Plaquette Renormalized Tensor Network States: Application to Frustrated Systems

Ying-Jer Kao
高英哲
Department of Physics and Center for Quantum Science and Engineering National Taiwan University

Hsin-Chih Hsiao, Ji-Feng Yu (NTU)
Anders W. Sandvik, Ling Wang (Boston)
Frustration

- Large number of degenerate classical ground states
- Emergence of novel spin-disordered ground states due to quantum fluctuations
- Hard to study numerically: size limitation in Exact Diagonalization, sign problem in QMC (also fermion), dimension limit for DMRG (d=1)
Negative Sign Problem

\[ H = \sum_{\langle ij \rangle} H_{ij} = \sum_{\langle ij \rangle} (S_i^+ S_j^- + S_i^- S_j^+) \]

\[ Z = Tr \left[ e^{-\beta H} \right] = \sum_{\alpha} \sum_{n=0}^{\infty} \frac{(-\beta)^n}{n!} \langle \alpha | H^n | \alpha \rangle \]

\[ p(c) = (-1)^3 \langle \alpha | H_{12} H_{23} H_{31} | \alpha \rangle < 0 \]

\[ \langle A \rangle_{p(c)} = \frac{\langle A_s \rangle_{|p(c)|}}{\langle s \rangle_{|p(c)|}} \]

- Negative matrix elements
- Fluctuations in signs: ensemble average diverges at low temperature
Simulation of Quantum Systems

• Is it possible to classically simulate faithfully a quantum system?

• To represent a quantum state:

\[ |\psi\rangle = \sum_{i_1, i_2, \ldots, i_n = 1}^{d} c_{i_1 i_2 \ldots i_n} |i_1 i_2 \ldots i_n\rangle \]

classical representation requires \(d^n\) complex coefficients which grows exponentially with \(n\)

• Ground states of quantum many-body systems are usually entangled.
Entanglement

Complementary viewpoints on entanglement:

- Quantum information theory: a crucial resource to process and send information in novel ways
- Quantum many-body physics: entanglement gives rise to exotic phases of matter
- Numerical simulation of strongly correlated quantum systems: source of difficulties!!

What kind of superpositions appear in nature? symmetries, local interactions, little entanglement
Entanglement: Matrix Product State

- Between each nearest-neighbor lattice site, we introduce a $D$-dimensional maximally entangled state: $|\phi> = \frac{1}{\sqrt{D}} \sum_{\alpha=1}^{D} |\alpha\alpha>$
- At each lattice site there are two ends to these maximally entangled states
- Operator $A$: project $D^2$ to $d$ (AKLT GS: $D=2$, $d=3$)
  
  $$|\psi> = \sum_{i_1,\ldots,i_N=1}^{d} A_1(i_1) A_2(i_2) \cdots A_N(i_N) |i_1, i_2, \ldots, i_N>$$
- We reexpress the $2^N$ coefficients of in terms of about $2D^2 N$ parameters (linear in $N$)
- Key ingredient behind the success of DMRG

**Tensor product states**

- Extension similar ideas to higher dimension:
  - tensor product states, projected entangled paired state (PEPS)

\[
|\psi\rangle = \sum_{\{s\}} t \text{Tr}\{A(s_1)A(s_2) \cdots A(s_N)\}|s_1, s_2, \ldots, s_N\rangle
\]

\[
s = \pm 1
\]

\[
A_{ijkl}^{s_k}
\]

\[
l = 1 \ldots D
\]

\[
A_{ijkl}: \text{rank-4 tensor}
\]

F. Verstraete and J. I. Cirac, arXiv:cond-mat/0407066
Variational wave function

- Use tensor product state as trial wave function

\[ E = \frac{\langle \psi | H | \psi \rangle}{\langle \psi | \psi \rangle} \geq E_0 \]

\[ H = \sum_i H_i \]

\[ H_i = \hat{O}_i^0 + \hat{O}_i^1 \hat{O}_j^2 + \ldots \]

\[ \langle \psi | \psi \rangle = \sum_{s_1, s_2, \ldots} \sum_{i, j, k, l, \ldots} A_{i', j', k', l'}^{s_1*} A_{i, j, k, l}^{s_1} A_{k', m', n', o'}^{s_2*} A_{k, m, n, o}^{s_1} \cdots \]

\[ = t \text{Tr} [T \otimes T \otimes T \otimes \ldots] \]

