Loop Optimization for Tensor Network Renormalization

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1103/PhysRevLett.118.110504">http://dx.doi.org/10.1103/PhysRevLett.118.110504</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>American Physical Society</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Sun Jan 13 07:02:05 EST 2019</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/107448">http://hdl.handle.net/1721.1/107448</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>
Loop Optimization for Tensor Network Renormalization

Shuo Yang,1 Zheng-Cheng Gu,2,1 and Xiao-Gang Wen3,1

1Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada
2Department of Physics, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong
3Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

(Received 1 January 2017; published 15 March 2017)

We introduce a tensor renormalization group scheme for coarse graining a two-dimensional tensor network that can be successfully applied to both classical and quantum systems on and off criticality. The key innovation in our scheme is to deform a 2D tensor network into small loops and then optimize the tensors on each loop. In this way, we remove short-range entanglement at each iteration step and significantly improve the accuracy and stability of the renormalization flow. We demonstrate our algorithm in the classical Ising model and a frustrated 2D quantum model.

DOI: 10.1103/PhysRevLett.118.110504

Introduction.—In recent years, the tensor network (TN) approach [1,2] has become a powerful theoretical [3–24] and computational [8,25–60] tool for studying condensed matter systems. Many physical quantities, including the partition function of a classical system, the Euclidean path integral of a quantum system, and the expectation value of physical observables, can be expressed in terms of tensor networks. Evaluating these quantities is reduced to the contraction of a multidimensional tensor network. In the two-dimensional case, many algorithms [8,32,37–41,43,45–50,53–57] have been developed to implement the approximate tensor contractions. Among these, the tensor renormalization group approach introduced by Levin and Nave [38] and its generalizations [8,22,39,43–47,55,56,61] have unique features: the tensor contraction is based on a fully isotropic coarse-graining procedure. Moreover, when applying the method to a system on a finite torus, the computational cost is lower than those based on matrix product states (MPS) [32,37,41,48–50,53,54].

However, the Levin-Nave tensor network renormalization (TRG, also referred as LN-TNR here) [38] is based on the singular value decomposition (SVD) of local tensors, which only minimizes the truncation errors of tree tensor networks. Several improvements [45–47] have taken into account the effect of the environments, but they are still essentially based on tree tensor networks. These approaches cannot completely remove short-range entanglements during the coarse graining process. For example, in the 2D TN calculation of a partition function (or a path integral) TNR based on simple SVD cannot simplify the corner-double-line (CDL) tensor [38], despite the CDL tensor describing a product state that should be simplified to a one-dimensional tensor. In Ref. [8], this issue was seriously discussed. The authors pointed out that to further remove short-range entanglement, it is crucial to optimize the tensor configurations that contain a loop. However, due to the computational cost, only a crude iterative method is used to implement the loop optimization strategy. We refer to that method as Gu-Wen tensor network renormalization (TEFR, also referred as GW-TNR here). Reference [8] showed that GW-TNR can simplify CDL tensors, resulting in a simple fixed-point tensor for gapped or short-range correlated phases. This led to the discovery of symmetry-protected topological (SPT) order. Recently, Refs. [55,56] introduced a method based on the multiscale entanglement renormalization ansatz (MERA) [33] to completely remove short-range entanglement, even in critical systems. This approach is referred to as Evenbly-Vidal TNR (EV-TNR).

In this Letter, we develop a new practical and accurate algorithm called Loop-TNR, which can optimize looplike tensor configurations more effectively than GW-TNR. Loop-TNR can completely remove the short-range entanglement within a loop at each coarse-graining step, for both on- and off-critical systems. The performance of Loop-TNR is greater than EV-TNR, and it has a lower computational cost. To demonstrate this, we computed the central charge and scaling dimensions of the critical Ising model, and then examined the accuracy and stability of these data when undergoing coarse-grained transformations. All TNR methods can produce accurate central charge and scaling dimensions. However, their stabilities are significantly different. Loop-TNR and EV-TNR provide good stability (their data remain accurate after tens of iterations), while LN-TNR has the worst stability (its data remain accurate only for a few iterations).

Our results suggest that all TNR approaches can produce a fixed-point tensor which appears as the low-index part of the tensor (with a proper choice of basis). The high-index part is not represented by the fixed-point tensor, and can be considered to be the “junk” part of the tensor. As we perform more TNR iterations, the junk part may grow and eventually destroy the fixed-point tensor at low indices. The accuracy of an algorithm represents the accuracy of the fixed-point tensor at low indices. Its stability represents the
growth rate of the junk part of the tensor. We have found that Loop-TNR can significantly reduce the growth rate of the junk part. Moreover, Loop-TNR can be used to compute physical measurements of 2D projected entangled-pair states (PEPS) with high accuracy.

