Deep Quantum Geometry of Matrices

Phys. Rev. X 10, 011069

Xizhi Han and Sean A. Hartnoll

Jan 2021

Stanford University

Roadmap

1. Motivation
2. Matrix model background
3. Numerical results
4. Variational Monte Carlo and machine learning background
5. Wavefunction ansatz
6. Entanglement on the emergent geometry

Motivation

Why machine learning matrix models?

- String theorist?
  - Non-perturbative string physics
  - Black hole microstates and dynamics
  - Holography

Motivation

Why machine learning matrix models?

- String theorist?
  - Non-perturbative string physics
  - Black hole microstates and dynamics
  - Holography
- Machine learning theorist?
  - Representability theorems of neural networks
  - Arena for ML physics architectures
  - Clear numerical advantage

Motivation

Why machine learning matrix models?

- String theorist?
  - Non-perturbative string physics
  - Black hole microstates and dynamics
  - Holography
- Machine learning theorist?
  - Representability theorems of neural networks
  - Arena for ML physics architectures
  - Clear numerical advantage
- IfQ?
  - Tests of QI understandings
  - Complexity of holographic states
  - ...

Basic Example

- What is matrix model?

$$H = \mathbf{tr}(P^2 + X^2 + g^2 X^4)$$

Here X and P are N-by-N hermitian matrices with commutator $[P_{ij}, X_{kl}] = \delta_{il}\delta_{jk}$.

Basic Example

- What is matrix model?

$$H = \mathrm{tr}(P^2 + X^2 + g^2 X^4)$$

  Here X and P are N-by-N hermitian matrices with commutator $[P_{ij}, X_{kl}] = \delta_{il}\delta_{jk}$.

- Where is the emergent geometry?

$$\begin{aligned} Z &= \int dX\, e^{-\,\mathrm{tr}(X^2+X^4)} = \int dx_i \Delta(x_i) e^{-\sum(x_i^2+x_i^4)} \\ &= \int dx_i e^{-\sum(x_i^2+x_i^4)+\sum_{i\neq j}\log|x_i-x_j|} \end{aligned}$$

  The Jacobian $\Delta(x_i) = \prod_{i<j}(x_i - x_j)^2$. Large N is a large number of eigenvalue "particles" with repulsive interaction in an external potential.

Background

- N D-particles in AdS_4

Background

- N D-particles in AdS_4
- One dimensionless parameter ν proportional to fluxes supporting AdS_4

$$F_{tijk} \sim \Omega \epsilon_{ijk}, \quad \nu \sim \frac{\Omega l_s}{g_s^{1/3}} \sim \frac{1}{g_s^{1/3}} \frac{l_s}{l_{\mathrm{AdS}}}$$

- Gravitational backreaction becomes important when

$$\frac{N}{\nu^3} \gtrsim 1$$

- Small ν: gravitational collapse; Large ν: fuzzy sphere

Mini-BMN Hamiltonian

$$H_B = \mathrm{tr}\left(\frac{1}{2}\Pi^i\Pi^i - \frac{1}{4}[X^i, X^j][X^i, X^j] + \frac{1}{2}\nu^2 X^i X^i + i\nu\epsilon^{ijk}X^i X^j X^k\right)$$

$$H_F = \mathrm{tr}\left(\lambda^\dagger \sigma^k [X^k, \lambda] + \frac{3}{2}\nu\lambda^\dagger\lambda\right) - \frac{3}{2}\nu(N^2 - 1)$$

The Hamiltonian $H = H_B + H_F$ describes a quantum mechanical system of $i = 1, 2, 3$ bosonic and $\alpha = 1, 2$ fermionic N-by-N traceless hermitian matrices.

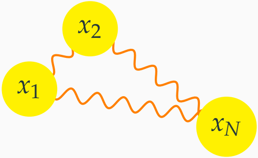
Mini-BMN Hamiltonian

$$H_B = \mathrm{tr}\left(\frac{1}{2}\Pi^i\Pi^i - \frac{1}{4}[X^i, X^j][X^i, X^j] + \frac{1}{2}\nu^2 X^i X^i + i\nu\epsilon^{ijk}X^i X^j X^k\right)$$

$$H_F = \mathrm{tr}\left(\lambda^\dagger\sigma^k[X^k, \lambda] + \frac{3}{2}\nu\lambda^\dagger\lambda\right) - \frac{3}{2}\nu(N^2-1)$$

The Hamiltonian $H = H_B + H_F$ describes a quantum mechanical system of $i = 1, 2, 3$ bosonic and $\alpha = 1, 2$ fermionic N-by-N traceless hermitian matrices.

$$\begin{pmatrix} x_1^i & y_{12}^i & \dots & y_{1N}^i \\ y_{12}^{i*} & x_2^i & \dots & y_{2N}^i \\ \vdots & \vdots & \ddots & \vdots \\ y_{1N}^{i*} & y_{2N}^{i*} & \dots & x_N^i \end{pmatrix}$$



$$[X^i, X^j]^2 \sim (x_a^i - x_b^i)^2 |y_{ab}^j|^2$$

Emergent Fuzzy Sphere in Mini-BMN

The bosonic potential can be conveniently written as

$$V(X) = \frac{1}{4}\,\mathrm{tr}\left[\left(\nu\epsilon^{ijk}X^k + i[X^i, X^j]\right)^2\right]$$

and is minimized by

$$[X^i, X^j] = i\nu\epsilon^{ijk}X^k\,.$$

Emergent Fuzzy Sphere in Mini-BMN

The bosonic potential can be conveniently written as

$$V(X) = \frac{1}{4}\,\mathrm{tr}\left[\left(\nu\epsilon^{ijk}X^k + i[X^i, X^j]\right)^2\right]$$

and is minimized by

$$[X^i, X^j] = i\nu\epsilon^{ijk}X^k\,.$$

So classically, three bosonic matrices satisfy the $\mathfrak{so}(3)$ algebra:

$$X^i = \nu J^i,$$

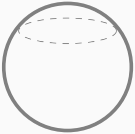
where J^i are representations of the algebra $[J^i, J^j] = i\epsilon^{ijk}J^k$ of dim N.

