Tensor Renormalization Group Approach to Two-Dimensional Classical Lattice Models

Michael Levin[1] and Cody P. Nave[2]

[1]*Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA*
[2]*Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

(Received 29 November 2006; published 19 September 2007)

We describe a simple real space renormalization group technique for two-dimensional classical lattice models. The approach is similar in spirit to block spin methods, but at the same time it is fundamentally based on the theory of quantum entanglement. In this sense, the technique can be thought of as a classical analogue of the density matrix renormalization group method. We demonstrate the method—which we call the tensor renormalization group method—by computing the magnetization of the triangular lattice Ising model.

DOI: 10.1103/PhysRevLett.99.120601 PACS numbers: 05.10.Cc, 03.67.−a

Introduction.—The density matrix renormalization group (DMRG) technique has proved extraordinarily powerful in the analysis of one-dimensional quantum systems [1,2]. Thus, it is natural to try to develop an analogous renormalization group method in higher dimensions. Such a method could solve many currently intractable problems (such as the 2D Hubbard model).

Recent work has focused on generalizing DMRG to higher-dimensional quantum systems [3]. But it is also natural to try to generalize to higher-dimensional *classical* lattice models. While classical real space renormalization group methods (such as block spin methods [4]) have been around for many years, they have never achieved the generality or precision of DMRG.

In this Letter, we address this problem in the two-dimensional case. We use ideas from quantum information theory to develop a numerical renormalization group method that can effectively solve any two-dimensional classical lattice model. The technique—which we call the tensor renormalization group (TRG) method—has no sign problem and works equally well for models with complex weights.

Accurate numerical methods based on transfer matrices [5,6] have already been developed for 2D classical systems. The advantage of the approach described here is that it is a *fully isotropic* coarse-graining procedure, similar in spirit to block spin methods. It is thus naturally suited to investigating universal long distance physics. Also, on a more theoretical level, the method reveals the relationship between classical RG and quantum entanglement. Finally, if only for its simplicity, we feel that the method is a useful numerical tool in two dimensions as well as a natural candidate for higher-dimensional generalizations.

Tensor network models.—The tensor renormalization group method applies to a set of classical lattice models called "tensor network models" [7]. Many well-known statistical mechanical models, such as the Ising model, the Potts model, and the six vertex model, can be written naturally as tensor network models. In fact, as we show later, all classical lattice models with local interactions can be written as tensor network models [8].

To describe a tensor network model on the honeycomb lattice, one must specify a (cyclically symmetric) tensor T_{ijk}, with indices i, j, k running from 1 to D for some D. The corresponding tensor network model has a degree of freedom $i = 1, \ldots, D$ on each bond of the honeycomb lattice. The weight for a configuration $(i, j, k, \ldots)$ is given by

$$e^{-S(i,j,k,\ldots)} = T_{ijk} T_{ilm} T_{jnp} T_{kqr} \ldots, \tag{1}$$

where the product includes a tensor for each site of the lattice (Fig. 1). The partition function is the sum of all of the weights:

$$Z = \sum_{ijk\ldots} e^{-S(i,j,k,\ldots)} = \sum_{ijk\ldots} T_{ijk} T_{ilm} T_{jnp} T_{kqr} \ldots. \tag{2}$$

In other words, the partition function is obtained by taking the product of all of the tensors, contracting the pairs of indices on each bond.

The TRG method.—The tensor renormalization group method is a way to compute the partition function Z using a real space RG flow. We explain the method in the case of the honeycomb lattice. Each coarse-graining iteration is made up of two separate steps. The first is approximate, and the second is exact (Fig. 2). The first step is to find a tensor S such that

$$\sum_n S_{lin} S_{jkn} \approx \sum_m T_{ijm} T_{klm} \tag{3}$$

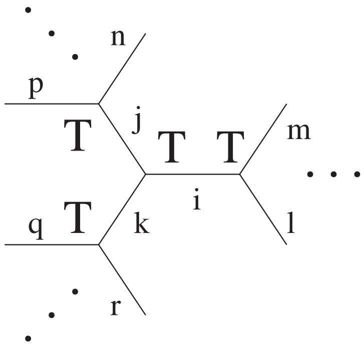
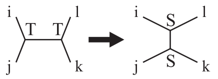


FIG. 1. A tensor network model on the honeycomb lattice.

