Solving quantum many-body Hamiltonians with artificial neural networks

Yusuke NOMURA
Univ. of Tokyo

Collaborators: Andrew S. Darmawan, Youhei Yamaji, Giuseppe Carleo, Masatoshi Imada

YN, A. Darmawan, Y. Yamaji, and M. Imada, PRB 96, 205152 (2017)
Artificial Neural Networks in Condensed Matter Physics

**Phase classification** (discriminative model)

T. Ohtsuki and T. Ohtsuki, JPSJ 85, 123706 (2016).
A. Tanaka and A. Tomiya, JPSJ 86, 063001 (2017).
T. Ohtsuki and T. Ohtsuki, JPSJ 86, 044708 (2017).
N. Yoshioka et al., arXiv:1709.05790.

**Many-body solver** (generative model)

See also: H. Saito, JPSJ 86, 093001 (2017);
H. Saito and M. Kato, JPSJ 87, 014001 (2018);
YN et al., PRB 96, 205152 (2017)

**Monte Carlo speed up** (generative model)

L. Huang and L. Wang, PRB 95, 035105 (2017)
L. Wang, PRE 96, 051301 (2017)
(Related) G.Torlai and R. G. Melko PRB 94, 165134 (2016)
Restricted Boltzmann machine (RBM)

Energy function

\[ E(\sigma, h) = - \sum_{i=1}^{N} a_i \sigma_i - \sum_{i=1}^{N} \sum_{k=1}^{M} W_{ik} \sigma_i h_k - \sum_{k=1}^{M} b_k h_k \]

\[ \sigma \in \{0, 1\}^N \quad h \in \{0, 1\}^M \]

Boltzmann distribution

\[ p(\sigma, h) = \frac{e^{-E(\sigma, h)}}{Z} \]

\[ Z = \sum_{\sigma, h} e^{-E(\sigma, h)} \]

Marginal distribution

\[ \tilde{p}(\sigma) = \sum_{h} p(\sigma, h) \]

Single hidden layer + interlayer coupling only \(\rightarrow\) restricted Boltzmann machine (RBM)

Marginal distribution \(\tilde{p}(\sigma)\) can represent any distribution over \(\{0,1\}^N\) with infinite M

Paul Smolensky (1986)

Using artificial neural network to solve quantum many body problems

Interaction $W_{ij}$

Variational wave function

$\Psi(\sigma^z) = \sum_{\{h_j\}} \exp\left(\sum_i a_i \sigma_i^z + \sum_{i,j} \sigma_i^z W_{ij} h_j + \sum_j b_j h_j\right)$

$\sigma^z = (\sigma_1^z, \sigma_2^z, \ldots, \sigma_N^z) : \text{real space spin config.}$

$h_j = \pm 1 : \text{spin of hidden neuron}$

Quantum correlations among physical spins via artificial neural network

Single hidden layer + interlayer coupling only $\rightarrow$ restricted Boltzmann machine (RBM)

$\Psi(\sigma^z) = e^{\sum_i a_i \sigma_i^z} \times \prod_j 2 \cosh\left(b_j + \sum_i W_{ij} \sigma_i^z\right)$

See also: H. Saito, JPSJ 86, 093001 (2017);
Optimization strategy

many-body wave function = vector with exponentially large dimension
→ extract essential pattern from machine leaning and represent it with polynomial number of parameters
※ For the moment, we will consider positive-definite wave function → wave function can be regarded as probability density function

1. We know exact \( \psi (x) \)
   → exact \( \psi (x) \) = teacher \ ((supervised learning))

2. We do not know the form of \( \psi (x) \), but we can observe real-space configuration \( x \) generated according to \( \psi (x) \)
   → density estimation: estimate underlying probability density function \ (unsupervised learning) \)

3. We do not know the form of \( \psi (x) \), we cannot observe real-space configuration \( x \) either
   (most challenging)
   → finding unknown ground sate

we teach machine “rule of the game” (commutation relation, Hamiltonian, measurement of energy, ⋅⋅⋅)
optimize parameters following variational principle = minimization of energy
Example: 1D Antiferromagnetic Heisenberg model (8 site)

\[
H_{\text{Heisenberg}} = J \sum_{(i,j)} \sigma_i \cdot \sigma_j = J \sum_{(i,j)} (\sigma_i^x \sigma_j^x + \sigma_i^y \sigma_j^y + \sigma_i^z \sigma_j^z) \quad (J > 0)
\]