Emergent Fuzzy Sphere in Mini-BMN

The emergence of the $\mathfrak{so}(3)$ algebra motivates a correspondence between matrices and fields on a sphere:

$$\begin{pmatrix} x_1^i & y_{12}^i & \cdots & y_{1N}^i \\ y_{12}^{i*} & x_2^i & \cdots & y_{2N}^i \\ \vdots & \vdots & \ddots & \vdots \\ y_{1N}^{i*} & y_{2N}^{i*} & \cdots & x_N^i \end{pmatrix}$$

Matrix	Fields
M	$f(\theta,\phi)$
$\alpha M_1 + \beta M_2$	$\alpha f_1(\theta,\phi) + \beta f_2(\theta,\phi)$
$M_1 M_2$	$f_1 \star f_2 = f_1(\theta,\phi) f_2(\theta,\phi) + \ldots$
$[X^i, M]$	$L^i f(\theta,\phi)$
$\frac{1}{N}\,\mathrm{tr}\, M_1^\dagger M_2$	$\frac{1}{4\pi}\int d\Omega\, f_1^*(\theta,\phi) f_2(\theta,\phi)$
$I, X^i, \ldots$	$1, x^i, \ldots$



Example

$R^2 = (x^i)^2 = \frac{1}{4\pi}\int d\Omega\, x^{i*} x^i = \frac{1}{N}\,\mathrm{tr}(X^i)^\dagger X^i = \frac{1}{4}\nu^2(N^2-1)$

Noncommutative Gauge Theory on the Fuzzy Sphere

Now consider quantum fluctuations around the classical solution:

$$X^i = \nu J^i + \sqrt{\frac{4\pi}{N\nu}} A^i,$$

Noncommutative Gauge Theory on the Fuzzy Sphere

Now consider quantum fluctuations around the classical solution:

$$X^i = \nu J^i + \sqrt{\frac{4\pi}{N\nu}} A^i,$$

the bosonic potential can be rewritten as

$$V(X) = \frac{1}{4} \operatorname{tr} \left[\left(\nu \epsilon^{ijk} X^k + i[X^i, X^j] \right)^2 \right] = \frac{4\pi\nu}{4N} \operatorname{tr} \left(F^{ij} \right)^2 ,$$

where

$$F^{ij} = i \left([J^i, A^j] - [J^j, A^i] \right) + \epsilon^{ijk} A^k + i \sqrt{\frac{4\pi}{N\nu^3}} [A^i, A^j].$$

Noncommutative Gauge Theory on the Fuzzy Sphere

Now consider quantum fluctuations around the classical solution:

$$X^i = \nu J^i + \sqrt{\frac{4\pi}{N\nu}} A^i,$$

the bosonic potential can be rewritten as

$$V(X) = \frac{1}{4}\,\mathrm{tr}\left[\left(\nu\epsilon^{ijk}X^k + i[X^i, X^j]\right)^2\right] = \frac{4\pi\nu}{4N}\,\mathrm{tr}\left(F^{ij}\right)^2,$$

where

$$F^{ij} = i\left([J^i, A^j] - [J^j, A^i]\right) + \epsilon^{ijk}A^k + i\sqrt{\frac{4\pi}{N\nu^3}}[A^i, A^j].$$

Correspondingly,

$$f^{ij} = i\left(L^i a^j - L^j a^i\right) + \epsilon^{ijk}a^k + i\sqrt{\frac{4\pi}{N\nu^3}}[a^i, a^j]_\star,$$

$$H_B = \nu \int d\Omega \left(\frac{1}{2}(\pi^i)^2 + \frac{1}{4}(f^{ij})^2\right).$$

Solvable if $N\nu^3 \to \infty$!

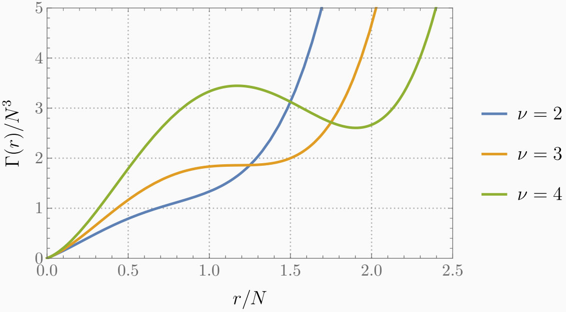
Recap

- Supersymmetric matrix quantum mechanics of 3 bosonic and 2 fermionic SU(N) matrices, with one (dimensionless) mass deformation parameter ν.
- In the limit $N\nu^3 \to \infty$, a fuzzy sphere with a noncommutative U(1) gauge theory emerges, and the theory is solvable in this limit.
- The emergent gauge field is the fluctuation around the classical solution $X^i = \nu J^i$, under the matrix-field correspondence $M \leftrightarrow f(\theta, \phi)$.

