Generative Models and Statistical Predictions in String Theory

Jim Halverson
Northeastern University

Deep Learning and Physics 2019
@ Yukawa Institute for Theoretical Physics

based on: 1911.xxxxx with C. Long [arXiv + GitHub]
see also: 1809.08279 with F. Rühle
String Theory

- quantum theory of gravity, candidate TOE. can give rise to semi-realistic particle-cosmo.
- extra dimensions → compactify.
  geometry and topology determine 4d physics.
- landscape: many solutions / vacua of theory.
  e.g. Kreuzer-Skarke CY3s, or $10^{755}$ F-theory geometries.
- Upshot: if string theory is true, “fundamental physics” is a complex system, many metastable states.
  (as in, for example, spin glasses, protein folding)
Strings and Machine Learning

- unexpected patterns and correlations.
- deep searches.
- generating conjectures.
- improve + accel. predictions.
- faking actual (theoretical) data.

... and much more. Learning rapidly, wide open O(1) questions. Not everything works, but probably still much to discover.

Summer 2017: [He] [Ruehle] [Carifio, J.H., Nelson, Krioukov] [Krefl, Song]

[Hashimoto, Sugishita, Tanaka, Tamiya] [Cole, Shiu] [Carifio, Cunningham, J.H., Krioukov, Long, Nelson] [Mutter, Parr, Vaudrevange] [Erbin, Krippendorf] [Klaewer, Schlechter] [Liu] [Wang, Zhang] [Jinno] [Bull, He, Jejjala, Mishra]

[He, Jejjala, Nelson] [Rudelius] [Altman, Carifio, J.H., Nelson] [Jejjala, Kar, Parrikar] [Dias, Frazer, Westphal] [Bull, He, Jejjala, Mishra] [Hashimoto] [J.H., Nelson, Ruehle] [He, Lee] [He, Kim] [Brodie, Constantin, Deen, Lukas]

+ ... (sorry if I forgot! let me know and I'll add you next time around)  

Review: [Ruehle, 2019]
Landscape Statistics

- Recall: if string theory is true, physics is complex. i.e. many metastable states, must ask: “why one like ours?”

- Predictions must be statistical. e.g., [Douglas]

\[ \mathbb{E}[O] = \frac{\sum_i U(i) A(i) O(i)}{\sum_i U(i) A(i)} \]

- Ensembles of vacua. Compute expected value of observables in ensemble.

- **Problem:** don’t know full U, A, or \{i\}, and even if we did in theory, in practice we’d be complexity-limited.
What we’re up against

- $10^{755}$ F-theory geometries:
  - $\sim 1800$ axions,
  - $\sim 700$ gauge sectors.

- Reheating critical due to DM concerns.

- **Result:** at $\sim 188$ axions, asymmetric reheating.

* Couldn’t do large N calculation!
Where we’re going

- Complexity.
  affects practical matters and dynamics.

- Learning random matrix approximations.
  simulate string theory to get statistical predictions.

- Examples: Kähler metrics.
  the physics case.
  simulating with conditional Wasserstein GANs.
Complexity

large N is hard. not just b/c # vacua, but also complexity of physics tasks.
Complexity 101

The Good:

- polytime to solve = P
- polytime to check = NP
- NP-hard if at least as hard as any problem in NP.

The Bad:

- intuitively, finding solutions is harder than checking solutions.
  i.e. expect P ≠ NP.
- in that case, **NP-hard problems are exp-time.**
Complexity in String Theory

The ugly:

- Bousso-Polchinski NP-hard. 
  [Denef, Douglas]
- Instantons undecidable. 
  [Cvetič, J.H., Garcia-Etxebarria]
- Finding string scalar potentials and their minima both (co) NP-hard. 
  [JH, Rühl]

- **Dynamics:** if a process is exp-time, how did nature achieve it? 
  [Denef,Douglas] [Denef, Douglas, Greene, Zukowski]

Ways out?

- avoid worst-case instances, entirely or by selection. 
  e.g. protein folding

- allow error: approximate. 
  e.g. FPTAS for NP-hard. polytime in N and 1/ε.
String vacua encode physical observables in ensembles of matrices and tensors.

Can we machine learn random matrix approximations of them?

- Cody Long
Learning Random Matrix Approximations

**Goals:**
1) interpolate rel. training set.  
2) extrapolate to large $N$
Random Matrices

\[ F = A^T A \]

entries of \( A \) drawn from \( \Omega(0, 1/N^{1/2}) \), are random vars.