**Loop-TNR algorithm.**—The Loop-TNR algorithm has the same purpose as GW-TNR [8]—to eliminate local entanglement on a loop and determine the correct structures of fixed-point tensors. However, Loop-TNR significantly improves the numerical stability and accuracy of the renormalization group (RG) flow, especially for critical systems. The following illustrates the three main steps of the Loop-TNR algorithm. The first and last steps are exact, and the second is approximate. The method is discussed with regard to a square lattice, but generalizations to other lattices are straightforward.

The Loop-TNR methods begin with an entanglement filtering step [Figs. 1(a) and 1(g)] with two important features. First, it provides a canonical gauge for every tensor, and filters out the local entanglement of off-critical systems. More specifically, two projectors are inserted on each bond shown in Fig. 1(g). These projectors are constructed in an iterative way based on QR decompositions [62]. Subsequently, the tensors are redefined by combining the original tensors with the nearest projectors [see Fig. 1(g)] to complete the filtering step. In the Supplemental Material [62], we show that this approach can completely remove the CDL tensors. Thus, for off-critical systems containing CDL tensors (with gauge transformations), our method can simplify the tensors and reduce the bond dimensions. Although there is no bond reduction in critical systems, the canonical gauge provided by this method can enhance the performance of the following step. This step is quite efficient because the overall computational cost scales as $O(\chi^5)$, where $\chi$ is the bond dimension of the tensor.

In the next step the tensor network must be deformed from a square lattice to a square-octagon lattice [see Fig. 1(c)], as in the LN-TNR algorithm. However, approximations are necessary to avoid increasing the bond dimensions of the octagons. In the LN-TNR algorithm, this is achieved by minimizing the following single-site cost functions:

$$f = \|T_1 \cdot T_2 \cdot T_3 \cdot T_4 - S_1 \cdot S_2 \cdot S_3 \cdot S_4 \cdot S_5 \cdot S_6 \cdot S_7 \cdot S_8\|^2,$$

where the shaded square is deformed to an octagon. Since the cost function is now defined on a loop, we can remove the short-range entanglement inside this loop and significantly improve the accuracy, especially for critical systems. Furthermore, there is an efficient way to find the optimal $S$ tensors by viewing each patch as a wave function made up of matrix product states (MPS) with periodic boundary conditions. The eight dotted lines shown in Fig. 1(h) are the physical legs of the MPS, and the solid lines are the virtual legs of the MPS. Minimizing the cost function is equivalent to minimizing the distance between two MPS. Thus, $S$ tensors can be optimized using the well-developed variational MPS method [2,29,62]. The computational cost of this step scales as $O(\chi^6)$. The final step is the same as that of the LN-TNR algorithm. As shown in Fig. 1(e), a coarse-grained square lattice is obtained by contracting the tensor.

![FIG. 1. Three key steps of the Loop-TNR algorithm. (a) The entanglement filtering step. Projectors are inserted to eliminate local entanglements on the squares labeled with gray circles [see (g) for details]. (c) The loop optimization step. Each of the shaded squares is deformed to a octagon made up of 8 rank-3 tensors with bond dimensions no greater than $\chi$. The best approximation is found by minimizing the cost function in (h), i.e.,](image-url)
over the inner indices within the circles. The overall computational cost of all the steps only scales as $O(\chi^6)$, which is significantly more efficient than other improved LN-TNR methods, such as SRG/HOSRG algorithms $[O(\chi^7) - O(\chi^8)]$ [45–47], and EV-TNR algorithms $[O(\chi^7)]$ [55] and $O(\chi^6)$ [56,63]. Below, we demonstrate the advantages of the Loop-TNR algorithm using the classical Ising model on a square lattice.

**Classical Ising model.**—The partition function of the 2D classical Ising model is given by $Z = \sum_\{\sigma\} \exp \{\beta \sum_{(i,j)} \sigma_i \sigma_j\}$. It can also be expressed as the contraction of a 2D tensor network with $\chi = 2$ [38]. In this model, the spins are localized on the links of the square lattice. Each local tensor $T = T_{u;i,l,d,r}^{\text{Ising}}$ has the following nonzero components:

\[
T_{1,2,1,2}^{\text{Ising}} = e^{-4\beta}, \quad T_{2,1,2,1}^{\text{Ising}} = e^{-4\beta}, \quad T_{1,1,1,1}^{\text{Ising}} = e^{4\beta},
\]

\[
T_{2,2,2,2}^{\text{Ising}} = e^{4\beta}, \quad \text{others} = 1.
\]