Solvable Case One: $\nu = \infty$, $R = 0$

As $N|\nu|^3 \to \infty$, the theory is solvable with

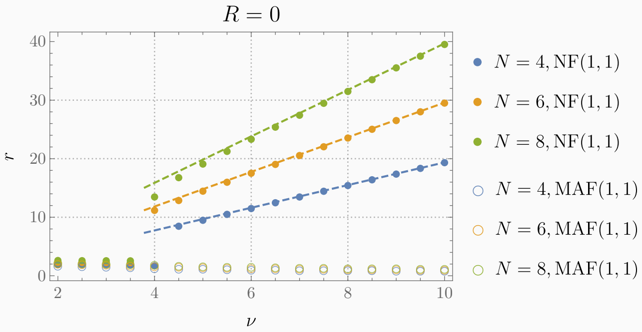
- Radius of the emergent sphere $r_0 \sim \frac{1}{2} N|\nu|$
- Ground state energy $E \sim \frac{2}{3} N^3 |\nu|$
- First-order phase transition at $\nu \approx 3$ (one-loop)



Solvable Case One: $\nu = \infty$, $R = 0$

As $N|\nu|^3 \to \infty$, the theory is solvable with

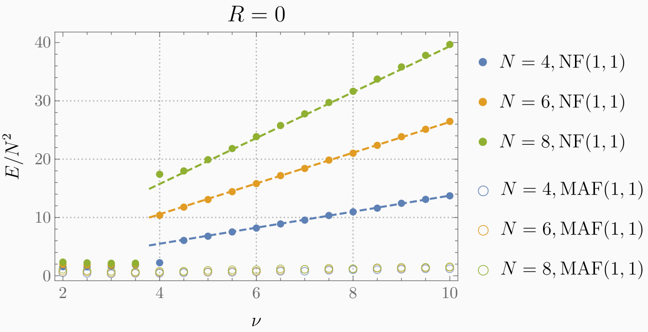
- Radius of the emergent sphere $r_0 \sim \frac{1}{2}N|\nu|$
- Ground state energy $E \sim \frac{2}{3}N^3|\nu|$
- First-order phase transition at $\nu \approx 3$ (one-loop)



Solvable Case One: $\nu = \infty$, $R = 0$

As $N|\nu|^3 \to \infty$, the theory is solvable with

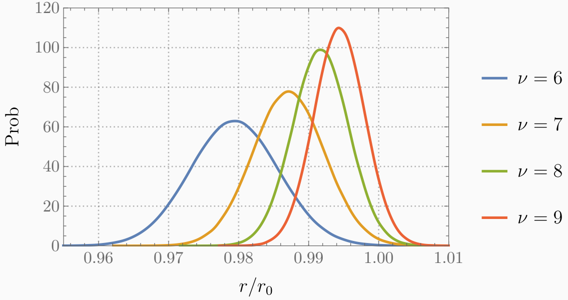
- Radius of the emergent sphere $r_0 \sim \frac{1}{2}N|\nu|$
- Ground state energy $E \sim \frac{2}{3}N^3|\nu|$
- First-order phase transition at $\nu \approx 3$ (one-loop)



Solvable Case One: $\nu = \infty$, $R = 0$

As $N|\nu|^3 \to \infty$, the theory is solvable with

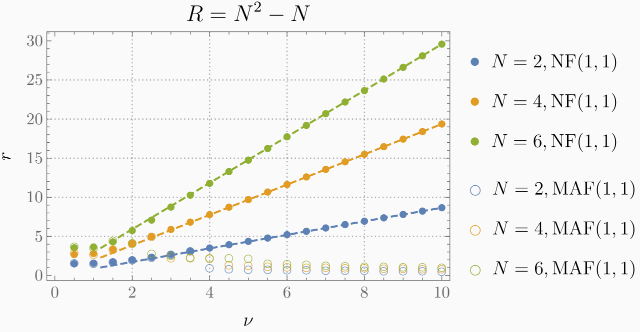
- Radius of the emergent sphere $r_0 \sim \frac{1}{2}N|\nu|$
- Ground state energy $E \sim \frac{2}{3}N^3|\nu|$
- First-order phase transition at $\nu \approx 3$ (one-loop)



Solvable Case Two: $\nu = \infty$, $R = N^2 - N$

As $N|\nu|^3 \to \infty$, the theory is solvable with

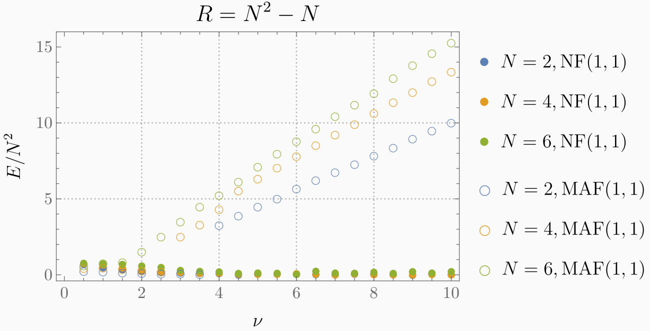
- Radius of the emergent sphere $r_0 \sim \frac{1}{2}N|\nu|$
- Ground state energy $E/N^2 \to 0$
- No first-order phase transitions to one-loop



Solvable Case Two: $\nu = \infty$, $R = N^2 - N$

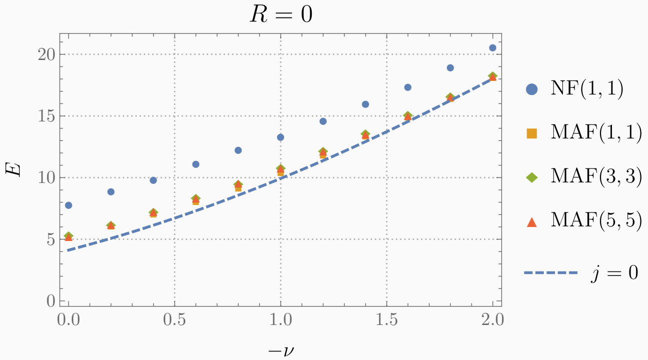
As $N|\nu|^3 \to \infty$, the theory is solvable with

- Radius of the emergent sphere $r_0 \sim \frac{1}{2}N|\nu|$
- Ground state energy $E/N^2 \to 0$
- No first-order phase transitions to one-loop