- Double tensor

\[ T = \sum_{s, s'} A^{s'}* \otimes A^s \]
**Variational Energy**

\[ H = \sum_i H_i \]

\[ H_i = \hat{O}_i^0 + \hat{O}_i^1 \hat{O}_j^2 + \ldots \]

\[ \langle \psi | H | \psi \rangle = t \text{Tr} \left[ T_i^0 \otimes T \otimes T \otimes \ldots \right] + t \text{Tr} \left[ T_i^1 \otimes T_j^2 \otimes T \otimes \ldots \right] + \ldots \]

\[ T^a = \sum_{s,s'} A^{s'*} \otimes A^s \langle s' | \hat{O}^a | s \rangle \]

- **Computationally intensive**: \( D^{N_{\text{bond}}} \)
- **Direct computation of contraction is impossible**: approximation (RG schemes)
Tensor Contraction

- Contracting the internal indices, the four-leg tensor can be viewed as a single tensor.

- External link dimension becomes $D^2$ after one contraction; exponential growth as we keep contracting $D^4, D^8, \ldots$.

- Computationally intensive; Impossible to store intermediate results.

- Need some RG scheme.
Tensor entanglement renormalization

- Singular Value Decomposition: rank-3 tensors $S$

\[
T_{\alpha\beta\mu\nu} \approx \sum_{\gamma' = 1}^{D_{\text{cut}}} S_{\mu\nu\gamma'}^1 S_{\alpha\beta\gamma'}^3
\]

\[
T_{\alpha\beta\mu\nu} \approx \sum_{\gamma' = 1}^{D_{\text{cut}}} S_{\alpha\nu\gamma'}^2 S_{\mu\beta\gamma'}^4
\]

\[
T'_{\gamma\sigma\lambda\rho} \approx \sum_{\alpha\beta\mu\nu} S_{\beta\alpha\gamma}^1 S_{\mu\beta\sigma}^2 S_{\nu\mu\gamma}^3 S_{\alpha\nu\rho}^4
\]

- Introduce cut-off for the bond-contraction

Tranverse ising model

\[ H = - \sum_{\langle ij \rangle} \sigma_i^z \sigma_j^z - h \sum_i \sigma_i^x \]

- $D=2$, $D_{\text{cut}}=18$, non-MF results is obtained
- Method is not variational
  - Locally optimized
  - SVD truncation
- Translational invariance explicitly broken

Distribution of Singular Values

\[ T_{\alpha \beta \mu \nu} = M_{\alpha \beta, \mu \nu} = \sum_{\gamma} U_{\alpha \beta, \gamma} \lambda_{\gamma} V^{T}_{\gamma, \mu \nu} \quad \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \]

Transverse Ising Model, 8x8, h=3.08

Fully Frustrated XX model, 8x8, h=1

Frustrated Systems Remain Difficult
Renormalization of an 8-index plaquette tensor using auxiliary 3-index tensors $S$. 

8-index tensor: $D^8$  
4-index tensor: $D_{\text{cut}}^4$  
8-index dbl-tensor: $D_{\text{cut}}^{16}$  
4-index dbl-tensor: $D_{\text{cut}}^8$
Plaquette renormalized trial wave function

\[ |\psi\rangle = \sum_{\{s\}, i, j, k, l, ...} (A_1^{s_1})_{i, j, k, l} (A_2^{s_2})_{k, m, n, o} (S_1)_{m, j, p} (S_2)_{q, p, r} \cdots |s_1 s_2 \cdots s_N\rangle \]

\[ = \sum_{\{s\}} tTr(A_1^{s_1} \otimes A_2^{s_2} \otimes S_1 \otimes S_2 \otimes \cdots \otimes A_N^{s_N}) |s_1 s_2 \cdots s_N\rangle \]

- We treat the elements in the A and S tensors as variational parameters, and minimize the total energy.
- Principal axis method: derivative-free, but computationally expensive.
**Expectation values**

\[
H = \sum_{i} \hat{O}_{i}^{0} + \sum_{\langle i,j \rangle} \hat{O}_{i}^{1} \hat{O}_{j}^{2} + \sum_{\langle \langle i,j \rangle \rangle} \hat{O}_{i}^{1} \hat{O}_{j}^{2} + \ldots
\]

Normalization factor:

\[
\langle \psi \mid \psi \rangle = t \text{Tr} \left[ T_{1} \otimes T_{2} \otimes R_{1} \otimes \cdots \otimes T_{N} \right]
\]

