obied, Oogurri, Spodyneiko, Vafa, 18][Agrawal, Obied, Steinhardt, Vafa, 18] [Garg, Krishna, 19][Ooguri, Palti, Shiu, Vafa, 19],[Danielsson, Van Riet,18], [Danielsson,19]...

Introduction

Machine learning allows to explore large amounts of data for specific patterns. It provides more exhaustive checks than standard programing in shorter times. Artificial neural networks (ANNs) algorithms, inspired in the biological learning process have been used in several fields.

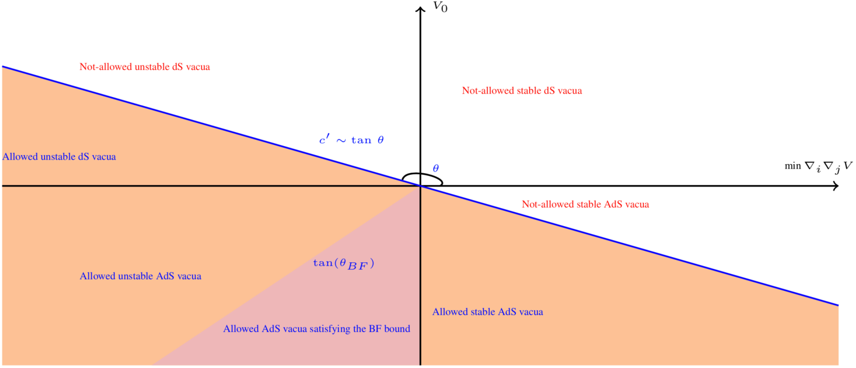
The use of machine learning techniques to solve classification problems with supervised learning applied to high energy physics has seen many developments in recent years [Ruehle,17][Carifio, Halverson, Krioukov,Nelson,17][He,18[Erbin,Krippendorf,18][Mutter,Parr,vaudrevange,19] [Halverson,Nelson,Ruehle,19] [Ashmore,He,Ovrut,19][Parr,Vaudrevange,20] [Gal,Jejjala,Mayorga-Peña,20]...

Genetic algorithms(GA) for function extremization have also been efficiently employed in the search for string theory vacua [Damian, Loaiza-Brito,13],[Damian, Díaz-Barrón,Loaiza-Brito, Sabido,13], [Abel,Rizos,14], [Ruehle,17], [Cole, Schachner, Shiu,19], [Ruehle,20], [AbdusSalam, Cicoli, Quevedo, Shukla,20]...

Introduction

We consider Type IIB string theory compactification on an isotropic torus with geometric and non geometric fluxes [Shelton, Taylor, Wecht,05] [Aldazabal, Camara, Font, Ibañez,06] [Blaback, Danielsson, Dibitetto,13] [Blumenhagen, Deser, Plauschinn, Rennecke, Schmid,13] [Blumenhagen, Gao, Herschmann, Shukla,13][Blaback, Danielsson, Dibitetto,Vargas,15] [Cribiori,Kallosh,Linde,19] ...
Employing supervised machine learning, consisting of a ANN coupled to a GA, we determine more than sixty thousand flux configurations yielding a scalar potential with at least one critical point. Stable AdS vacua with large moduli masses and small vacuum energy as well as unstable dS vacua with small tachyonic mass and large energy are absent, in accordance to the Refined de Sitter Conjecture.
Hierarchical fluxes [Damian,Loaiza,19] [Betzler,Plauschinn,19] [Cabo Bizet, Damian,Loaiza-Brito,Mayorga Peña,19]... favor perturbative solutions with small values of the vacuum energy and moduli masses, as well as scenarios with the lightest modulus mass much smaller than the AdS vacuum scale.

Employing ANN+GA we explore: the agreement of effective models with the refined dS conjecture, the 6 different zones. θ defines the slope of the line dividing the regions of the Swampland from the Landscape. In pink: AdS vacua satisfying the BF bound defined by $tan(\theta_{BF}) = -2/3$.