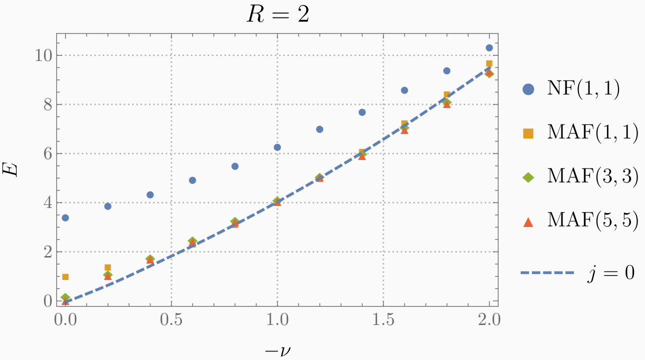
Solvable Case Three: $N = 2$

For two-by-two matrices, the mini-BMN was solved in different sectors in arXiv:1701.07511.



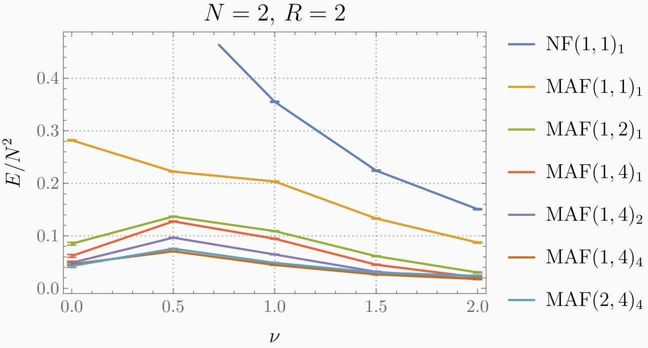
Solvable Case Three: $N = 2$

For two-by-two matrices, the mini-BMN was solved in different sectors in arXiv:1701.07511.



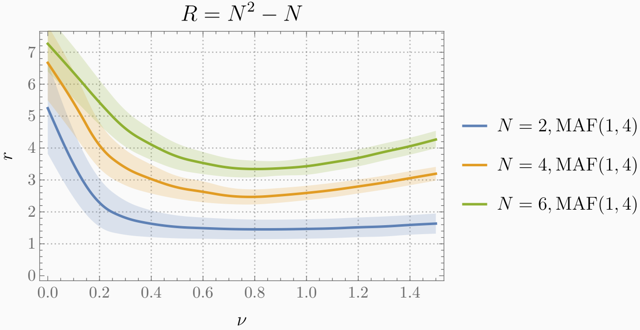
Exploration near $\nu = 0$

- Small ν (gravitational) regime is more quantum and difficult
- Complexity of the network is necessary to get better results

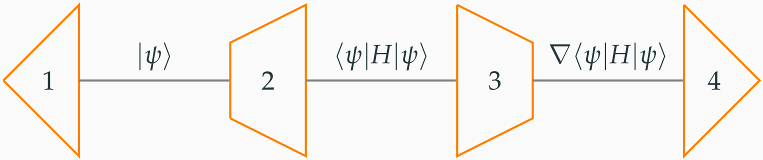


Exploration near $\nu = 0$

- Small ν (gravitational) regime is more quantum and difficult
- Complexity of the network is necessary to get better results



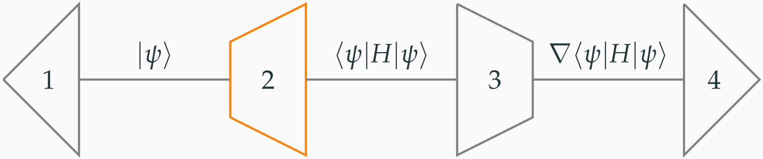
Variational Monte Carlo



1. Parametrize the variational wavefunction $\psi_\theta(X)$ by parameters θ
2. Estimate the expectation value of energy by Monte Carlo samples
3. Evaluate the gradient of the energy with respect to θ
4. Apply the gradient to the parameters via gradient descent

$$\theta \to \theta - \alpha \nabla_\theta \langle\psi_\theta|H|\psi_\theta\rangle$$

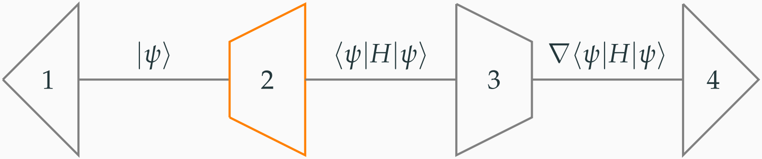
Monte Carlo Estimate of Energy



- Bosonic potential:

$$\langle\psi|V(X)|\psi\rangle = \int dX\, |\psi(X)|^2 V(X) = \mathbb{E}_{X\sim|\psi|^2}[V(X)]$$

Monte Carlo Estimate of Energy



- Bosonic potential:

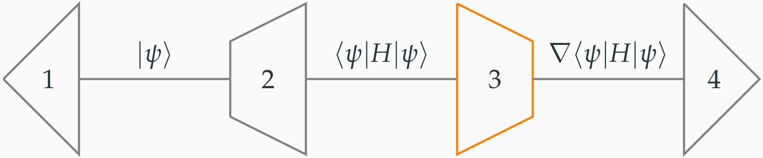
$$\langle\psi|V(X)|\psi\rangle = \int dX\,|\psi(X)|^2 V(X) = \mathbb{E}_{X\sim|\psi|^2}[V(X)]$$

- Kinetic terms:

$$\langle\psi|\operatorname{tr}\Pi^2|\psi\rangle = \sum_{ij}\int dX\,\left|\frac{\partial\psi}{\partial X_{ij}}\right|^2 = \sum_{ij}\int dX\,|\psi(X)|^2\left|\frac{\partial\ln\psi}{\partial X_{ij}}\right|^2$$