Gauge transformation:

\[
\sigma^{xy} \rightarrow -\sigma^{xy}
\]

for one of sublattice

Optimization following variational principle:

\[
\langle \mathcal{H} \rangle = \frac{\langle \Psi | \mathcal{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \geq E_0 \quad \text{E}_0: \text{ground state energy}
\]

Energy:

\[
\langle \mathcal{H} \rangle = \sum_x p(x) E_{\text{loc}}(x)
\]

\[
p(x) = \frac{|\Psi(x)|^2}{\sum_x |\Psi(x)|^2}
\]

\[
E_{\text{loc}}(x) = \sum_{x'} \langle x | H | x' \rangle \frac{\Psi(x')}{\Psi(x)}
\]

Wave function (real and positive for any x)

RBM interaction parameter

\[\psi(x)\]

\[X\]

\[\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\]

\[\downarrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\]

Exact

Initial RBM

Optimized RBM

Initial

Optimized

\[W\]

\[\sigma_1\]

\[\sigma_2\]

\[\sigma_3\]

\[\text{Heisenberg Lattice}\]

\[\sigma_4\]
Using artificial neural network to solve quantum many body problems


1D Transverse-field Ising model
80 spins, periodic boundary condition
h: transverse field (h=1: critical)

1D AF Heisenberg model
80 spins, periodic boundary condition

2D AF Heisenberg model
10x10 square lattice
periodic boundary condition

\[ \alpha: \text{hidden variable density} = \frac{\text{(# hidden units)}}{\text{( # physical spins)}} \]

\[ \varepsilon_{\text{rel}} = \frac{E_{\text{NQS}}(\alpha) - E_{\text{exact}}}{|E_{\text{exact}}|} \]
Properties of RBM wave function

D.-L. Deng et al., PRX 7, 021021 (2017); J. Chen et al., PRB 97, 085104 (2018)
Y. Huang and J. E. Moore, arXiv:1701.06246

1. Complex RBM $\rightarrow$ can be applied to general wave function

2. Universal approximator
   $\rightarrow$ k nonzero complex amplitudes $\psi(x)$ can be represented by RBM with k hidden units

3. Short-range RBM
   $\rightarrow$ can be mapped onto entangled plaquette states (EPS)
   $\rightarrow$ area-law entanglement entropy

   \[
   \Psi(\sigma^z) = \prod_{p=1}^{\mu} C_p(\sigma^z_p)
   \]

4. Long-range RBM
   $\rightarrow$ can be mapped onto string-bond state (= product of MPS)
   $\rightarrow$ volume-law entanglement entropy

   \[
   \Psi(\sigma^z) = \prod_i \text{Tr} \left[ \prod_{j \in i} A_{i,j}(\sigma^z_j) \right]
   \]

   \[
   A_{i,j}(\sigma^z_j) = \begin{pmatrix}
   e^{b_i/N+W_{ij}\sigma^z_j} & 0 \\
   0 & e^{-b_i/N-W_{ij}\sigma^z_j}
   \end{pmatrix}
   \]
RBM vs diagonal SBS

RBM wave function

\[ \Psi(\sigma^z) = \prod_i \text{Tr} \left[ \prod_{j \in i} A_{i,j}(\sigma^z_j) \right] \]

\[ A_{i,j}(\sigma^z_j) = \begin{pmatrix} e^{b_i/N + W_{ij}\sigma^z_j} & 0 \\ 0 & e^{-b_i/N - W_{ij}\sigma^z_j} \end{pmatrix} \]

RBM vs diagonal SBS

\[
H_1 = J \sum_{\langle i,j \rangle} S_i \cdot S_j + J \chi \sum_{\langle i,j,k \rangle_\sigma} S_i \cdot (S_j \times S_k)
\]

J=1, J \chi=1, square lattice, 10x10, open boundary

\[ \rightarrow 98\% \text{ overlap with Laughlin state in 4x4 lattice} \]
NetKet: open-source package

https://www.netket.org
How to improve RBM wave function?

