Markov Chains for Tensor Network States

Sofyan Iblisdir
University of Barcelona

September 2013
arXiv 1309.4880
Main issue discussed here: how to find good tensor network states?

- The optimal state problem
- Markov chains
- Multiplicative approximations
- *Examples* with MPS and PEPS
- Discussion and possible extensions (low and high energy eigenstates, 2D time evolution, finite temperature).
Tensor network states seem to accurately describe a wide variety of strongly correlated systems.

Profound reason: Nature doesn’t saturate entanglement; it makes sense to try and represent quantum states with states where entanglement follows well defined patterns.

- Gapped 1D systems: matrix product states (DMRG).
- Scale invariant 1D systems: entanglement renormalisation
- Two-dimensional systems: projected entangled pair states

Systematic investigation of condensed matter systems with such states has proven hugely successful over the past 20 years. For instance:

- Progress on specific models (e.g. Kagomé-Heisenberg, $J_1 - J_2$)
- Better understanding of why some systems are so hard to study.
We are interested in an issue that can be informally stated as follows:

Given a Hamiltonian $H$ (in one or in two dimensions), and a family $T$ of tensor network states, find the best approximation $\phi^* \in T$ to some ground state $\psi^*$ of $H$.

Existing methods: DMRG (most popular), Euclidean evolution, site-by-site tensor improvement.

In practice, these methods have proven very powerful; what has been achieved cannot be over-rated. It would however be wrong to consider the optimal state problem as "closed", let alone "solved".
1D. Despite its success, it is generally impossible to actually prove that DMRG indeed outputs a good solution; we can never be quite sure we haven’t been led to a local minimum. Worse, one can construct situations where DMRG is known to get stuck in sub-optimal solutions. The situation is the same with other methods.

The optimal state problem is NP-hard\(^1\).

The case of uniformly gapped quantum Hamiltonians seems to be more tractable\(^2\); \(\exists\) a poly-time certifiable algorithm. But the method requires prior knowledge about the gap and is tedious to implement.

The situation seems even worse in two dimensions. What do we know exactly?

The relevance of the problem is not only complexity-theoretic. See for example the MERA-DMRG comparisons about the Heisenberg model on a Kagomé lattice\(^3\):

\[
E_{DMRG}^{old}(\text{spin liquid}) > E_{MERA}(\text{not spin liquid}) > E_{DMRG}^{new}(\text{spin liquid}).
\]

One possible strategy to make progress is to enlarge the set of methods available to search optimal states. Certainly, this will to new difficulties. But hopefully, some insight can be gained from them.

---

\(^1\)N. Schuch, I. Cirac, F. Verstrate, PRL 100, 250501 (2008)

\(^2\)Z. Landau, U. Vazirani, T. Vidick, pre-print arXiv:1307.5143 and ref. therein

\(^3\)S. Depenbrock et al. PRL 109, 067201 (2012)
We have considered the use of Markov chains concatenated into a simulated annealing algorithm.

- Connects the optimal state problem to sampling. Perhaps it will allow us to benefit from decades of efforts and abundant literature.
- Gain new perspectives on our problem. Presumably, the issue of certifying a solution is a global optimum can be related to that of bounding mixing times (convergence) for Markov chains.
- When optimality cannot be guaranteed, it is reasonable to try and get answer through independent ways and compare; such weak forms of verification are commonplace in hard combinatorial problems such as the 3-SAT.
- Seems flexible, might well allow for extensions to other problems: high energy eigenstates, finite temperature, time evolution.
- Relatively easy to implement: no need to compute environments in 2D for instance.

N.B. ∃ earlier work on Monte-Carlo for tensor networks, but for a wholly different purpose: approximate contraction scheme; optimisation was still ”greedy”. Here standard contraction will do, but different search method, where moves to worse configurations are allowed.

---

4
dMetropolis algorithm has turned 60 this year!
Consider a chain made of $n$ particles interacting via $H$ (we assume open boundary conditions). Assume we want the best approximation of the ground state in terms of an MPS with fixed bond dimension $\chi$:

$$|\psi\rangle = \sum_{s_1 \ldots s_n} \text{tr}[A_1(s_1) \ldots A_n(s_n)]|s_1 \ldots s_n\rangle,$$

(1)

Our first task will be to define a convenient configuration space. Consider the map $\varphi : U(d\chi)^\otimes n \to \text{Hilb} : \omega \to |\omega\rangle$:

We choose the configuration space to be $\Omega \equiv U(d\chi)^\otimes n$. 

Figure 1: Possible preparation of a matrix product state (open boundary conditions) of $nd$-dimensional quantum particles. A $\chi$-dimensional ancilla is prepared in a reference state $|1\chi\rangle$, and interacts sequentially with each of the considered particle, also prepared in a reference state $|1d\rangle$). The ancilla is eventually projected on $|1\chi\rangle$ and disregarded.