Non geometric flux compactification

For type IIB string theory compactification with non geometric fluxes on an isotropic T^6 torus the effective 4D $\mathcal{N}=1$ theory can be obtained in terms of a superpotential(W) and a Kähler potential (K):

$$W = P_1(U) - iSP_2(U) + iTP_3(U), \quad (1)$$

with U, S and T being the complex structure, axio-dilaton and Kähler moduli. $P_i(U), i = 1,2,3$ are polynomial and their coefficients are given in terms of NS-NS h_j, R-R f_i and non geometric b_k fluxes:

$$\begin{aligned} P_1 &= f_1 + 3If_2U - 3f_3U^2 - If_4U^3, \quad (2)\\ P_2 &= h_1 + 3Ih_2U - 3h_3U^2 - Ih_4U^3, \\ P_3 &= 3b_1 + I(2b_2 + b_3)U - (2b_4 + b_5)U^2 - Ib_6U^3. \end{aligned}$$

h_j and b_k determine the relevance of S, T to the scalar potential respectively, suggesting that the hierarchy between moduli masses can be reached by implementing hierarchies among the fluxes.

Non geometric flux compactification

The Kähler potential reads

$$K = -\ln(S + S^*) - 3\ln(U + U^*) - 3\ln(T + T^*). \qquad (3)$$

We decompose the scalar fields in terms of their real and imaginary components as: $U = u + iv$, $S = s + ic$ and $T = t + i\tau$, where u, v, s, c, t, τ are real.
The corresponding scalar potential can be computed as

$$V = e^K \left(|D_I W|^2 K^{IJ} - 3|W|^2 \right). \qquad (4)$$

The potential has extremal values when SUSY is preserved for AdS or Minkowski vacua. Non-geometric fluxes stabilize the Kähler moduli T. There are constraints from the tadpoles and Bianchi Identities. The validity of the perturbative approximation implies $s = 1/g_s > 1$. Searching for SUSY vacua, it follows that:

$$D_U W = D_S W = D_T W = 0. \qquad (5)$$

To search for minima of the scalar potential the method of differential evolution is employed [Storn, Price,97]. In our work it has been implemented in Mathematica.

For n variables, the algorithm starts with a population set of m points $x_j, j = 1...m$, with n components and $m \gg n$. In each iteration 3 random points $u, v, w \neq j$ are selected, and $x_s = x_w + s(x_u - x_w)$, $s \in \mathbb{R}, 0 < s < 2$ is constructed. The point j is updated to x_j^{new}, which is x_j with probability ρ and x_s with probability $1 - \rho$.

Using a cost function f, if $f(x_j^{new}) < f(x_j)$ then the point j is replaced: $x_j \to x_j^{new}$.

Convergence: Is measured by the distance between the new best point and old best point with a cutoff, and by the difference $f(x_j^{new}) - f(x_j)$ been also smaller than a cutoff.

Genetic algorithms

This is a global optimization method with a fast convergence, which has been employed already in the search of string theory vacua.

In the current problem for a fixed flux configuration:

- There are 6 variables.
- The set of points is constructed by the real field values of the moduli:

$$x_j = (u_j, v_j, s_j, c_j, t_j, \tau_j),\ j = 1...m,\ m = 16.$$

- The cost function is the scalar potential $f(x_j) = V(u_j, v_j, s_j, c_j, t_j, \tau_j)$.
- The number of iterations is bounded by 250.

Artificial Neural Network

There are layers denoted by l. We consider a single internal layer.

Each neuron is connected with all the neurons in the next layer through a weight factor w_{jk}. The response of the ANN a_j^l at level l is related with level $l-1$ via the weighted connection

$$a_j^l = \sigma\left(\sum_k w_{jk}^l a_k^{l-1} + b_j^l\right) , \tag{6}$$

the sum is over all neurons k in the $l-1$ layer, b^l is named the bias factor.

Consider $l = 1 \ldots n$, a_j^l represents a neuron j at level l, a_j^0 represents the input data of the network with j the number of elements, a_j^n represents the output of the network with j the number of classifications. In our case for $l = n$ we have $j = 1$.

The function $\sigma(x) = \frac{1}{1+\exp(y)}$ is the activation function which is a hyperbolic sigmoid function.