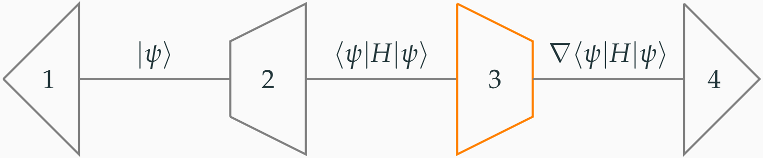
$$= \mathbb{E}_{X\sim|\psi|^2}\left[\sum_{ij}\left|\frac{\partial\ln\psi}{\partial X_{ij}}\right|^2\right]$$

Reinforcement Gradient of Energy



Objective: minimize $E_\theta = \langle\psi_\theta|H|\psi_\theta\rangle = \mathbb{E}_{X\sim|\psi_\theta|^2}[\epsilon_\theta(X)]$.

Reinforcement Gradient of Energy

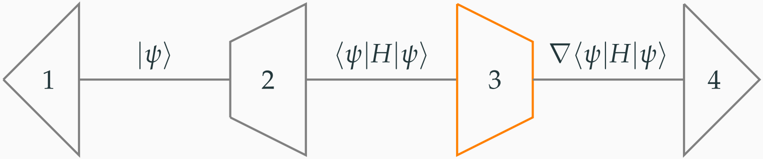


Objective: minimize $E_\theta = \langle\psi_\theta|H|\psi_\theta\rangle = \mathbb{E}_{X\sim|\psi_\theta|^2}[\epsilon_\theta(X)]$.

Gradient:

$$\nabla_\theta E_\theta = \mathbb{E}_{X\sim|\psi_\theta|^2}[\nabla_\theta \epsilon_\theta(X)] + \mathbb{E}_{X\sim|\psi_\theta|^2}[2\epsilon_\theta(X)\nabla_\theta \ln|\psi_\theta(X)|]$$

Reinforcement Gradient of Energy



Objective: minimize $E_\theta = \langle\psi_\theta|H|\psi_\theta\rangle = \mathbb{E}_{X\sim|\psi_\theta|^2}[\epsilon_\theta(X)]$.

Gradient:

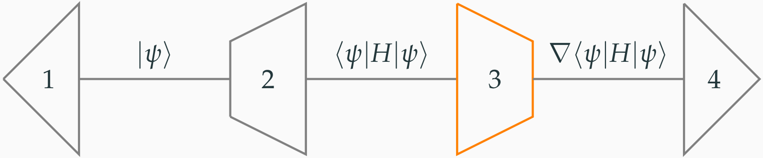
$$\nabla_\theta E_\theta = \mathbb{E}_{X\sim|\psi_\theta|^2}[\nabla_\theta \epsilon_\theta(X)] + \mathbb{E}_{X\sim|\psi_\theta|^2}[2\epsilon_\theta(X)\nabla_\theta \ln|\psi_\theta(X)|]$$

To minimize variance, the second term can be rewritten as

$$\mathbb{E}_{X\sim|\psi_\theta|^2}[2(\epsilon_\theta(X) - E_\theta)\nabla_\theta \ln|\psi_\theta(X)|]$$

because $\mathbb{E}[2\nabla_\theta \ln|\psi|] = \nabla_\theta \int dX |\psi|^2 = 0$.

Reinforcement Gradient of Energy



Objective: minimize $E_\theta = \langle\psi_\theta|H|\psi_\theta\rangle = \mathbb{E}_{X\sim|\psi_\theta|^2}[\epsilon_\theta(X)]$.

Gradient:

$$\nabla_\theta E_\theta = \mathbb{E}_{X\sim|\psi_\theta|^2}[\nabla_\theta \epsilon_\theta(X)] + \mathbb{E}_{X\sim|\psi_\theta|^2}[2\epsilon_\theta(X)\nabla_\theta \ln|\psi_\theta(X)|]$$

To make it work without having to normalize the wavefunction, the second term is also

$$\mathbb{E}_{X\sim|\psi_\theta|^2}[2(\epsilon_\theta(X) - E_\theta)\nabla_\theta \ln(Z_\theta|\psi_\theta(X)|)]$$

for any function Z_θ of θ, as $\mathbb{E}[(\epsilon - E)\nabla Z] = \mathbb{E}[\epsilon - E]\nabla Z = 0$.

Recap

For any (not necessarily normalized) wavefunction $\psi_\theta(X)$,

$$E_\theta = \mathbb{E}_{X\sim|\psi_\theta|^2}[\epsilon_\theta(X)]$$

$$\nabla_\theta E_\theta = \mathbb{E}_{X\sim|\psi_\theta|^2}[\nabla_\theta \epsilon_\theta(X)] + \mathbb{E}_{X\sim|\psi_\theta|^2}[2(\epsilon_\theta(X) - E_\theta)\nabla_\theta \ln|\psi_\theta(X)|]$$

Recap

For any (not necessarily normalized) wavefunction $\psi_\theta(X)$,

$$E_\theta = \mathbb{E}_{X\sim|\psi_\theta|^2}[\epsilon_\theta(X)]$$

$$\nabla_\theta E_\theta = \mathbb{E}_{X\sim|\psi_\theta|^2}[\nabla_\theta \epsilon_\theta(X)] + \mathbb{E}_{X\sim|\psi_\theta|^2}[2(\epsilon_\theta(X) - E_\theta)\nabla_\theta \ln|\psi_\theta(X)|]$$