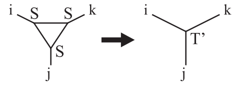
(we explain how to find such a tensor S at the end of this section). The relation (3) is a very useful property. Geometrically, it means that we can reconnect the lattice, making the replacement



(4)

wherever we like, without affecting the partition function Z. Applying this to the bonds shown in Fig. 2(a), we change the honeycomb lattice to the new lattice shown in Fig. 2(b). The partition function is now given by contracting the S tensors on the new lattice.

The second step is now clear. We group together triplets of neighboring points replacing them by a single lattice point with a coarse-grained tensor T':



$.$ (5)

Here the tensor T' is given by contracting over the three bonds of the triangle:

$$T'_{ijk} = \sum_{pqr} S_{kpq} S_{jqr} S_{irp}. \tag{6}$$

Making this replacement everywhere gives a new (coarser) honeycomb lattice [see Fig. 2(c)]. This completes the coarse-graining transformation. The end result is that the number of points in the lattice has decreased by a factor of 3 and T has been replaced by T'.

Iterating this procedure, one can compute the partition function of an arbitrarily large finite lattice. Thermodynamic observables and correlation functions can be obtained by taking numerical derivatives of $F = -\log(Z)$ or by evaluating the free energy of more general models where the tensors T_{ijk} vary from site to site. An additional feature is that the tensors T converge to a fixed point tensor T^*—whose physical significance we explain in the next section. The method is not limited to the honeycomb lattice and can easily be implemented on other lattices (see Fig. 3).

To complete the discussion, we address the issue of finding a tensor S which satisfies (3). The first step is to think of $\sum_m T_{ijm} T_{klm}$ as a $D^2 \times D^2$ matrix M: $M_{li,jk} = \sum_m T_{ijm} T_{klm}$. It is also useful to think of the tensor S_{lin} as a $D^2 \times D$ matrix $S_{li,n}$. Then the problem of satisfying (3) is the problem of finding a $D^2 \times D$ matrix S such that $M = SS^T$. In general, this factorization cannot be done exactly, since M typically has rank D^2, while SS^T has rank at most D.

However, an approximate solution can be obtained as follows. The idea is to choose the matrix S that minimizes the error, $|M - SS^T|^2$; the optimal S can then be found using the singular value decomposition of M. In more detail, first, one writes $M_{li,jk} = \sum_n s_n U_{li,n} V^*_{jk,n}$ (here s_n are the singular values and U, V are unitary matrices). Second, one truncates the matrices $U_{li,n}$, $V_{jk,n}$ keeping only those columns corresponding to the largest D singular values. The result is $D^2 \times D$ matrices $\tilde{U}_{li,n}$, $\tilde{V}_{jk,n}$. Finally, one sets $S_{lin} = \sqrt{s_n}\tilde{U}_{li,n}$. This gives the required factorization—provided that we adjust the phase ambiguity $U_{li,n} \rightarrow U_{li,n} e^{i\phi_n}$, $V_{jk,n} \rightarrow V_{jk,n} e^{-i\phi_n}$ appropriately. In practice, it is often more convenient to ignore the phase adjustment issue and set $S^A_{lin} = \sqrt{s_n}\tilde{U}_{li,n}$, $S^B_{jkn} = \sqrt{s_n}\tilde{V}^*_{jk,n}$. The result is a factorization $\sum_n S^A_{lin} S^B_{jkn} \approx \sum_m T_{ijm} T_{klm}$, where S^A and S^B differ by some phase factors. The TRG procedure can be applied as before—the only difference being that we have to keep track of two different tensors T^A, T^B for the A and B sublattices.

We will show in the next section that the error for this optimal decomposition is independent of the number of iterations and can be made arbitrarily small by increasing D. Indeed, the error vanishes as $\epsilon \sim \exp[-\text{const} \times (\log D)^2]$—the same scaling behavior as the truncation error in DMRG [2].

Physical picture.—In this section, we explain the physics behind the TRG method and give a physical interpretation of the fixed point tensor T^*. We begin by showing how arbitrary classical models in two dimensions can be thought of as tensor models on the honeycomb lattice. For concreteness, we frame our discussion around the case of the square lattice Ising model, $Z = \sum_{\{\sigma_i\}} \exp(K \sum_{\langle ij \rangle} \sigma_i \sigma j)$. Consider the partition function Z_R for some finite region R in the plane. One way to compute Z_R is to triangulate R, dividing it into triangles of size L much larger than the lattice spacing l [Fig. 4(a)].

