Markov Entropy Decomposition: A Variational Dual for Quantum Belief Propagation

David Poulin[1] and Matthew B. Hastings[2,3]

[1]*Département de Physique, Université de Sherbrooke, Québec, J1K 2R1, Canada*
[2]*Department of Physics, Duke University, Durham, North Carolina 27708, USA*
[3]*Microsoft Research, Station Q, Elings Hall, University of California, Santa Barbara, California 93106, USA*

(Received 9 December 2010; revised manuscript received 25 January 2011; published 24 February 2011)

We present a lower bound for the free energy of a quantum many-body system at finite temperature. This lower bound is expressed as a convex optimization problem with linear constraints, and is derived using strong subadditivity of von Neumann entropy and a relaxation of the consistency condition of local density operators. The dual to this minimization problem leads to a set of quantum belief propagation equations, thus providing a firm theoretical foundation to that approach. The minimization problem is numerically tractable, and we find good agreement with quantum Monte Carlo calculations for spin-$\frac{1}{2}$ Heisenberg antiferromagnet in two dimensions. This lower bound complements other variational upper bounds. We discuss applications to Hamiltonian complexity theory and give a generalization of the structure theorem of [P. Hayden *et al.*, Commun. Math. Phys. **246**, 359 (2004).] to trees in an appendix.

DOI: 10.1103/PhysRevLett.106.080403 PACS numbers: 05.30.−d, 02.70.−c, 03.67.−a, 05.10.−a

Describing the properties of a local quantum system is perhaps the central problem of theoretical physics. However, the exponential growth of the Hilbert space with system size makes it prohibitive to even write down the state of a system with even a modest number of sites. For this reason, variational methods, such as matrix product states used in density matrix renormalization group [1–4] and their higher dimensional generalizations [5,6], are a central tool, describing a state with a small number of parameters, allowing a practical optimization of the energy.

All these methods provide an upper bound to the free energy and the quality of the approximation cannot be assessed directly. In this Letter, we present a *lower* bound to the free energy that nicely complements variational approaches. We use strong subadditivity (SSA) of von Neumann entropy [7] to approximate the system's entropy by a local quantity. This approximation is exact when the system is a Markov network [8]—i.e., when its long-range correlations arise due to correlations over shorter distances—but in general provides a lower bound to the true entropy. By relaxing the consistency constraints on the reduced density operators of the systems, we find a formula for the free-energy expressed as a convex minimization problem with linear constraints.

Our formula for the free energy is similar to the Bethe free energy [9]—and its generalization by Kikuchi [10]—but differs by a crucial ordering of the lattice sites. This distinction is responsible for the lower bound obtained by our method, in contrast to Bethe's and Kikuchi's approximations which are uncontrolled. The dual of the minimization problem provides a set of quantum belief propagation equations similar to those presented in [8,11–13]. This connection provides a solid theoretical foundation to understand the success and limitations of quantum belief propagation. Similar connections [14] and algorithms [15] have been found in the classical setting.

Markov entropy decomposition.—Consider a lattice of N spins that we label from 1 to N. The labeling of the sites chosen will determine the order in which we apply our procedure later. The Hamiltonian of the system is a sum of geometrically local terms $H = \sum_X h_X$ where X labels subsets of $\{1, \ldots, N\}$ and locality means that $h_X = 0$ when the radius of X is larger than some constant w. Given the density matrix ρ of the system, we can compute the average energy $E(\rho) = \mathrm{Tr}(\rho H) = \sum_X \mathrm{Tr}(\rho_X h_X)$ from knowledge of only the reduced density matrices $\rho_X \equiv \mathrm{Tr}_{\bar{X}}\rho$ on small local regions, that can be obtained from the partial trace of ρ over the complement $\bar{X}$ of X.

At finite temperature T, we are interested in the system's free energy $F(T) \equiv \min_\rho\{E(\rho) - TS(\rho)\}$. Unlike the energy, the entropy $S(\rho) \equiv -\mathrm{Tr}(\rho \log\rho)$ cannot be evaluated in general from knowledge of only the reduced density matrices ρ_X over regions X of finite radius. We define an approximate way of doing this evaluation, following the derivation of [15] in the classical setting. For every site k, define a subset of sites $\mathcal{N}_k$ consisting of "neighboring" sites. There is no unique prescription for the choice of $\mathcal{N}_k$, but it is useful to imagine that they consist of a set of sites located within a finite distance from k. With trivial manipulations, we can rewrite the entropy of the system in the form of an "entropy chain rule" $S(\rho) = \sum_{k=1}^N S(k|\{<k\})$ where the conditional entropy of a region X given region Y is $S(X|Y) \equiv S(X \cup Y) - S(Y)$, the entropy of any region X is denoted $S(X) \equiv S(\rho_X) = -\mathrm{Tr}(\rho_X \log\rho_X)$, and we use the notation $\{<k\} = \{1, 2, \ldots, k-1\}$.