- We need samples of the wavefunction to evaluate

$$\mathbb{E}_{X\sim|\psi_\theta|^2}[F(X)] = \frac{1}{K}\sum_{i=1}^{K} F(X_i).$$

Recap

For any (not necessarily normalized) wavefunction $\psi_\theta(X)$,

$$E_\theta = \mathbb{E}_{X\sim|\psi_\theta|^2}[\epsilon_\theta(X)]$$

$$\nabla_\theta E_\theta = \mathbb{E}_{X\sim|\psi_\theta|^2}[\nabla_\theta \epsilon_\theta(X)] + \mathbb{E}_{X\sim|\psi_\theta|^2}[2(\epsilon_\theta(X) - E_\theta)\nabla_\theta \ln|\psi_\theta(X)|]$$

- We need samples of the wavefunction to evaluate

$$\mathbb{E}_{X\sim|\psi_\theta|^2}[F(X)] = \frac{1}{K}\sum_{i=1}^{K} F(X_i).$$

- We need the explicit functional form of $|\psi_\theta(X)|$ to efficiently evaluate the gradient.

Recap

For any (not necessarily normalized) wavefunction $\psi_\theta(X)$,

$$E_\theta = \mathbb{E}_{X\sim|\psi_\theta|^2}[\epsilon_\theta(X)]$$

$$\nabla_\theta E_\theta = \mathbb{E}_{X\sim|\psi_\theta|^2}[\nabla_\theta \epsilon_\theta(X)] + \mathbb{E}_{X\sim|\psi_\theta|^2}[2(\epsilon_\theta(X) - E_\theta)\nabla_\theta \ln|\psi_\theta(X)|]$$

- We need samples of the wavefunction to evaluate

$$\mathbb{E}_{X\sim|\psi_\theta|^2}[F(X)] = \frac{1}{K}\sum_{i=1}^{K} F(X_i).$$

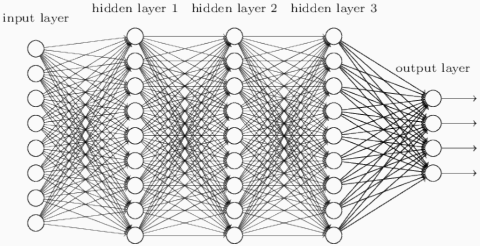
- We need the explicit functional form of $|\psi_\theta(X)|$ to efficiently evaluate the gradient.
- Efficient sampling and evaluating $|\psi_\theta(X)|^2 \Rightarrow$ Generative Flows

Building Blocks: Fully-Connected Neural Networks

The neural network defines a function $F : x \mapsto y$ mapping an input vector x to an output vector y via a sequence of affine and nonlinear transformations:

$$F = A_\theta^m \circ \tanh \circ A_\theta^{m-1} \circ \tanh \circ \cdots \circ \tanh \circ A_\theta^1 \,.$$

Here $A_\theta^1(x) = M_\theta^1 x + b_\theta^1$ is an affine transformation. The hyperbolic tangent nonlinearity then acts elementwise on $A_\theta^1(x)$.



[1]

[1] Figure from neuralnetworksanddeeplearning.com

Generative Flow One: Normalizing Flows

Normalizing flows give an efficient way of parametrizing complicated probability distributions ($|\psi_\theta(X)|^2$ in our case). For any reversible transformation F and $y = F(x)$:

$$p_y(y_0) = p_x(x_0)|\det DF|_{x_0}^{-1},$$

where $x_0 = F^{-1}(y_0)$.

Generative Flow One: Normalizing Flows

Normalizing flows give an efficient way of parametrizing complicated probability distributions ($|\psi_\theta(X)|^2$ in our case). For any reversible transformation F and $y = F(x)$:

$$p_y(y_0) = p_x(x_0)|\det DF|_{x_0}^{-1},$$

where $x_0 = F^{-1}(y_0)$.

In practice, F can be parametrized by neural networks composed of reversible affine and nonlinear transformations.

Generative Flow One: Normalizing Flows

Normalizing flows give an efficient way of parametrizing complicated probability distributions ($|\psi_\theta(X)|^2$ in our case). For any reversible transformation F and $y = F(x)$:

$$p_y(y_0) = p_x(x_0) |\det DF|_{x_0}^{-1},$$

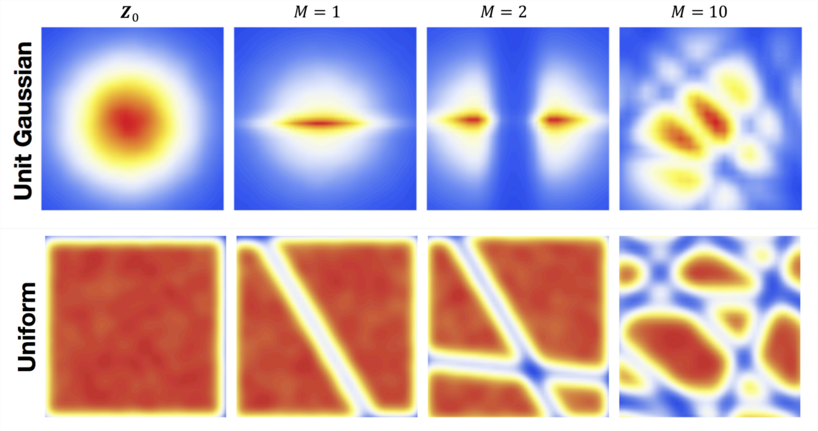
where $x_0 = F^{-1}(y_0)$.