In the level zero we fix the fluxes as the input:

$$a_j^0 = F_j/|F|, j = 1...10, \quad (7)$$
$$a^0 = (F_1, F_2, F_3 ... F_{10}).$$

In the level 1 we obtain:

$$a_j^1 = \sigma\left(\sum_{k=1}^{10} w_{jk}^1 a_k^0 + b_j^l\right), j = 1...N, \quad (9)$$

where the number of neurons runs from 1 to N. In the last layer there is a single output, because we are testing a single parameter of classification:

$$a_j^2 = \sigma\left(\sum_{k=1}^{N} w_{jk}^2 a_k^1 + b_j^l\right), j = 1. \quad (10)$$

Here l takes the values $l = 0, 1, 2$. We explore from $N = 1....35$ number of neurons in layer $l = 1$.

10 networks are trained for every number of neurons N.

The training set is composed of sets of 10 independent fluxes $(a_j^0)_i$, $i = 1...K$ and a $t_i = 1$ for a positive outcome and $t_i = 0$ for a negative outcome. i.e. the target value for a set of fluxes is given by: 0 or 1.

The weights are selected in random way and updated in every step. The distance between the output a_i^n and the target value t_i for all the trained cases is measured.

The weights and bias factors are determined in such a way that the mean squared error(MSE) defined as

$$\mathrm{MSE} = \frac{1}{No.cases} \sum_{i=1}^{No.cases} \left(a_1^2(i) - t_i\right)^2 \,, \tag{11}$$

is diminished, where $a_1^2(i)$ is the response of the ANN to the entrance $a_j^0(i)$ and the t_i is the target value.

Levenberg-Marquardt optimization is employed, which is a deterministic algorithm for non-linear systems that is able to find local minima iteratively.

Artificial Neural Network

The ANN was implemented employing Matlab resources.

In our case: From the set of fluxes, 80% is used for training, 10% for validation test and 10% for a posterior test.

Taking around 40000 configurations for training, with two kinds of data: (A) Random and (B) Hierarchical fluxes.

Two criteria: (I) A stable critical point, which can be AdS or dS (No tachyon). (II) A dS critical point, which can be maximum, minimum or saddle ($\Lambda > 0$).

The input layer consists of 10 neurons, the hidden layer consists of 12 neurons(optimized) for the criterium I and 23(optimized) for the criterium II, and the output layer consists of 1 neuron.

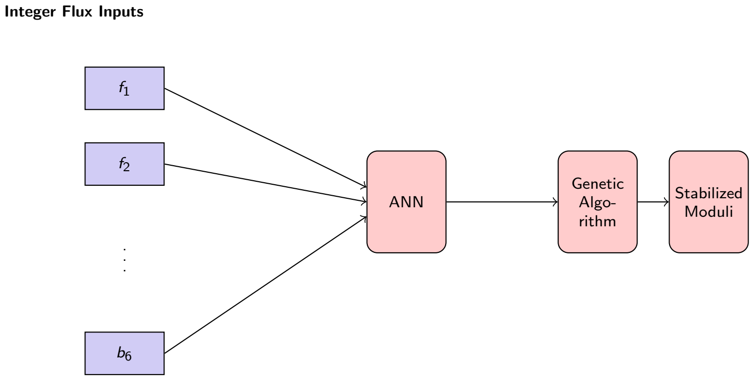
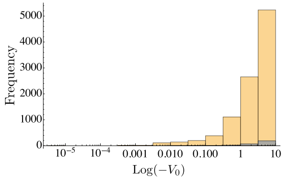


Figure:
Flow chart of the vacua search procedure. A flux configuration is the input for the neural network. The outcome is whether or not the fluxes under consideration lead to a scalar potential in the effective theory with critical points. If the outcome is positive, then one employs the Genetic Algorithm in order to find the critical point(s).

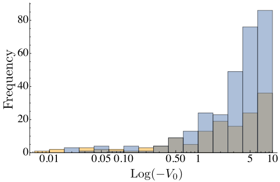
I) Stable critical points. From 10^6 cases the network selects 66000, the GA confirms 20779 critical points, 9872 with no tachyons. To compare: originally from 40000 cases the genetic algorithm found 298 stable points.

II) dS critical points. From 10^6 cases the network selects 55000, the GA confirms 4944 critical points, 140 AdS with no tachyons and 2744 dS.