1. Combine concepts from machine learning and physics
   
   YN, A. Darmawan, Y. Yamaji, and M. Imada, PRB 96, 205152 (2017)

2. Adding additional hidden layer (deep Boltzmann machine)
   

(3. Extension to fermion-boson coupled Hamiltonians)
How to improve RBM wave function?

1. Combine concepts from machine learning and physics

   YN, A. Darmawan, Y. Yamaji, and M. Imada, PRB 96, 205152 (2017)

2. Adding additional hidden layer (deep Boltzmann machine)


(3. Extension to fermion-boson coupled Hamiltonians)
Restricted Boltzmann machine (RBM) wave function

\[ |\Psi\rangle = \sum_x |x\rangle \mathcal{N}(x) \phi_{\text{product}}(x) \]

Neural-network correlation factor:
\[ \mathcal{N}(x) = \sum_{\{h_k\}} \exp\left(\sum_i a_i \sigma_i + \sum_{i,k} W_{ik} \sigma_i h_k + \sum_k b_k h_k\right) \]

Product state: \( \phi_{\text{product}}(x) = 1 \)

Fermi-sea-based RBM (F-RBM) for fermions
\[ |\Psi\rangle = \sum_x |x\rangle \mathcal{N}(x) \phi_{\text{Fermi-sea}}(x) \]

Neural-network correlation factor can be efficiently calculated because neuron spins are noninteracting
\[ \mathcal{N}(x) \equiv \prod_k 2 \cosh\left(b_k + \sum_i W_{ik} \sigma_i\right) \times e^{\sum_i a_i \sigma_i} \]
RBM+PP wave function
restricted Boltzmann machine + pair-product

YN, A. Darmawan, Y. Yamaji, and M. Imada, PRB 96, 205152 (2017)

Product-basis RBM (P-RBM)

visible layer

hidden layer

visible layer

hidden layer

combine concepts from machine learning (RBM) and physics (pair-product (PP) state)

$$|\Psi\rangle = \sum_x |x\rangle \mathcal{N}(x) \phi_{\text{product}}(x)$$

no entanglement if hidden layer is absent

$$|\Psi\rangle = \sum_x |x\rangle \mathcal{N}(x) \phi_{\text{pair}}(x)$$

Pair-Product state (geminal wave function):

$$|\phi_{\text{pair}}\rangle = \left( \sum_{i,j=1}^{N_{\text{site}}} \sum_{\sigma,\sigma'=\uparrow,\downarrow} f_{i,j}^{\sigma\sigma'} c_{i,\sigma}^{\dagger} c_{j,\sigma'}^{\dagger} \right)^{N_e/2} |0\rangle$$

direct entanglement in visible layer

$\rightarrow$ help RBM to learn ground state
RBM+PP wave function
restricted Boltzmann machine + pair-product

1. RBM+PP wave function

\[ |\Psi\rangle = \sum_x |x\rangle \mathcal{N}(x) \phi_{\text{pair}}(x) \]

cf. many variable VMC wave function

\[ |\Psi\rangle = \sum_x |x\rangle \mathcal{P}_G(x) \mathcal{P}_J(x) \phi_{\text{pair}}(x) \]

Gutzwiller Jastrow

D. Tahara and M. Imada JPSJ 77, 114701 (2008)

2. neural-network correlation factor

\[
\mathcal{N}(x) = \begin{cases} 
\prod_i 2 \cosh \left( b_i + \sum_j W_{i,j} (2S^z_j) \right) & \text{Heisenberg model, } S = \frac{1}{2}, \ 2S^z_j = \pm 1 \\
\prod_i 2 \cosh \left( b_i + \sum_{j,\sigma} W_{i,(j,\sigma)} (2n_{j,\sigma} - 1) \right) & \text{Hubbard model, } 2n_{j,\sigma} - 1 = \pm 1 
\end{cases}
\]

(number of visible variables) = \( N_{\text{site}} \) (Heisenberg)