Consider one of the triangles. Imagine summing over all of the lattice degrees of freedom within the triangle. The

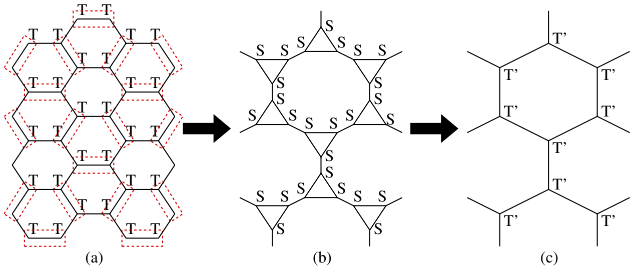


FIG. 2 (color online). A TRG transformation on the honeycomb lattice.

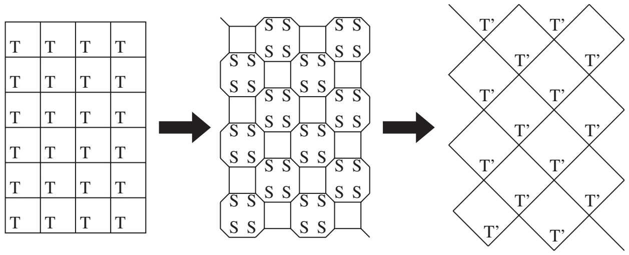
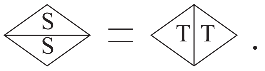


FIG. 3. A TRG transformation on the square lattice.

result will yield some number $\Psi = \Psi(\{\sigma_{i_n}\})$ that depends on the values of $\{\sigma_{i_n}\}$ at the boundary of the triangle [Fig. 4(b)]. It is convenient to separate out these boundary degrees of freedom into three groups $\{\sigma_{a_n}\}, \{\sigma_{b_n}\}, \{\sigma_{c_n}\}$ corresponding to the three sides a, b, c of the triangle. Denoting $\{\sigma_{a_n}\}$ schematically by α, and similarly for $\{\sigma_{b_n}\}, \{\sigma_{c_n}\}$, we can think of Ψ as a three-index tensor $\Psi = \Psi_{\alpha\beta\gamma}$.

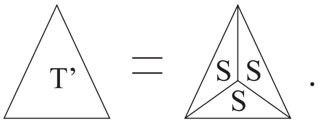
To obtain the partition function for the region R, we simply need to glue together all of the triangles and then sum over the spins at their boundaries: $Z_R = \sum_{\alpha\beta\gamma\delta\epsilon\ldots} \Psi_{\alpha\beta\gamma}\Psi_{\alpha\delta\epsilon}\ldots$. This is nothing but the partition function of a tensor model (2). (Actually, a more careful analysis shows that the resulting tensor model has different tensors Ψ^A, Ψ^B for the A, B sublattices).

It is useful to think about the tensor renormalization group transformation in this context. Recall that the first step is to find tensors S satisfying $\sum_n S_{\text{lin}} S_{jkn} \approx \sum_m T_{ijm} T_{klm}$. Thinking in terms of the triangles, the right-hand side is simply the partition function of the rhombus obtained by gluing two triangles together. Thus, a solution S can be constructed by setting S equal to the partition function of one of the obtuse triangles obtained by dividing the rhombus in the other direction:



(7)

Thus, the first step of the renormalization transformation process simply changes the triangulation as shown in Fig. 5(b). The second step also has a simple interpretation. Examining the definition of T', it is not hard to see that T' is simply the partition function of a large equilateral triangle, obtained by gluing together three obtuse triangles:



(8)

Thus, the second step simply glues together triplets of obtuse triangles to form larger equilateral triangles as shown in Fig. 5(c). In this way, the TRG method builds up larger and larger triangles.

However, there is one subtlety. As we build up larger and larger triangles, the corresponding tensors will have indices with larger and larger ranges, increasing from $2^{L/l}$ to $2^{\sqrt{3}L/l}$ to $2^{3L/l}$ and so on. Yet, the TRG method insists on approximating these tensors by a tensor with a fixed range D. How can this approximation possibly be accurate?

To answer this question, we must explain the physical meaning of $\Psi_{\alpha\beta\gamma}$. Let Ψ be a tensor obtained from the partition function of a very large triangle. Writing out the labels α, β, γ explicitly, we can write Ψ as a function $\Psi(\{\sigma_{a_n}\}, \{\sigma_{b_n}\}, \{\sigma_{c_n}\})$ of the spins on the three sides of the triangle. In fact, this function should be thought of as a *wave function* for a one-dimensional quantum spin system with spins living on the boundary of a triangle. This interpretation comes from thinking of the original two-dimensional classical model (e.g., the Ising model) as a $(1 + 1)$-dimensional quantum model. We think of the direction parallel to the boundary of the triangle as space and the (radial) direction perpendicular to the boundary as time. We imagine constructing a one-dimensional transfer matrix or quantum Hamiltonian H living on the boundary of the triangle, whose (radial) time evolution generates the two-dimensional classical model in question. Then in this picture, the function $\Psi(\{\sigma_{a_n}\}, \{\sigma_{b_n}\}, \{\sigma_{c_n}\})$ is the result of evolving H for a long time (the triangle is large). Hence, Ψ is simply the ground state of H—up to exponentially small corrections.