Two-body interaction:

\[
\langle \psi \mid \hat{O}_{i}^{1} \hat{O}_{j}^{2} \mid \psi \rangle = t \text{Tr} \left[ T_{1} \otimes T_{2} \otimes R_{1} \otimes \cdots \otimes T_{i}^{1} \otimes T_{j}^{2} \otimes \cdots \otimes T_{N} \right]
\]

\[
T_{j} = \sum_{s_{j}} A_{j}^{s_{j}} \otimes A_{j}^{s_{j}}
\]

\[
T_{j}^{a} = \sum_{s_{j}' s_{j}} A_{j}^{s_{j}'} \otimes A_{j}^{s_{j}} \langle s_{j}' \mid \hat{O}_{j}^{a} \mid s_{j} \rangle
\]

\[
R_{i} = S_{i}^{*} \otimes S_{i}
\]
Plaquette-Renormalization of TNS

- Effective reduced tensor network for a 8x8 lattice
- Summing over all unequivalent bonds and sites
- Method is variational
- Optimize $T$ and $R$ globally
- Size scaling: $L^2 \log(L)$
Transverse Ising Model

- Assume translational invariance: all initial T’s are the same.
- Globally optimized T and R.
- $h_c = 3.33$ (3.04, QMC)
- $m_z \sim (h-h_c)\beta$, $\beta \sim 0.40$
- $h \sim h_c$, $\beta \sim 0.50$ mean-field like. (Sanvik’s talk).

<table>
<thead>
<tr>
<th>$L$</th>
<th>$h$</th>
<th>$D$</th>
<th>$E_{\text{var}}/N$</th>
<th>$E/N$ (exact)</th>
<th>$\Delta E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>3.0</td>
<td>2</td>
<td>-3.1978372</td>
<td>-3.2155081</td>
<td>$5.4955 \times 10^{-3}$</td>
</tr>
<tr>
<td>8</td>
<td>3.0</td>
<td>2</td>
<td>-3.1717845</td>
<td>-3.19750(QMC)</td>
<td>$8.0437 \times 10^{-3}$</td>
</tr>
</tbody>
</table>
Transverse Ising Model

- Spins at different lattice sites inside the plaquette have different environments.
- We use different tensors inside a plaquette.

<table>
<thead>
<tr>
<th>$L$</th>
<th>$h$</th>
<th>$D$</th>
<th>$E_{\text{var}}/N$</th>
<th>$E/N$</th>
<th>$\Delta E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>3.0</td>
<td>2</td>
<td>$-3.1978372$</td>
<td>$-3.2155081$ (exact)</td>
<td>$5.4955 \times 10^{-3}$</td>
</tr>
<tr>
<td>8</td>
<td>3.0</td>
<td>2</td>
<td>$-3.1717845$</td>
<td>$-3.19750$ (QMC)</td>
<td>$8.0437 \times 10^{-3}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$L$</th>
<th>$h$</th>
<th>$D$</th>
<th>$E_{\text{var}}/N$</th>
<th>$E/N$</th>
<th>$\Delta E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>3.0</td>
<td>2</td>
<td>$-3.2044358$</td>
<td>$-3.2155081$ (exact)</td>
<td>$3.4434 \times 10^{-3}$</td>
</tr>
<tr>
<td>4</td>
<td>3.0</td>
<td>3</td>
<td>$-3.2152333$</td>
<td>$-3.2155081$ (exact)</td>
<td>$8.546 \times 10^{-5}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$L$</th>
<th>$h$</th>
<th>$D$</th>
<th>$E_{\text{var}}/N$</th>
<th>$E/N$</th>
<th>$\Delta E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>3.0</td>
<td>2</td>
<td>$-3.17712$</td>
<td>$-3.19750$ (QMC)</td>
<td>$6.3737 \times 10^{-3}$</td>
</tr>
<tr>
<td>8</td>
<td>3.044</td>
<td>2</td>
<td>$-3.21404$</td>
<td>$-3.23627$ (QMC)</td>
<td>$6.869 \times 10^{-3}$</td>
</tr>
</tbody>
</table>
$J_1$-$J_2$ Heisenberg Model

\[ H = J_1 \sum_{\langle ij \rangle} \mathbf{S}_i \cdot \mathbf{S}_j + J_2 \sum_{\langle \langle ij \rangle \rangle} \mathbf{S}_i \cdot \mathbf{S}_j \]

\[ M^2 = \langle \frac{1}{N} \sum_j f(j) \tilde{S}_j \rangle \]

\[ f(j) = \exp[i \tilde{Q} \cdot \tilde{R}_j], \quad \tilde{R}_j = (x_j, y_j) \]

$M_1^2$ at $\tilde{Q} = (\pi, \pi)$

$M_2^2$ at $\tilde{Q} = (\pi, 0)$ or $\tilde{Q} = (0, \pi)$

Error in energy

- Improvement in accuracy as $D$ increase.
- $D$ also sets a length scale.