= \( 2N_{\text{site}} \) (Hubbard)
Ability of RBM to represent Gutzwiller-Jastrow factor

---

**Gutzwiller factor at site i**

$$P_G = \exp(-gn_i^\uparrow n_i^\downarrow)$$

**rewrite**

except for constant factor and one-body potential

$$P_G = \exp\left(-\frac{g}{4}\sigma_{2i}\sigma_{2i-1}\right)$$

$$(\sigma_{2i},\sigma_{2i-1}) = (2n_i^\uparrow - 1, 2n_i^\downarrow - 1)$$

---

**RBM form**

$$P_G = \exp\left(-\frac{g}{4}\sigma_{2i}\sigma_{2i-1}\right)$$

**rewrite**

except for constant factor

$$P_G = \sum_{h=\pm 1} \exp(W_1\sigma_{2i}h + W_2\sigma_{2i-1}h)$$

$$= 2\cosh(W_1\sigma_{2i} + W_2\sigma_{2i-1})$$

$$\begin{cases} 
W_1 = \frac{1}{2}\text{arcosh}(\exp(|g|/2)) \\
W_2 = -\text{sgn} \ g \times W_1 
\end{cases}$$
Result for

2D Heisenberg model

2D Hubbard model

Parameters are numerically optimized following variational principles: \( \langle \Psi_W | \mathcal{H} | \Psi_W \rangle \geq (\text{Ground State Energy}) \)

→ finding set of parameters \( W \) which minimize energy using machine learning techniques

- stochastic reconfiguration method (condensed-matter physics community)
  S. Sorella, PRB 64, 024512 (2001)

- natural gradient (artificial intelligence community)
Application to 2D antiferromagnetic Heisenberg model

8x8 square lattice with periodic boundary condition
YN, A. Darmawan, Y. Yamaji, and M. Imada, PRB 96, 205152 (2017)
tensor network data: L. Wang et al., PRB 83,134421 (2011);

$\alpha = (\# \text{ hidden units})/(\# \text{ physical spins})$

- RBM+PP substantially improves accuracy compared to P-RBM
- Improving reference function helps RBM
Application to 2D Hubbard model

8x8 square lattice, half-filling (periodic anti-periodic)

YN, A. Darmawan, Y. Yamaji, and M. Imada, PRB 96, 205152 (2017)

\[ \alpha = (# \text{ hidden units})/(# \text{ physical spins}) \]

RBM+PP substantially improves accuracy compared to F-RBM

Improving reference function helps RBM
Application to 2D Hubbard model

8x8 square lattice, half-filling (periodic anti-periodic)
YN, A. Darmawan, Y. Yamaji, and M. Imada, PRB 96, 205152 (2017)

RBM+PP result for spin structure factor

<table>
<thead>
<tr>
<th>$S(\pi, \pi)/N_{\text{site}} \times 10^2$</th>
<th>$\alpha = 2$</th>
<th>$\alpha = 8$</th>
<th>$\alpha = 32$</th>
<th>mVMC</th>
<th>AF-QMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$U/t = 4$</td>
<td>3.078(5)</td>
<td>3.057(5)</td>
<td>3.021(5)</td>
<td>3.107(4)</td>
<td>2.92(2)</td>
</tr>
<tr>
<td>$U/t = 8$</td>
<td>5.233(9)</td>
<td>5.206(9)</td>
<td>5.20(1)</td>
<td>5.30(1)</td>
<td>5.0(1)</td>
</tr>
</tbody>
</table>

U dependence of RBM+PP energy ($\alpha=32$)

RBM+PP works better in larger U, in contrast to mVMC
Discussion

Representability of RBM

\[
N'(x) = \begin{cases} 
\prod_i 2 \cosh \left( b_i + \sum_j W_{i,j} (2S_j^z) \right) & \text{(Heisenberg model, } S = \frac{1}{2}, \ 2S_j^z = \pm 1) \\
\prod_i 2 \cosh \left( b_i + \sum_{j,\sigma} W_{i,j\sigma} (2n_{j\sigma} - 1) \right) & \text{(Hubbard model, } 2n_{j\sigma} - 1 = \pm 1) 
\end{cases}
\]

\( N(x) \) with real variables can represent any probability distribution over \( x \) with infinite \( \alpha \)


Representability of RBM+PP for ground-state wave functions

- In bipartite Heisenberg model, a gauge transformation makes wave function positive definite
  \( \rightarrow \) relative error should go to zero as \( \alpha \) goes to infinity

- In Hubbard model, nodal structure of wave function is important
  \( \rightarrow \) PP wave function needs to take account of nodes
  Introduction of complex RBM helps to improve nodal structure?