Figure 2: Schematic illustration of a Markov-Haar move. In this example, where $d=2$ and $\chi=1$, the configuration space is $U(2)^\otimes n$ and can be represented as a set of $n$ points on a $2^{-}$-sphere. The plain (resp. dotted) arrow represents the current (resp. candidate) configuration.
Let \( h(\omega) = \langle \omega | H | \omega \rangle / \langle \omega | \omega \rangle \): energy associated with \( \omega \). Sampling according to

\[
\pi^{(\beta)}(\omega) \equiv \frac{e^{-\beta h(\omega)}}{\int d\omega' e^{-\beta h(\omega')}}, \ \omega \in \Omega, \tag{2}
\]

would be an invaluable resource. However, except for a few trivial cases, direct sampling of \( \pi^{(\beta)} \) seems to be impossible. But a Markov chain can be constructed:

1. Draw \( n \) i.i.d. unitary matrices \( \{ v_k, k = 1 \ldots n \} \in U(d \chi)^{\otimes n} \) according to the Haar measure\(^5\).
2. From the current configuration \( \omega = \{ u_k, k = 1 \ldots n \} \), construct a candidate \( \omega' = \{ u_k v_k^\alpha : k = 1 \ldots n \}, 0 < \alpha < 1 \).
3. Accept the proposed move \( \omega \to \omega' \) with (Metropolis) probability

\[
P^{(\beta)}_{acc} (\omega \to \omega') = \min\{1, e^{-\beta (h(\omega') - h(\omega))} \}. \tag{3}
\]

\(^5\)F. Mezzadri, Not. Amer. Math. Soc. 54, 592 (2007).
1. Draw \( \{ v_k, k = 1 \ldots n \} \in U(d\chi) \otimes n \sim \text{Haar measure} \).

2. \( \omega' = \{ u_k v_k^\alpha : k = 1 \ldots n \}, \ 0 < \alpha < 1 \).

3. \( P_{\text{acc}}^{(\beta)}(\omega \rightarrow \omega') = \min\{1, e^{-\beta(h(\omega') - h(\omega))}\} \).

A picture:

\[
\begin{align*}
(1) & \quad (2) & \quad (n) \\
\hspace{0.5cm} & \quad \quad & \\
\hspace{1cm} & \quad \quad & \\
\end{align*}
\]

- Implementation: draw random numbers and unitaries, evaluate \( h(\omega) \).
- \( \alpha \) should be tuned to achieve good trade-off between modest and ambitious moves.
- Strongly ergodic: \( P_{\text{trans}}(\omega \rightarrow \omega') > 0 \ \forall \omega, \omega' \).
- Reversible: \( \pi^{(\beta)}(\omega) P_{\text{trans}}(\omega \rightarrow \omega') = \pi^{(\beta)}(\omega') P_{\text{trans}}(\omega' \rightarrow \omega) \).
That the Markov chain is defined on a (connected) compact space, strongly ergodic and reversible implies that $P_{trans}$ is a contraction:

$$\|P_{\tau}^{\text{trans}} \pi_0 - \pi^{(\beta)}\|_{TV} \leq 2 \eta(\beta)^\tau, \quad \eta(\beta) < 1.$$ 

As $\beta \to \infty$, the distribution

$$\pi^{(\beta)}(\omega) \equiv \frac{e^{-\beta h(\omega)}}{\int d\omega' e^{-\beta h(\omega')}} , \quad \omega \in \Omega,$$

(4)

exponentially concentrates on minimal energy states. So if $\beta$ is increased sufficiently slowly with time, we get a simulating annealing scheme, converging towards a global optimum. We don’t know how slow is slow enough; finding a useful bound on $\eta(\beta)$ is a difficult problem.
What should we reasonably expect? Asking for, say, 14-digit precision is probably too much. Following requirements common in Monte-Carlo literature, a simulation will be considered efficient if it yields a multiplicative approximation in polynomial time. Let $e_0$ denote the true g.s. energy. We will be content if the Metropolis algorithm yields an estimate $e'_0$ satisfying

$$\text{Proba}[|e'_0 - e_0| < \delta |e_0|] > 2/3,$$  

in a time that grows polynomially with $n, \chi, \delta^{-1}$.

Remark: Computing the partition function of the ferromagnetic Ising model (with a magnetic field) is known to be #P-complete. Jerrum and Sinclair\textsuperscript{6} have however been able to construct an efficient multiplicative for the problem, through a clever design of Markov chains. It would be interesting to exhibit quantum analogues of this situation.