<table>
<thead>
<tr>
<th>$L$</th>
<th>$D$</th>
<th>$E_{\text{var}}/N$</th>
<th>$E/N$ (exact/ QMC)</th>
<th>$\Delta E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>2</td>
<td>-0.649491</td>
<td>-0.701781 (exact)</td>
<td>$7.451 \times 10^{-2}$</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>-0.633135</td>
<td>-0.673487 (QMC)</td>
<td>$5.991 \times 10^{-2}$</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>-0.630080</td>
<td>-0.669976 (QMC)</td>
<td>$5.954 \times 10^{-2}$</td>
</tr>
</tbody>
</table>

- $J_2 = 0$
Building block: plaquette

- 12 internal sums. Maximum computational effort: $D^8$.

- Each free index and summation contributes $D$.

$$T^{(1),\alpha\beta}_{\gamma\delta} = \sum_{\alpha_1,\alpha_2,\ldots,\delta_1,\delta_2,ijkl} S^{(1),\alpha}_{\alpha_1\alpha_2} S^{(1),\beta}_{\beta_1\beta_2} S^{(1),\gamma}_{\gamma_1\gamma_2} S^{(1),\delta}_{\delta_1\delta_2} T^{(0),\alpha_1}_{\gamma_2i} T^{(0),\alpha_2}_{i\delta_1} T^{(0),\beta_1}_{j\delta_2} T^{(0),\beta_2}_{\gamma_1k}$$
**Computational Costs**

- Globally optimized $T$ and $R$.
- Contraction calculation is highly parallelizable.
- IBM Blue Gene/L at BU. $D=2$, takes weeks to optimize.
- Bottleneck: plaquette internal contractions.
- Take advantage of GPU (Graphic Processing Unit).

Wang, Kao, Sandvik, arXiv:0901.0214
GPU supercomputing

• The GPU is specialized for compute-intensive, massively data parallel computation

nVidia Tesla 10-Series Processor

NTU CQSE GPU cluster
16x Tesla S1070 = 64 Teraflops
GPU kernel Strategy

- **Inputs:** \( T \) (\( D^4 \)), \( S \) (\( D^3 \)) (unchanged during contraction)
  - Load into texture memory (optimized for 2D read)

- **Output:** \( T' \) (\( D^4 \))

- **Thread block geometry up to 3D** (\( t_x,t_y,t_z \)), grid of blocks geometry up to 2D (\( b_x,b_y \)).
  - Reshape \( T \) into a \( D^2 \times D^2 \) matrix, \( S \) into a \( D \times D^2 \) matrix.
  - Use \( D^2 \times D^2 \) thread blocks, each block has \( D \times D \) threads. Each block computes an element of \( T' \). Use block indices \((b_x,b_y)\) to represent external indices.
Benchmark: One plaquette contraction

- CPU code: Intel Core 2 Duo E4700 2.6GHz
  - D=6, 2.1 GFLOPS
  - D=8, 3.4 GFLOPS
- GPU, Single Precision code (unoptimized):
  - D=6, 24.9 GFLOPS (GTX8800), 49.7 GFLOPS (GTX280)
  - D=8, 92.2 GFLOPS (GTX280)
- 10x to 30x speedup. A factor of at least 8 performance hit for DP.
- GPU code requires optimization
  - Bank conflict in shared memory R/W
  - Low stream multiprocessor occupancy
  - Current scheme D bound by shared memory size
  - Global memory bandwidth hiding
Conclusion

- Tensor network states are promising candidates to understand frustrated quantum spin systems.
- In plaquette renormalized tensor network representation, no approximations are made when contracting the effective renormalized tensor network.
- Non-MFT results even with the smallest possible non-trivial tensors and truncation ($D = 2$)
- Larger internal bond dimension $D$ is necessary to get the right physics.
- GPU can potentially speed up the computationally intensive part of the calculation.
- Use Monte Carlo sampling in the future.