Distributions of the stable vacua with random fluxes generated by the GA (blue bars) and by the ANN+GA (yellow bars). Intersection appears as gray bars. (a): I. (b): II.



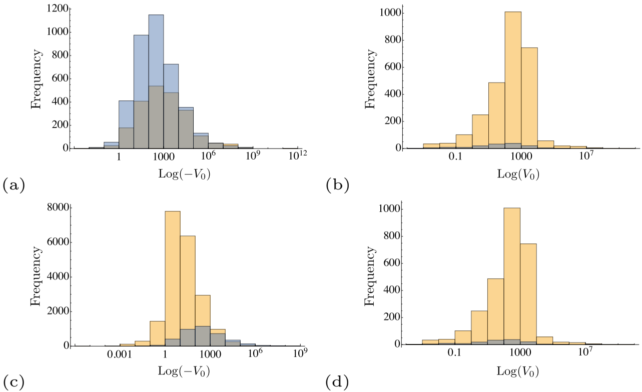
(a)



(b)

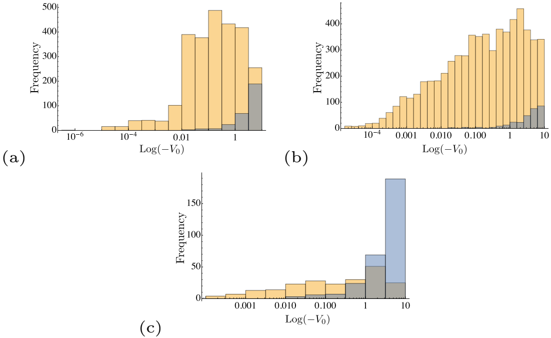
ANN for non-hierarchical fluxes

Distributions of critical points with random fluxes obtained by the ANN+GA (yellow bars) and those found by the GA (blue bars). Intersection appear in grey. (a) AdS vacua, II, (b) dS vacua, II, (c) AdS vacua, I, (d) dS vacua, I. a) differs from the others in that the amount of solutions found by the GA and the ANN+GA are comparable. Cases with at least one tachyon are included.



ANN for hierarchical fluxes

Criterium I: Distributions of free tachyon vacua for hierarchical fluxes: (a) Kähler (T), (b) complex structure (U) and (c) axio-dilaton(S) as the lightest modulus. The critical points obtained by the ANN+GA are given in yellow bars, and those found by the GA are given in blue bars.

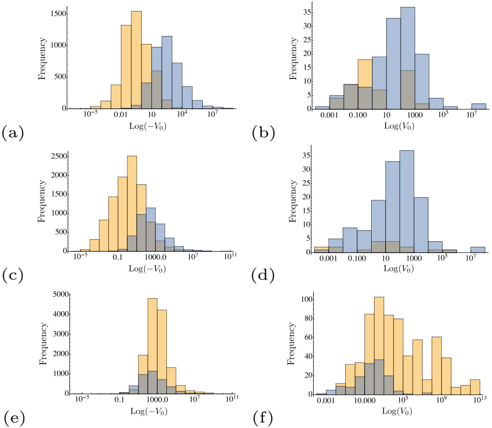


(a) (b) (c)

ANN for hierarchical fluxes

Criterium II: Distributions of unstable critical points for hierarchical fluxes.

(a) AdS, case T, (b) dS, case T, (c) AdS, case S, (d) dS, case S, (e) AdS, case U, (f) dS, case U. All dS vacua are unstable. The critical points obtained by the ANN+GA are given in yellow, and those found by the GA are given in blue. These histograms include vacua with at least one tachyon.