The first step is to compute the free energy of this model with $2^{50}$ spins, so that it saturates to the value of the thermodynamic limit. Figure 2 shows the relative error of the free energy per site at and away from the critical temperature $T_c$. At the critical point [see Fig. 2(a)], the error of Loop-TNR decays much faster than the error of LN-TNR. When $\chi \leq 16$, the error of Loop-TNR decays almost exponentially with $\chi$. This demonstrates a significant improvement over LN-TNR. In Fig. 2(b), the errors of Loop-TNR remain almost constant for all temperatures near the critical point. When $\chi = 8$, Loop-TNR has an accuracy in the order of $10^{-7}$. At the same point LN-TNR has an accuracy of $10^{-4}$–10$^{-5}$. Other improved methods, such as SRG and HOSRG [45–47], can reduce the error by up to 3 orders of magnitude at off-critical conditions, but by only 1 order of magnitude at criticality. The recently proposed EV-TNR algorithm [55] can achieve the same accuracy with the same “effective” bond dimensions in the octagon (but a larger overall bond dimension [62]). However, Loop-TNR has a lower computational cost than EV-TNR.

![FIG. 2. Comparison of the relative errors of the free energy per site computed using LN-TNR and Loop-TNR. Results were obtained on a square lattice with $2^{50}$ spins. (a) Relative error as a function of bond dimension $\chi$ at the critical point. (b) Relative error as a function of temperature for off-critical Ising models.](image)

After applying several steps of Loop-TNR, we obtain an approximate fixed-point tensor with proper normalization and gauge fixing, which encodes the low-energy physics of the critical system. To prevent gauge fixing at the final step, $C_4$ lattice symmetry may be imposed on the RG flow. This produces a single rank-3 tensor that is approximately invariant at criticality [62].

As proposed in Ref. [8], the transfer matrix shown in Fig. 3(g) can be constructed, and the central charge and lowest scaling dimensions determined from the eigenvalues of the transfer matrix. When $\chi = 24$ and with $2^{18}$ spins, these conformal data have extremely high accuracies (up to five digits):

<table>
<thead>
<tr>
<th>$c$</th>
<th>$h_1$</th>
<th>$h_2$</th>
<th>$h_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loop-TNR: 0.500001 0.1250001 1.000006 1.124994</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EV-TNR: 0.50001 0.1250004 1.00009 1.12492</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exact: 1/2 1/8 1 9/8</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![FIG. 3. Comparison of central charge and scaling dimensions for LN-TNR and Loop-TNR at different iteration steps. The red dotted line denotes the central charge, the blue (light gray) solid lines denote the scaling dimensions in the $Z_2$-odd sector, and the black solid lines denote the scaling dimensions in the $Z_2$-even sector. In the $L = 2 (L = 4)$ case, a transfer matrix is constructed using two (four) columns of tensors [shown in (g) and (h)]. The central charge and scaling dimensions are determined from the eigenvalues of the transfer matrix [8].](image)
For comparison, the central charge and the scaling dimensions obtained using EV-TNR under the same conditions are given [55] (here $\chi$ denotes the largest bond dimension used in that scheme).

In addition to improving the accuracy of the central charge and scaling dimensions, Loop-TNR also significantly improves their stabilities. Figure 3 compares the results from LN-TNR and Loop-TNR. In the LN-TNR case shown in the left-hand column, the high-level scaling dimensions start to merge with the low-level scaling dimensions after a few iteration steps. This indicates that the high-index “junk” starts to merge quickly with the low-index approximate fixed-point tensor [62]. In Fig. 3(a), the $h = 2$ and $h = 2.125$ scaling dimensions are destroyed by the junk after 10 iteration steps. Correspondingly, LN-TNR fails to produce the accurate scaling dimensions, even for primary fields. In general, both stability and accuracy deteriorate at higher scaling dimensions (or, equivalently, higher-index tensor elements).

The conformal data are significantly improved using Loop-TNR. As shown in the right-hand column of Fig. 3, these data remain accurate up to 40 iteration steps in the case of $\chi = 16$, and even longer when $\chi = 32$. Moreover, the high-index junk is well separated from the low-index scaling dimensions. By increasing $\chi$, a greater number of scaling dimensions beyond the primary fields can be resolved from the approximate fixed-point tensors. As shown in Figs. 3(d) and 3(b), the $h = 3$ and $h = 3.125$ scaling dimensions are clearly visible in the $\chi = 32$ simulation, while they are difficult to distinguish from the high-index junk when $\chi = 16$.