If we assume that the original classical model (e.g., the Ising model) is not critical, then Ψ is the ground state of a *gapped* Hamiltonian. Gapped ground states in one dimension have an important property: They are only weakly entangled. More specifically, it is known that the density matrix $\rho(x, L)$ of a region of size x converges to a fixed density matrix ρ_∞ as $x, L \to \infty$. Moreover, the size of the mth eigenvalue λ_m of ρ_∞ falls off rapidly with increasing m: $\lambda_m \sim \exp[-\text{const} \times \log(m)^2]$ [9].

It is this property which guarantees the accuracy of the TRG method. Indeed, as a consequence of this property, one can factor Ψ into a product of three spin states on the three sides of the triangle:

$$\Psi(\sigma_a, \sigma_b, \sigma_c) \approx \sum_{ijk=1}^{D} T_{ijk} \Psi_A^i(\sigma_a) \Psi_B^j(\sigma_b) \Psi_C^k(\sigma_c) \quad (9)$$

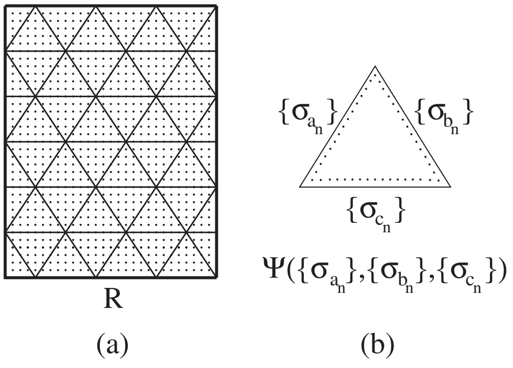


FIG. 4. After (a) dividing R into triangles, (b) the partition function of each triangle can be written as a function $\Psi(\{\sigma_{a_n}\}, \{\sigma_{b_n}\}, \{\sigma_{c_n}\})$ of the boundary spins.

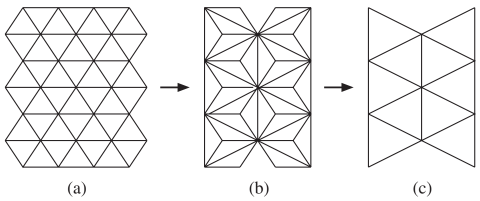


FIG. 5. The tensor renormalization group transformation can be viewed as a two-step change in the triangulation of R.

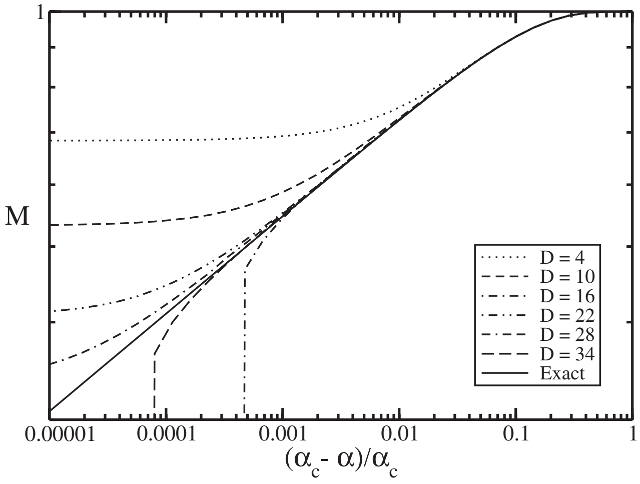


FIG. 6. The magnetization of the triangular lattice Ising model as a function of $\alpha = e^{-2\beta J}$, obtained using the TRG method.

with high accuracy. Such a factorization can be obtained by choosing Ψ_A^i, Ψ_B^i, Ψ_C^i to be the ith largest eigenstates of the density matrices of sides A, B, C and letting T_{ijk} be the matrix elements of the tensor $\Psi(\sigma_a, \sigma_b, \sigma_c)$ between these states. By the discussion above, the error ϵ of this representation is (i) independent of the size L of the triangle and (ii) decreases rapidly with increasing D ($\epsilon \sim \exp[-\text{const} \times (\log D)^2]$).