Results

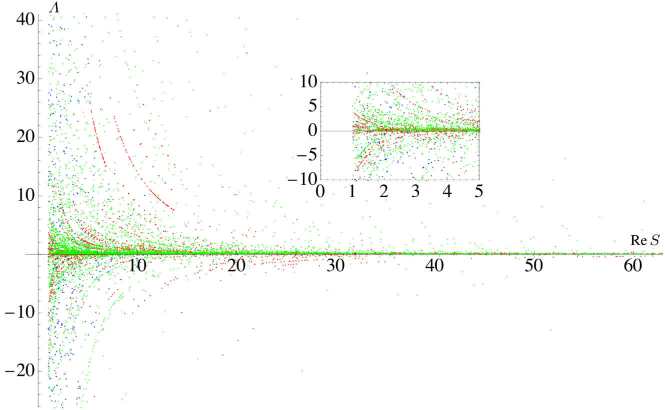
- Perturbative regime $g_s \ll 1$ is associated to a small minima values of the scalar potential $|\Lambda| \ll 1$, suggesting a correspondence $\Lambda = \pm \exp(-Re(S))$.

- There is compatibility with the refined dS conjecture. All the vacua explored in the plane Λ vs. $\min(m^2)$ are in the region $V_{min} \leq -\min(m^2)/c' + c''$.

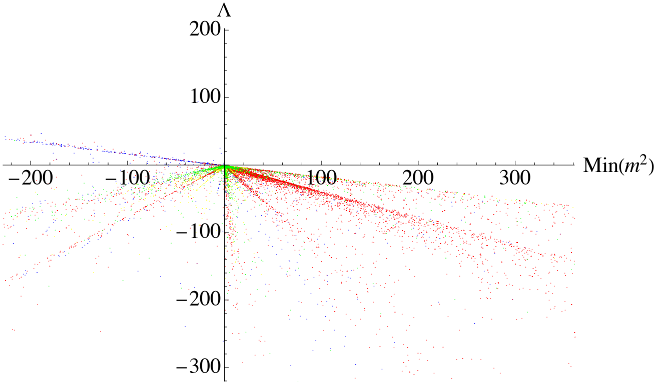
- There is an AdS scale separation: $\min(m^2) L^2_{AdS} \leq c$, with c of order 1.

Results

Perturbative regime and small cosmological constant: Λ vs. g_s at the critical point for cases produced by the ANN+GA. Red and Blue points correspond to vacua in case A, while yellow and green dots are related to case B. The smaller the string coupling the smaller the value for the cosmological constant Λ.

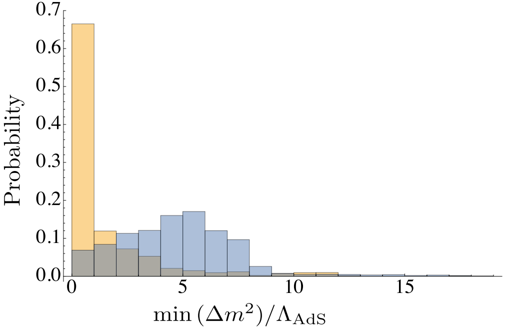


Redefined dS conjecture: Distribution of the extrema for the scalar potential. Red and blue points represent critical points of case A whereas green and yellow points correspond to critical points of case B.



Results

AdS scale separation: Histogram of the scale separation between the lightest modulus and the corresponding value of the cosmological constant in Planck units $\min(m^2/\Lambda_{\mathrm{AdS}})$. Yellow bars correspond to hierarchical fluxes while blue bars correspond to non-hierarchical fluxes.



Conclusions

We have implemented a vacuum search through an ANN coupled to a GA, reporting more than 60.000 flux configurations yielding to a scalar potential with at least one critical point.

We tested probabilistically some of our model's features in the light of recent Swampland conjectures.

Generic flux configurations produce vacua with two features:

- The Refined dS Conjecture is fulfilled and the relation $\min \nabla_i \nabla_j V \leq -c' V$ with c' of order 1 is graphically proved.
- A statistical correlation is observed favoring a small value of Λ in models exhibiting a small g_s.

Conclusions

There is a greater probability to find vacua with $\Lambda < 1$ (and $g_s \ll 1$)from a hierarchical flux configuration. This is best in the case of a heavy complex structure scalar field.

Also for hierarchical fluxes the probability of having a stable AdS vacuum in which the lightest modulus mass squared is smaller than Λ_{AdS} increases.

The vacua exploration employing ANN+GA in this case was successful and the method is ready to apply to more extensive datasets employing more refined selection criteria. We plan to apply similar techniques to more generic cases, for example type IIB flux compactifications in CICY varieties.


Thank you!