\textsuperscript{6}M. Jerrum, A. Sinclair, SIAM J. Comput. 22 1087 (1993).
A case study: the 1D Ising model

\[ H = -\sum_{k=1}^{n} \sigma_k^z - \sum_{k=1}^{n-1} \sigma_k^x \sigma_{k+1}^x. \]  

(6)

A linear schedule has been assumed: \( \beta(t) = \beta_0 + \kappa t. \)

| \( \chi \) | HM (\( \beta < \infty \)) | HM (\( \beta = \infty \)) | EE |
|---|---|---|---|
| 2  | -50.354476 | -50.503292 | -50.503303 |
| 4  | -50.489549 | -50.511711 | -50.567694 |
| 8  | -50.338708 | -50.514589 | -50.569413 |
| 16 | -50.354480 | -50.502324 | -50.569425 |

All computations made on a single core desktop computer. Computations stopped at \( \sim 10^{-3} \) relative accuracy. Data not ”converged”, better figures if longer times are used. The longest computation took \( \sim 48 \) hours, could be reduced down to \( \sim 3 \) hours (double complex \( \rightarrow \) single precision real). At this stage, we shouldn’t aim at a comparison with prior methods (would be too unfair!):

- Many improvements possible (some of them very mundane). We make no claim of optimality; a lot of room for improvement.
- Usual methods are particularly efficient for this Hamiltonian.
- These first results are encouraging.
The scheme can be extended to PEPS ($n_x \times n_y$ square lattice). Again a unitary configuration space can be associated to the problem:

$$\Omega \equiv U(d\chi^3)^{n_x n_y}$$

For any search scheme, there is an important additional difficulty when moving from MPS to PEPS; accuracy when estimating mean values. Let $h(\omega), \delta_h(\omega)$ denote an estimate for $\langle \omega | H | \omega \rangle$ together with a bound on the error. Similarly, let $n(\omega), \delta_n(\omega)$ denote an estimate for $\langle \omega | \omega \rangle$ together with a bound on the error.

Definition: a PEPS is **pointless** if $|h(\omega)| \leq \delta_h(\omega), f n(\omega) \leq \delta_n(\omega)$ or both. A PEPS is **contractible** if it is not pointless.

The Monte-Carlo scheme can be designed so as to avoid pointless PEPS; stabilises computations. We consider the distribution:

$$\pi^{(\beta)}(\omega) \equiv \frac{e^{-\beta g(\omega)}}{\int d\omega' e^{-\beta g(\omega')}} , \omega \in \Omega,$$

$$g(\omega) = \begin{cases} +\infty & \text{if } \omega \text{ is pointless}, \\ \max_{a,b \in \{0,1\}} \left\{ \frac{h(\omega)+(-)^a \delta_h(\omega)}{n(\omega)+(-)^b \delta_n(\omega)} \right\} & \text{else}. \end{cases}$$
We have tested the method with the Ising model, defined on a $7 \times 7$ square lattice, with parameters $\chi = 2, \chi_c = 4$.

\[
H = -J_x \sum_{\langle k,l \rangle} \sigma^x_k \sigma^x_l - h_z \sum_k \sigma^z_k, \tag{7}
\]

\[
\begin{array}{cccc}
J_x & h_z & \text{HM} & \text{EE}^7 \\
1 & 1/2 & -85.825 & -85.869(3) \\
1 & 1 & -91.492 & -91.520(8) \\
1/2 & 1 & -57.192 & -57.377(3) \\
\end{array}
\]

Each computation made again with a single core desktop. Time: 7 days with double precision complex: $\sim$ half a day with simple precision reals. No claim of optimality, can be significantly improved.

\footnote{Thanks to M. Lubasch for the EE data}
Discussion and possible extensions

- Markov chains for tensor network states introduced. No environment; very easy implementation.
- The method is certainly worth further exploration.
- It would be interesting to explore connections between the analysis of Markov chains and complexity issues regarding quantum Hamiltonians.
- First results encouraging; linear schedules seem to be possible. How well does it compare to existing methods? We don’t know but there are many possibilities for improvement:
  - Non-monotonic schedules? (seems very powerful)
  - Metropolis scaling? (trade-off between modest and ambitious moves)
  - Beyond the simple Haar-Markov chain? We are currently working on biased priors; can dramatically boost computations; compare say single site update and the Wolff algorithm for classical Ising models.
  - Other configuration spaces
- The method is very flexible.
  - Trees and MERAS?
  - Fermions? (Plain Jordan-Wigner, Mapping to Hamiltonians of spins, fermionic PEPS)
  - Low and high energy excited states? Change the figure of merit:
    \[ h'(\omega) = \theta(h(\omega) - e_1)\theta(e_2 - h(\omega)) + \lambda|\langle \omega | H^2 | \omega \rangle - \langle \omega | H | \omega \rangle^2|. \]
  - Spectral decomposition and time evolution for PEPS?
  - Finite temperature?