Thus, even though the exact tensor $\Psi_{\alpha\beta\gamma}$ has an exponentially large range, one can make a change of basis so that, in that basis, Ψ can be accurately approximated by a tensor T_{ijk} whose indices have a fixed range D. The TRG method can be thought of as a numerical technique for (approximately) constructing this tensor T_{ijk} for larger and larger triangles. The fixed point T^* is the value of this tensor in the limit of an infinitely large triangle.

The above analysis was based on the assumption that the classical model was not critical. If instead the classical model is critical, the associated quantum states are *gapless* ground states. Gapless ground states are more entangled then their gapped counterparts. The entanglement entropy $S = -\text{Tr}(\rho \log\rho)$ of a region of size x in a system of size L grows logarithmically with the region size $S \sim \log(x)$ [10]. This means that the factorization (9) will always break down when the triangle is sufficiently large. Thus, in principle, the TRG method—like DMRG [2]—breaks down at criticality.

A simple example.—In this section, we demonstrate the method with a simple example: the triangular lattice Ising model $Z = \sum_{\{\sigma\}} \exp(\beta J \sum_{\langle ij\rangle} \sigma_i \sigma_j)$. Note that the Ising model partition function can be written as a sum over domain wall configurations where the domain walls live on the bonds of the honeycomb lattice. This domain wall model can be easily realized by a tensor network with $D = 2$. We think of the state $i = 1$ as denoting "no domain wall" and $i = 2$ as denoting "domain wall." Then the tensor with nonzero components

$$T^{111} = 1, \qquad T^{122} = T^{212} = T^{221} = \alpha, \qquad (10)$$

$\alpha = e^{-2\beta J}$, gives rise to the correct Boltzmann weight.

Applying the TRG method to the above tensor, we compute the free energy per unit spin $F = -\frac{1}{N\beta}\log(Z)$ in the thermodynamic limit $N \to \infty$. The magnetization M can be obtained by taking numerical derivatives of F (though we need to use a more complicated tensor T to represent an Ising model with an external magnetic field H). Increasing D, the computation rapidly converges to the exact result [11] except in an increasingly narrow interval around the critical point $\alpha_c = 1/\sqrt{3}$ (Fig. 6).

This interval becomes so narrow that one can study the critical point itself. For example, the magnetization curve for $D = 34$ predicts an α_c within 10^{-4} of the exact result. One can even estimate the critical exponent β from the scaling behavior of the magnetization. We find $\beta = 0.12$, not far from the exact value $\beta = 1/8$. However, as explained earlier, the TRG method (like DMRG) is best suited to studying systems *off criticality.* An interesting question for further research is whether the TRG method can be modified so that its (almost exponential) accuracy is uniform, both away from and near critical points.

The authors thank Michael Freedman and Frank Verstraete for motivating work on this problem and Nihat Berker for useful discussions. This work was supported by the Harvard Society of Fellows (M. L.) and by NSF Grant No. DMR-0517222 (C. P. N.).

[1] S. R. White, Phys. Rev. Lett. **69**, 2863 (1992).
[2] U. Schollwöck, Rev. Mod. Phys. **77**, 259 (2005).
[3] F. Verstraete and J. Cirac, arXiv:cond-mat/0407066.
[4] K. G. Wilson, Rev. Mod. Phys. **47**, 773 (1975).
[5] T. Nishino, J. Phys. Soc. Jpn. **64**, 3598 (1995).
[6] V. Murg, F. Verstraete, and J. Cirac, Phys. Rev. Lett. **95**, 057206 (2005).
[7] Y. Shi, L. Duan, and G. Vidal, Phys. Rev. A **74**, 022320 (2006).
[8] The argument given in the text applies equally well to models with continuous degrees of freedom. In this case, while the *exact* tensor network description may involve infinitely large tensors, after a change of basis the tensors can be truncated to a finite size D with a small error $\epsilon \sim \exp[-\text{const} \times (\log D)^2]$ [cf. Eq. (9)].
[9] K. Okunishi, Y. Hieida, and Y. Akutsu, Phys. Rev. E **59**, R6227 (1999).
[10] G. Vidal, J. Latorre, E. Rico, and A. Kitaev, Phys. Rev. Lett. **90**, 227902 (2003).
[11] R. Houtappel, Physica (Utrecht) **16**, 425 (1950).