Quantum entropy S obeys SSA [7], which implies

$$S(k|\{<k\}) \leq S(k|\{<k\} \cap \mathcal{N}_k) = S(k|\mathcal{M}_k), \qquad (1)$$

where we define the "Markov shield" of site k by $\mathcal{M}_k = \{<k\} \cap \mathcal{N}_k$, see Fig. 1. We can define the Markov entropy $S_M(\rho) \equiv \sum_{k=1}^{N} S(k|\mathcal{M}_k)$ which upper bounds the system's entropy. Because each term in that sum can be computed from the reduced density matrices on site k and its Markov shield, the Markov entropy, unlike the entropy S, is suitable for direct numerical calculations.

Returning to the free-energy calculation, we now have the bound $F(T) \geq F_M(T) \equiv \min_\rho\{E(\rho) - TS_M(\rho)\}$. The Markov free energy F_M of any given state is equal to its true free energy if SSA is saturated with the given choice of Markov shields as shown in Fig. 1. Because both E and S_M can be evaluated from the density matrix of constant-size regions X, we can express $F_M(\rho) = E(\rho) - TS_M(\rho)$ as a function of some set of reduced density operators $\{\rho_X\}$ and write $F_M(T) = \min_{\{\rho_X\}\in\Omega} F_M(\{\rho_X\})$, where Ω denotes the set of *consistent* reduced density matrices that are all obtainable from some global density matrix ρ, i.e., $\Omega \equiv \{\{\rho_X\}: \exists \rho, \quad \rho_X = \mathrm{Tr}_{\bar{X}}\rho, \quad \forall\ X\}$.

Unfortunately, verifying consistency of a set of reduced density matrices $\{\rho_X\}$ is a difficult problem, it is complete for the quantum Merlin-Arthur complexity class [16], so it is very unlikely that Ω can be characterized efficiently. Thus, we will make one more approximation and enlarge the set Ω to the set $\tilde{\Omega}$ of all *locally* consistent reduced density matrices that agree on overlapping regions, i.e., $\tilde{\Omega} \equiv \{[\rho_X]: \mathrm{Tr}_{\overline{X\cap Y}}\rho_X = \mathrm{Tr}_{\overline{X\cap Y}}\rho_Y, \ \forall\ (X, Y)\}$. Since all reduced density matrices in Ω are derived from one global ρ, it should be clear that $\Omega \subset \tilde{\Omega}$, and as a consequence

$$F_{\mathrm{MED}}(T) \equiv \min_{\{\rho_X\}\in\tilde{\Omega}} F_M(\{\rho_X\}) \leq F_M(T) \leq F(T). \qquad (2)$$

Equation (2) defines our numerical method which we call the Markov entropy decomposition (MED) scheme. The Markov free energy $F_M(\{\rho_X\})$ to be minimized to evaluate $F_{\mathrm{MED}}(T)$ is a convex function [17] over the cone of semipositive operators $\{\rho_X\}$ subject to some linear constraints specified in the definition of $\tilde{\Omega}$. Thus, it is suitable for numerical optimization. There is no unique prescription for the choice of $\mathcal{M}_k$, but given two choices, the one which leads to the higher free energy is to be preferred due to the variational nature of our approach.

Numerical results on translationally invariant systems.—The procedure simplifies greatly when applied to translationally invariant systems. If we assume that all density matrices ρ_X are related by translational symmetry, the Markov free-energy is a function of a single density matrix. We have numerically investigated this method with a spin-$\frac{1}{2}$ antiferromagnetic Heisenberg model on an infinite two-dimensional square lattice. The correlation length diverges exponentially in $1/T$ [18], and at low temperatures it becomes much larger than the size of our shield, thus providing a hard-case study. However, we expect on physical grounds that those long-range correlations will be Markovian because different Nèel-like orders are almost perfectly distinguishable over sufficiently large regions. For the same reason, the fact that the low energy states presumably break translational symmetry should not cause a problem because a mixture over the different symmetry breaking sectors restores the symmetry and does not change the value of F_{MED}. We have used a Markov shield of size 7 and 10, so that the main computational task of our program was exact diagonalization of (nonsparse) matrices of size 2^8 and 2^{11}, respectively. Figure 2 compares our results to other methods.