We have shown that for higher bond dimensions, the proper RG flow lasts longer. Thus, we believe that at the infinite $\chi$ limit, Loop-TNR can determine an infinite dimensional fixed-point tensor described by Ising CFT at the continuum limit (with proper normalization and gauge fixing). For instance, four columns of tensors may be used to construct the transfer matrix [shown in Fig. 3(h)], which is equivalent to using $\chi = 256$. As shown in Fig. 3(f), a greater number of scaling dimensions can be evaluated, and the accuracy is greatly improved. The result shown in Fig. 3(f) suggests that the complete information of a CFT is encoded in the approximate fixed-point tensor. If more tensors are used to construct the transfer matrix, it is possible to reconstruct the whole conformal tower to a given accuracy. Moreover, we have found evidence that the operator product expansion coefficients are also encoded in the low-index approximate fixed-point tensors. How to compute these coefficients will be discussed in future work. Because the central charge, scaling dimensions, and operator product expansion coefficients of primary fields constitute the complete set of data for a CFT, the low-index approximate fixed-point tensors can completely determine the low-energy physics with an emergent conformal symmetry. The high-index junk is subject to the conformal symmetry-breaking perturbations introduced by truncation errors, which cannot be prevented in any numerical simulations with a finite $\chi$.

Variational energy for a 2D quantum model.—Loop-TNR can compute the physical quantities of 2D projected entangled-pair states, especially those states with divergent correlation lengths. We tested our algorithm by calculating the variational energy of the $D = 3$ PEPS proposed in Ref. [64]. This is a variational resonating valence bond ansatz for the $J_1 - J_2$ antiferromagnetic Heisenberg model on a square lattice around the maximally frustrated regime ($J_2 = 0.5J_1$). The extrapolated ground state energy was obtained in Ref. [64] using the boundary MPS method [3,5,32,40,65], the value of which is shown as the black dash-dot line in Fig. 4. The results of LN-TNR and Loop-TNR were calculated using a 256-site system with periodic boundary condition. Since this PEPS has a divergent correlation length, the energy from LN-TNR is highly frustrated, and far from the accurate value. Conversely, the energy determined from Loop-TNR quickly converges to the accurate value. Here, only 20 sweeps were carried out when minimizing the cost function Eq. (1) by the variational MPS method [2,29]. Using more sweeps would have improved the results.

Conclusions and discussions.—We have developed the Loop-TNR algorithm, a coarse-graining transformation based on loop optimizations, to significantly improve the RG flow for both critical and off-critical systems. We demonstrated the advantage of Loop-TNR using the classical Ising model on a square lattice. High accuracy and stability of the central charge and the lowest scaling dimensions were observed at criticality. Furthermore, good accuracy was achieved in the computation of the variational energy of a frustrated 2D PEPS.

Thanks to the concept of loop optimization, we may integrate the well-developed 1D algorithms with LN-TNR to enhance its performance. The integration with iTEBD [31] gives rise to GW-TNR [8], the integration with MERA

![FIG. 4. Benchmark of the variational energy of the $D = 3$ PEPS proposed in Ref. [64] for the maximally frustrated $J_1 - J_2$ antiferromagnetic Heisenberg model on a square lattice (with $J_2 = 0.5J_1$). Here, we consider a 256 site system with PBC. Because the benchmark energy (dashed line) is an extrapolation for infinite systems, it could be slightly lower than the actual variational energy for 256 sites.](image-url)
[33] results in EV-TNR [55], and now the integration with variational MPS [29] leads to Loop-TNR. From the viewpoint of quantum field theory, our way of removing local entanglement is equivalent to integrating out local modes during the RG transformation. As a result, Loop-TNR works better than the algorithms based on tree tensor networks (such as LN-TNR and SRG/HOSRG), where the local modes are only removed by a hard cut. For future works, we will explore the structure of the fixed-point tensor for a CFT. The 3D generalization of Loop-TNR is also a promising direction, where the “loop optimization” will be replaced by the “membrane optimization.”

We thank Guifre Vidal and Glen Evenbly for earlier conversations about Refs. [55] and [8]. We thank the Institute for Advanced Study at Tsinghua University and the Kavli Institute for Theoretical Physics at University of California Santa Barbara for hospitality. S. Yang thanks Ling Wang and Michael Lubasch for explaining technical details in their previous publications [54] and [49]. Z. C. G. acknowledges start-up support, Direct Grant No. 4053163, and the funding from RGC/ECS(No. 2191110). X. G. W. acknowledges start-up support, Direct Grant No. 4053163, and the funding from RGC/ECS(No. 2191110). X. G. W. was supported by NSF Grant No. DMR-1506475 and NSFC 11274192. He is also supported by the BMO Financial Group and the John Templeton Foundation. This research was supported in part by the National Science Foundation under Grant No. PHY11-25915. Research at Perimeter Institute is supported by the Government of Canada through Industry Canada and by the Province of Ontario through the Ministry of Economic Development & Innovation.

[63] According to Ref. [56], one may reduce the cost of EV-TNR from $O(\chi^7)$ to $O(\chi^6)$ by minimizing an approximate (rather than exact) cost function. Using similar techniques, one can reduce the cost of Loop-TNR from $O(\chi^6)$ to $O(\chi^5)$.