---

![Heisenberg and Hubbard diagrams](image-url)
Short summary

**Summary of RBM+PP**

- outperforms the original RBM and mVMC method
- flexible applicability (both to bosons and fermions)
- No negative sign
  - can be applied to e.g. frustrated spin systems and doped Hubbard model, where QMC is not applicable

RBM+PP: a new powerful solver for strongly correlated quantum systems

YN, A. Darmawan, Y. Yamaji, and M. Imada, PRB 96, 205152 (2017)

**Perspectives**

- Application to other systems
- Introduction of another hidden layer (deep Boltzmann machine (DBM))

→ next part : exact DBM constructions to represent ground states of many-body Hamiltonians

G. Carleo, Y. Nomura, and M. Imada, arXiv:1802.09558 (see also N. Freitas et al., arXiv:1803.02118)
How to improve RBM wave function?

1. Combine concepts from machine learning and physics

   YN, A. Darmawan, Y. Yamaji, and M. Imada, PRB 96, 205152 (2017)

2. Adding additional hidden layer (deep Boltzmann machine)


(3. Extension to fermion-boson coupled Hamiltonians)
DBM (deep Boltzmann machine) wave function

\[ \Psi(\sigma) = \sum_{h,d} e^{\sum_i a_i \sigma_i^z + \sum_{i,j} \sigma_i^z W_{ij} h_j + \sum_j b_j h_j + \sum_{j,k} h_j W'_{jk} d_k + \sum_k b'_k d_k} \]
DBM representation of ground states

**DBM compared with RBM**

- **Pros**
  - much more flexible representability


- **Cons**
  - cannot trace out both h and d analytically
    (need to sample hidden spins to obtain wave function)

**Key idea**

- reproduce imaginary-time evolution by dynamically modifying DBM network
  (no need to perform stochastic optimization of parameters! everything deterministic!)

\[
|\Psi(\tau)\rangle = e^{-\mathcal{H}_1 \frac{\delta \tau}{2}} e^{-\mathcal{H}_2 \delta \tau} \ldots e^{-\mathcal{H}_2 \delta \tau} e^{-\mathcal{H}_1 \frac{\delta \tau}{2}} |\Psi_0\rangle
\]

- Physical quantities are measured by MC sampling of classical visible and hidden spins

**Novel class of quantum-to-classical mapping**

G. Carleo, Y. Nomura, and M. Imada, arXiv:1802.09558 (see also N. Freitas et al., arXiv:1803.02118)
Example: Transverse-Field Ising model

G. Carleo, YN, and M. Imada, arXiv:1802.09558

Hamiltonian: \[ \mathcal{H} = \mathcal{H}_1 + \mathcal{H}_2 \]

Interaction (classical): \[ \mathcal{H}_1 = \sum_{l<m} V_{lm} \sigma_l^z \sigma_m^z \]

Transverse-field: \[ \mathcal{H}_2 = -\sum_l \Gamma_l \sigma_l^x \]

How to express short time propagators by DBM?

Interaction propagator: \[ e^{-\delta \tau V_{lm} \sigma_l^z \sigma_m^z} |\text{DBM}\rangle \]

Transverse-field propagator: \[ e^{\delta \tau \Gamma_l \sigma_l^x} |\text{DBM}\rangle \]
Interaction propagator (diagonal)

\[ e^{-\delta \tau V_{lm} \sigma_i \sigma_m} |DBM\rangle \]

Initial network (arbitrary)
Interaction propagator (diagonal)

\[ e^{-\delta \tau V_{lm} \sigma_i^z \sigma_m^z} |DBM\rangle \]