In practice, F can be parametrized by neural networks composed of reversible affine and nonlinear transformations.

$$\begin{pmatrix} \exp x_1 & y_{12} & y_{13} & y_{14} \\ 0 & \exp x_2 & y_{23} & y_{24} \\ 0 & 0 & \exp x_3 & y_{34} \\ 0 & 0 & 0 & \exp x_4 \end{pmatrix}$$

Generative Flow One: Normalizing Flows

From arXiv: 1505.05770:



Generative Flow Two: Masked Autoregressive Flows

A more clever way of parametrizing a reversible transformation is to choose an ordering of the coordinates

$$
\begin{aligned}
x_1 &\sim p_1(x;\theta_0) \\
x_2 &\sim p_2(x;\theta_1(x_1)) \\
&\vdots \\
x_i &\sim p_i(x;\theta_{i-1}(x_1,\ldots,x_{i-1}))
\end{aligned}
$$

Generative Flow Two: Masked Autoregressive Flows

A more clever way of parametrizing a reversible transformation is to choose an ordering of the coordinates

$$\begin{aligned} x_1 &\sim p_1(x;\theta_0) \\ x_2 &\sim p_2(x;\theta_1(x_1)) \\ &\vdots \\ x_i &\sim p_i(x;\theta_{i-1}(x_1,\ldots,x_{i-1})) \end{aligned}$$

The joint probability distribution of x_i is a product of individual ones

$$p(x_1,x_2,\ldots) = p_1(x_1;\theta_0)p_2(x_2;\theta_1(x_1))\ldots$$

Generative Flow Two: Masked Autoregressive Flows

A more clever way of parametrizing a reversible transformation is to choose an ordering of the coordinates

$$
\begin{aligned}
x_1 &\sim p_1(x;\theta_0) \\
x_2 &\sim p_2(x;\theta_1(x_1)) \\
&\vdots \\
x_i &\sim p_i(x;\theta_{i-1}(x_1,\ldots,x_{i-1}))
\end{aligned}
$$

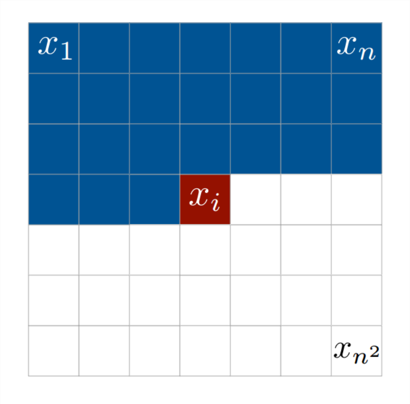
The joint probability distribution of x_i is a product of individual ones

$$p(x_1,x_2,\ldots) = p_1(x_1;\theta_0)p_2(x_2;\theta_1(x_1))\ldots$$

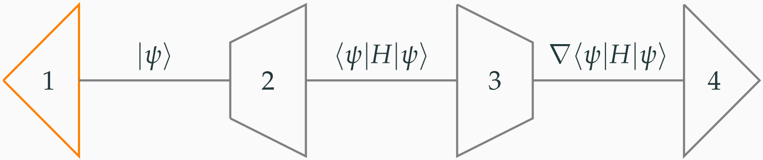
And θ_i can be parametrized by fully-connected neural networks without reversibility requirements.

Generative Flow Two: Masked Autoregressive Flows

To generate an image pixel-by-pixel:



Wavefunction Ansatz: Gauge Invariance



(Think the wavefunction as functions from bosonic matrices to fermion states.) The $\mathrm{SU}(N)$ gauge invariance of the wavefunction requires that

$$\psi(UXU^{-1}) = U\psi(X)U^{-1}.$$

One way to impose this is to fix a gauge $X = U\widetilde{X}U^{-1}$ where $\widetilde{X}$ is the gauge representative for X, and define

$$\psi(X) = U\widetilde{\psi}(\widetilde{X})U^{-1},$$

where $\widetilde{\psi}$ is a function of gauge representatives only.

Wavefunction Ansatz: Bosons

Objective: Sample $X \sim |\psi(X)|^2$ and evaluate $|\psi(X)|^2$ for any X.

Given: Generative flow $p_\theta(X)$, a probability distribution parametrized by θ, capable of evaluating $p_\theta(X)$ for any X and sample $X \sim p_\theta(X)$.

Wavefunction Ansatz: Bosons

Objective: Sample $X \sim |\psi(X)|^2$ and evaluate $|\psi(X)|^2$ for any X.

Given: Generative flow $p_\theta(X)$, a probability distribution parametrized by θ, capable of evaluating $p_\theta(X)$ for any X and sample $X \sim p_\theta(X)$.

Solution:

1. Sample gauge representatives $\widetilde{X} \sim p_\theta(\widetilde{X})$

Wavefunction Ansatz: Bosons

Objective: Sample $X \sim |\psi(X)|^2$ and evaluate $|\psi(X)|^2$ for any X.

Given: Generative flow $p_\theta(X)$, a probability distribution parametrized by θ, capable of evaluating $p_\theta(X)$ for any X and sample $X \sim p_\theta(X)$.

Solution:

1. Sample gauge representatives $\widetilde{X} \sim p_\theta(\widetilde{X})$
2. Sample a random gauge group element $g \in \mathrm{SU}(N)$

Wavefunction Ansatz: Bosons

Objective: Sample $X \sim |\psi(X)|^2$ and evaluate $|\psi(X)|^2$ for any X.

Given: Generative flow $p_\theta(X)$, a probability distribution parametrized by θ, capable of evaluating $p_\theta(X)$ for any X and sample $X \sim p_\theta(X)$.

Solution:

1. Sample gauge representatives $\widetilde{X} \sim p_\theta(\widetilde{X})$
2. Sample a random gauge group element $g \in \mathsf{SU}(N)$
3. Output samples $X = g\widetilde{X}$, then X follows the probability distribution $|\psi(X)|^2 \equiv p_\theta(\widetilde{X})/\Delta(\widetilde{X})$

Comments: $\Delta(\widetilde{X})$ is the size of the gauge orbit of $\widetilde{X}$. The functional form of $\Delta(\widetilde{X})$ must be known.