The MED free energy with the 10-site shield is in excellent agreement with quantum Monte Carlo (QMC) calculations for the entire temperature range. This agreement with QMC calculations is better than the one obtained from exact diagonalization (ED) of a 4×4 lattice. In fact, those diagonalization results are very well approximated by MED with a 7-site shield. Here we see the biggest advantage of MED: because of the constraints imposed on the minimization, the results converge to the thermodynamic limit faster than ED.

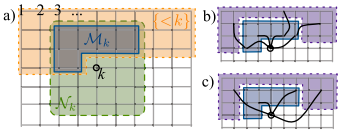

FIG. 1 (color online). (a) The Marrkov shield (shown in blue) $\mathcal{M}_k$ is the intersection of the neighborhood (green) of k and the sites preceding k (orange). (b) The entanglement (represented by black lines) between site k and the preceding sites is all mediated by the Markov shield: the state of the first k sites can be constructed by adding one extra spin to the state of the first $k-1$ site and coupling it only to the sites of the shield [19]. This turns inequality Eq. (1) into an equality. (c) There is direct entanglement between site k and the sites preceding k, so the Markov entropy is not equal to the true entropy, but it is an upper bound.

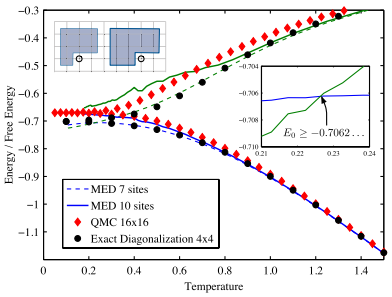

FIG. 2 (color online). Numerical results obtained from MED for the spin-$\frac{1}{2}$ Heisenberg antiferromagnet on a 2D square lattice. The energy (green) and free energy (blue) are obtained for a 7- and 10-site Markov shield, of shape illustrated in the upper left corner. The energy is evaluated as $\mathrm{Tr}(h\rho_2)$ from the two-body state ρ_2 minimizing Eq. (2). Results are compared to exact diagonalization of a 4×4 lattice and quantum Monte Carlo calculations. The crossing of energy and free-energy curve (negative entropy, 7-site shield) lower bounds to the ground energy.

Note that $F(T)$ is a monotonically decreasing function of T, equal to the ground-state energy E_0 at $T = 0$. Moreover, $F_{\mathrm{MED}}(T) \leq F(T)$ for all T. It follows that $\max_T F_{\mathrm{MED}}(T) \leq \max_T F(T) = E_0$. This maximum occurs at the point where the Markov entropy density $S_M(\{\rho_X\})$ goes negative, and $F_{\mathrm{MED}}(T) = E_{\mathrm{MED}}(T)$. The crossing point of F_{MED} and E_{MED} obtained with the 7-site shield gives a lower bound $E_0 \geq -0.7062\ldots$ to the true ground-state energy of the system.

We have used this technique to lower bound the ground-state energy of the one-dimensional model. Results obtained with a k-site neighborhood are in good agreement with ED results on a chain of length roughly $2k$ (with periodic boundary conditions). This can be understood from the fact that the ground-state entropy of a block of ℓ sites, $S(\ell)$, is an increasing function of ℓ for $\ell \leq k$, and then decreases to reach 0 when $\ell = 2k$ since the entire system is in a pure state. Thus, enforcing a positive Markov entropy density $S_M = S(k) - S(k-1)$ compels the system in our simulations to behave as it were on a lattice of size $2k$, even though we are manipulating states of k spins, providing some heuristic explanation for the improved convergence, compared to ED, seen above.

All these lower bounds on the ground-state energy and the lower bounds on the free energy, would be rigorous if the convex optimization problem were solved exactly. However, all our results are subject to numerical error. We used fairly elementary minimization methods (conjugate gradient) and more elaborate techniques that exploit the special features of this problem are likely to improve the results; we hope that this Letter will stimulate research in this direction. Numerical fluctuations are most prominent in the energy, while the free-energy curve is rather smooth. The fluctuations are largest near the specific heat peak; to understand this, consider the free energy $E - TS_M$ as a function of E, assuming for simplicity that S_M equals the correct entropy $S(E)$. At a minimum of F, $\frac{\partial F}{\partial E} = 0$, and $\frac{\partial^2 F}{\partial E^2} = \frac{1}{Tc}$ and so for large c, the basin around the minimum is shallow, increasing numerical error. We now describe an alternate approach, a dual problem, which connects to quantum belief propagation. If this dual problem could be turned into a *variational* dual problem (a concave function whose maximum equals the minimum of the Markov free energy), it would provide mathematically rigorous lower bounds on F.