Network after the propagator with new h and W

\[
W_{t[lm]} = \frac{1}{2} \text{arcosh} \left( e^{2|V_{lm}|\delta \tau} \right)
\]

\[
W_{m[lm]} = -\text{sgn}(V_{lm}) \times W_{t[lm]}
\]

RBM architecture is enough to represent classical interaction
Transverse-field propagator (off-diagonal)

\[ e^{\delta \tau \Gamma_i \sigma_i^x} |DBM\rangle \]
Transverse-field propagator (off-diagonal)

\[ e^{\delta \tau \Gamma_i \sigma_i^x} |DBM\rangle \]

(Intermediate step) new \( d \) and \( W' \)

\[ W'_{j[u]} = -W_{lj} \]

Deep layer makes it possible to derive analytical expression for quantum propagator
Transverse-field propagator (off-diagonal)

\[ e^{\delta \tau \Gamma_i \sigma_i^x} |DBM\rangle \]

(Intermediated step) modify \( W \)

\[ \bar{W}_{ij} = W_{ij} + \Delta W_{ij} = 0 \]
Transverse-field propagator (off-diagonal)

\[ e^{\delta \tau \Gamma_l \sigma_i^x} |DBM\rangle \]

new h, W, W' and obtain new network

\[ W_{l[u]} = \frac{1}{2} \text{arcosh} \left( \frac{1}{\tanh(\Gamma_l \delta \tau)} \right) \]

\[ W'_{[l][l]} = -W_{l[u]} \]
DBM construction for Heisenberg model

G. Carleo, YN, and M. Imada, arXiv:1802.09558

Initial State

Step 0
Initial State

New $d$ and $W$

Step 1
New $d$ and $W'$

Step 2
Modify $W$

Step 3
New $h, W, W'$ (real)

Step 4
New $h, W, W'$ (constraint)

Network after time evolution

$e^{-\delta T \sum_i \sigma_i \sigma_m} |DBM\rangle$
In this way, we can follow imaginary-time Hamiltonian evolution exactly within DBM framework

(# hidden units) $\propto$ (system size) $\times$ (imaginary time)
Numerical result (1)

G. Carleo, YN, and M. Imada, arXiv:1802.09558

1D Transverse-Field Ising

$N = 20$, $J \delta \tau = 0.01$

DBM reproduces exact time-evolution

1D Antiferromagnetic Heisenberg

$N = 80$, $J \delta \tau = 0.01$

from empty network
from pre-optimized RBM

better initial state => faster convergence
Discussion

DBM Representation Not Unique

In Heisenberg model, we have found (at least) 3 representations with different topology

1d3h => local W, non-local W'
2d6h => local W, local W' (equivalent to path-integral when h spins are traced out)
2d4h => non-local W, non-local W'

Possible sign problem

No sign problem for bipartite spin models

For frustrated system, negative signs appear (general W, W' become complex)

different representation give different amount of negative signs ?
Starting from pre-optimized state, we can reach ground state before negative signs become severe?
Numerical result (2)

2D $J_1$-$J_2$ Antiferromagnetic Heisenberg model

$N = 4 \times 4 = 16$, $J_2 \tau = 0.001$

$$|\Psi\rangle = \left[ \prod_i \exp(-H_i \delta \tau) \right]^{N_{\text{slice}}} |\text{pair-product (RVB)}\rangle$$

symbols: DBM
solid curves: exact

$J_2/J_1 = 0.4$
$J_2/J_1 = 0.0$
Summary

Show deterministic construction of DBM to represent ground states
   The number of hidden units grows linearly with system size and imaginary time, respectively

Additional hidden (deep) layer: “additional dimension” in statistical mechanics

DBM representation => New quantum-to-classical mapping

G. Carleo, Y. Nomura, and M. Imada, arXiv:1802.09558 (see also N. Freitas et al., arXiv:1803.02118)

Perspective

Application to Frustrated Spin Systems
   How does the negative-sign rate differ in 3 representations?

How to compress the network?
   An approximate mapping from DBM to RBM?
Future directions

- Calculations of Excited States
- Finite Temperature Calculations
- Dynamics
- Mutual understanding between Tensor and Neural networks