**Question:** learn accurate \( F \) from sophisticated ansätz?
Learning Random Matrix Approximations of String Theory

**Search ansatz space:**

- Restrict F by human cleverness, likely has params, optimize them.

- e.g. Bergman approx to Kähler. ([Donaldson, Tian, Yau])
  \[ \mathcal{K}(z, \bar{z}) = \frac{1}{k} \log f(z, \bar{z}), \quad f(z, \bar{z}) := s_\alpha(z)P_{\alpha\beta}s_\beta(z) \]

- Optimization possibilities:
  1) random walker
  2) reinforcement learning
  3) gradient descent
  (talk of [Nelson])

**Learn the ansatz:**

- Use very expressive form for F, i.e. can hit a lot of function space.

- F = a deep neural network.

- Training possibilities:
  1) simple network:
     decrease distance of eigenspectra
  2) generative adversarial network:
     competitive game until Nash eq.
Generative Adversarial Networks


[D] D discriminates real vs. fake data.

zero sum game until Nash equil.

\[ v(\theta^g, \theta^d) = \mathbb{E}_{x \sim p_{data}} \log d(x) + \mathbb{E}_{x \sim p_{model}} \log (1 - d(x)) \]

D gets \( v(\theta^g, \theta^d) \), G gets \( -v(\theta^g, \theta^d) \).

learns data distribution.

see [Erbin, Krippendorf] for a string application of GANs, as well as the present work and [J.H., Long, Ruehle] to appear.
Wasserstein GAN (Loss)

- improves training stability due to better behaved gradients.
- idea: use Wasserstein loss, a.k.a. earth-mover distance.
- Estimate with Kantorovich-Rubinstein duality.
- open source: hacked for current application in a morning.
Interpolating and Extrapolating: CGAN

- GAN: fakes data, given noise.
- What if there are multiple classes that appear in the data?
  CGAN: pass condition, generate certain types of data.
- e.g. MNIST: gen. a particular #.
- Q: extrapolate out of sample?

[Mirza, Osindero] 2014
Kähler Metrics

an example of demonstrating and learning structure using RMAs.
Concrete physics: 
Axion kinetic terms

\[ \mathcal{L} = -M_p^2 K_{ij} (\partial^\mu \theta^i)(\partial_\mu \theta^j) - V(\theta) \]
\[ - \sum_\alpha Q_\alpha^i \left( \tau^i G_\alpha^{\mu\nu} G_{\alpha\mu\nu} + \theta^i \tilde{G}_\alpha^{\mu\nu} G_{\alpha\mu\nu} \right) \]

\[ \mathcal{L} = -\frac{1}{2} \delta_{ij} (\partial^\mu \phi^i)(\partial_\mu \phi^j) - \frac{1}{4} F^{\mu\nu} F_{\mu\nu} \]
\[ - \frac{1}{4} c_i \phi^i \tilde{F}^{\mu\nu} F_{\mu\nu} - m_i^2 (\phi^i)^2. \]

- kinetic terms in QFT crucial for propagation of particles.
- **left:** axion kinetic terms with mixing matrix $K_{ij}$.
- **right:** field redefinition with “canonical kinetic terms”,
eigenvalues of $K$ in couplings of canonical ALPs $\phi$. 
where does $K_{ij}$ come from in string theory?

$$\mathcal{K} = -2 \log \mathcal{V}$$

$$\mathcal{V} = \int_B J \wedge J \wedge J$$

derivatives to get $K_{ij}$
Concrete physics: ALP cosmology

Physics:

1) moduli stabilization.

2) axion couplings, have many purposes / uses:
   - cosmological inflation
   - dark matter
   - axion-like particles (ALPs), search for in sun or galaxy clusters via ALP-photon coupling.

Abell 1689, credit: Chandra + Hubble

e.g. CAST (existing), IAXO (future)
ALP-photon couplings & Kahler metrics

- **result:** ALP-photon couplings depend crucially on metric eigenspectra.
so what is this data and how do we compute it?

Kahler metrics on the Kahler moduli space of Kreuzer-Skarke Calabi-Yau threefold hypersurfaces.

(the most popular ensemble of six-dimensional manifolds for extra dimensions in string theory).
Kähler Metric Data

A data point is:

1) particular CY3 type. Kreuzer-Skarke pushing CY3.
   (specified by fine regular star triangulation of reflexive 4d polytope).

2) Evaluate at special point. stretched Kähler cone apex:

   \[ \min \text{vol}(\Sigma_{ij} C_{ij}) \quad \text{s.t.} \quad \text{vol}(C_{ij}) > 0 \]

   were \( C_{ij} = D_i \cdot D_j \).