Wavefunction Ansatz: Fermions

- With fermions, $\widetilde{\psi}(\widetilde{X}) = |\widetilde{\psi}(\widetilde{X})||M(\widetilde{X})\rangle$, where

$$|M(X)\rangle \equiv \sum_{r=1}^{D} \prod_{a=1}^{R} \Big(\sum_{\alpha=1}^{2} \sum_{A=1}^{N^2-1} M_{A\alpha}^{ra}(X) \lambda_A^{\alpha\dagger} \Big) |0\rangle,$$

λ_A^α are matrices of fermions and $|0\rangle$ is the state with zero fermions.

Wavefunction Ansatz: Fermions

- With fermions, $\widetilde{\psi}(\widetilde{X}) = |\widetilde{\psi}(\widetilde{X})||M(\widetilde{X})\rangle$, where

$$|M(X)\rangle \equiv \sum_{r=1}^{D} \prod_{a=1}^{R} \Big(\sum_{\alpha=1}^{2} \sum_{A=1}^{N^2-1} M_{A\alpha}^{ra}(X) \lambda_A^{\alpha\dagger} \Big) |0\rangle,$$

  λ_A^α are matrices of fermions and $|0\rangle$ is the state with zero fermions.
- The fermionic part of the wavefunction is given by a complex tensor function $M_{A\alpha}^{ra}(X)$, parametrized by a fully-connected neural network as well.

Wavefunction Ansatz: Fermions

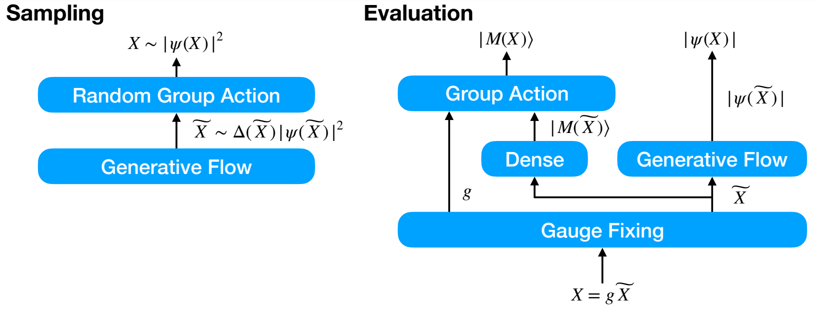
- With fermions, $\widetilde{\psi}(\widetilde{X}) = |\widetilde{\psi}(\widetilde{X})| |M(\widetilde{X})\rangle$, where

$$|M(X)\rangle \equiv \sum_{r=1}^{D} \prod_{a=1}^{R} \Big(\sum_{\alpha=1}^{2} \sum_{A=1}^{N^2-1} M_{A\alpha}^{ra}(X) \lambda_A^{\alpha\dagger} \Big) |0\rangle,$$

  λ_A^α are matrices of fermions and $|0\rangle$ is the state with zero fermions.

- The fermionic part of the wavefunction is given by a complex tensor function $M_{A\alpha}^{ra}(X)$, parametrized by a fully-connected neural network as well.
- Pros: easy to compute energy — no sign problems;
  Cons: not very efficient.

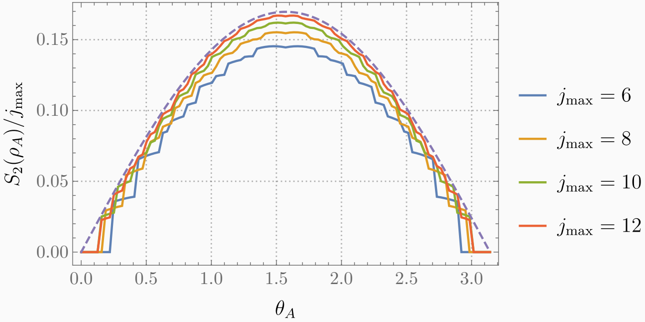
Recap



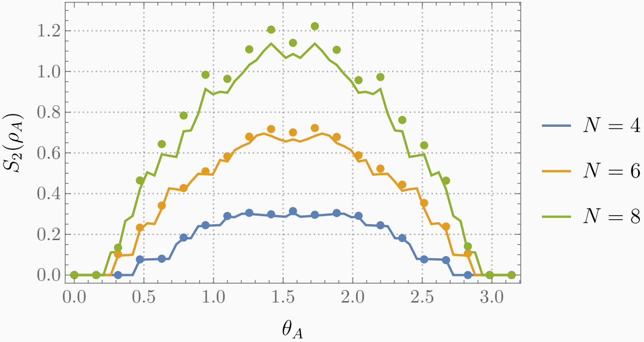
In this work $|\psi(X)|^2$ is represented by a normalizing flow (NF) or masked autoregressive flow (MAF), and the fermion wavefunction is a superposition of free fermion states.

Entanglement of Free Fields with $j \leq j_{\max}$

Find $P_A^{j_{\max}}$ in the space of fields with $j \leq j_{\max}$ as the projector that minimizes the distance $\|P_A^{j_{\max}} - P_A\|$.



At large $j_{\max}$ the curve approaches the boundary law $0.03 \times 2\pi \sin\theta_A$, shown as a dashed line.



Solid curves are for $\nu = \infty$ and dots are from numerics for $\nu = 10$.

Conclusion

- We demonstrate that it is possible to discover low-energy wavefunctions of matrix quantum mechanics with a deep learning approach.
- The variational ground state is quantitatively accurate in the solvable fuzzy sphere phase, from both observable and entanglement metrics.
- For the more interesting limit with emergent gravity, more physical insights / more efficient architectures are required.
- We also propose a way of evaluating entanglement on noncommutative geometries, which reveals an area-law entanglement on the emergent fuzzy sphere.
- Please check out our paper for more references.
  Thank you!