3) Compute \( K^{ij} \) at \( A \) from \( K = -2 \log(V) \).
Real Data

Structure

- batches of 64, 10x10 K in upper left of 0-padded 16x16.
- hot spots and voids.
- presumably due to topology and / or narrowing Kähler cones?
- see from eyes + string theory.
- how to discover structure?
- Is it human-interpretable?
Training

- Few thousand K-S $h^{11}=10$ CY3s.
- Choose subset to train on, generate Kahler potential.
- Test fitness of generated model.

**Question:** can GAN learn both features and the pad?
$h^{11} = 10$, $n_z = 5$
$h^{11} = 10, \ n_z = 5$
$$h^{11} = 10, \quad n_z = 5$$
Real Data

$ h^{11} = 10, \ n_z = 5 $
Blue: Real Data
Green: GAN @ Epoch
Orange: Naive Wishart

Note: not training eigenvalue loss. instead, learned matrix distribution, eigenvalue learning implicit. Distance = WD.
Real Data

Epoch 20

$h^{11} = 10, \ n_z = 5$

$log_{10}(\text{Eigenvalue})$

Distance = 1.14
Real Data

\( h^{11} = 10, \ n_z = 5 \)

\( h^{11} = 10, \ n_z = 5, \ \text{Distance} = 0.5 \)

\[\log_{10}(\text{Eigenvalue})\]
Real Data

$\hat{h}^{11} = 10, \ n_z = 5$

$\hat{h}^{11} = 10, \ n_z = 5, \ \text{Distance} = 0.22$

$\log_{10}(\text{Eigenvalue})$
Real Data

\[ h^{11} = 10, \ n_z = 5 \]

\[ h^{11} = 10, \ n_z = 5, \ \text{Distance} = 0.08 \]

\textit{paper version}: can get to ~.04
$n_z$ and Implicit correlations

Sculpting material = $n_z$, number of rand. vars in.

**note:**
1) inversion at late times.
2) initial values $\sim 10$.

**small $n_z$ good $\Rightarrow$ correlations.**
This was all at fixed N.

Can we interpolate and extrapolate in N?

often out of sample extrapolation too hard, but maybe actually doable, due to Reid’s fantasy?

(the fantasy that people believe: one big network of CY’s, related by topology changing transitions)
Interpolating & Extrapolating

- **setup:** take Wasserstein DCGAN, make conditional on $h^{11}$
- train on data across different $h^{11}$ values
- Q: do samples from train $h^{11}$ match eigenspectrum?
  
  Q: generate samples with $h^{11}$ not used in training, do they match e-spectrum?

($h^{11}$ between $h^{11}$ trains = interpolation)
($h^{11}$ above $h^{11}$ trains = extrapolation)
Interpolation Results

- **trained** on $h^{11} = 20, 30, 2500$ geometries each
- **test** $h^{11} = 20, 25, 30$ on unseen e-spectrum, $O$(thousands) geoms
- fairly good interpolation ($h^{11} = 25$)
- mediocre @ smaller $h^{11}$. correlated w/ 0-padding!?
Extrapolation Results

- **trained** on $h^{11} = 20, 25, 2500$ geometries each
- **test** $h^{11} = 20, 25, 30$ on unseen e-spectrum, $O$(thousands) geoms
- excellent extrapolation ($h^{11}=30$)
- again, mediocre smaller $h^{11}$. Padding issue clearer!
Predictions in string theory are statistical. Necessitated by complexity at large $N$, can we learn an approximation to it?

**Concrete:** arising from ensemble of matrices, can we learn random matrix approximation?

### Complexity
Instances of NP-hard and undecidable problems arise naturally in string theory. E.g. determining and minimizing scalar potentials.

### Learning RMT
Frameworks:
1) Optimize a clever ansatz.
2) Learn a new one.

Extrapolate out @ fixed $N$. **Interpolate and extrapolate** to other $N$, ideally large!

Here: use Wasserstein (c)GAN with DCGAN architecture.

### Correlations
High accuracy with few random vars? **Structure.** Have it in these examples.

What is controlling this in Kahler metrics?

Physics: directly related to ALP sun-searches.
Thanks!

Any questions?

Ask away, or at a coffee break, or feel free to write:

- @jhhalverson
- jhh@neu.edu

special thanks to A. Achúcarro, D. Robbins, L. McAllister, F. Rühle, G. Shiu, and A. Westphal for discussions on random matrices and string theory.