Dual problem: quantum belief propagation.—Consider a length-N spin chain and define density matrices ρ_k and σ_k associated to segments $k-n$ to k and $k-n$ to $k-1$, respectively. In this case, the minimization problem defined at Eq. (2) using a Markov shield containing the n sites preceding site k becomes $\sum_{k=n}^{N}(\mathrm{Tr}\{\rho_k[\hat{H}_k + \log\rho_k - I \otimes A_k - B_k \otimes I + \mu_k]\} - \mathrm{Tr}\{\sigma_k[\log\sigma_k - A_{k-1} - B_k + \nu_k]\})$ where for $k = n, \ldots N$, the matrices A_k and B_k and the scalars μ_k and ν_k are Lagrange multipliers used to enforce $\mathrm{Tr}_{k-n}\rho_k = \sigma_{k+1}$, $\mathrm{Tr}_k\rho_k = \sigma_k$, and the trace normalization of ρ_k and σ_k, respectively, and $A_N = 0$. This is simply Eq. (2) where we introduced an additional variable σ_k equal to the partial trace of ρ_k. Above, $\hat{H}_k$ is the part of the Hamiltonian supported on sites k to $k+n$ properly weighted to avoid double counting, and we have set temperature $T = 1$ to avoid cluttering equations. Taking derivatives with respect to ρ_k and σ_k yields $\hat{H}_k + \log\rho_k - IA_k - B_kI + \mu'_k = 0$ and $\log\sigma_k - A_{k-1} - B_k + \nu'_k = 0$ where $\nu'_k = \nu_k + 1$ and $\mu'_k = \mu_k + 1$, and we have dropped the $\otimes$ symbols. These equations, together with the constraints imposed on the reduced density matrices, give a set of self-consistent mean-field equations

$$A_{k-1} = \log(\mathrm{Tr}_k\rho_k) - B_k + \nu'_k = \log(\mathrm{Tr}_k e^{-\hat{H}_k + IA_k + B_kI - \mu'_k}) - B_k + \nu'_k \quad (3)$$

$$B_{k+1} = \log(\mathrm{Tr}_{k-n}\rho_k) - A_k + \nu'_{k+1} = \log(\mathrm{Tr}_{k-n} e^{-\hat{H}_k + IA_k + B_kI - \mu'_k}) - A_k + \nu'_{k+1}. \quad (4)$$

One can show that any solution to these equations is a minimum of the Markov free energy Eq. (2). Because this function is convex, the solution to Eqs. (3) and (4) is unique.

We can conceive an iterative procedure to approach solutions to Eqs. (3) and (4). Starting from an initial guess for the A_k and B_k, we obtain new guesses by inserting these values into Eqs. (3) and (4) which provides new values and recurse. Renaming $A_{k-1} = \log m_{k\to k-1}$ and $B_{k+1} = \log m_{k\to k+1}$, we recognize Eqs. (3) and (4) as almost the *belief propagation* prescription of [8,12]

$$m_{k\to k-1} \propto \mathrm{Tr}_k(\Lambda_k \odot m_{k+1\to k} \odot m_{k-1\to k}) \odot m^{-1}_{k-1\to k} \quad (5)$$

$$m_{k\to k+1} \propto \mathrm{Tr}_{k-n}(\Lambda_k \odot m_{k+1\to k} \odot m_{k-1\to k}) \odot m^{-1}_{k+1\to k} \quad (6)$$

$$\rho_k \propto \Lambda_k \odot m_{k+1\to k} \odot m_{k-1\to k} \quad (7)$$

where all proportionality constants can be set by normalization and $\Lambda_k = \exp(-\hat{H}_k)$. The $\odot$ product is defined by $A \odot B = \exp(\log A + \log B)$. We note a subtle difference between these belief propagation equations and those of [8,12]. If the action of the partial trace and the $\odot$ product were commutative as they are in the classical case, the two appearances of the term $m_{k-1\to k}$ in Eq. (5) would cancel, and similarly for $m_{k+1\to k}$ in Eq. (6). These cancellations were assumed in [8,12], based on heuristic arguments and numerical evidences. However, we see that they are required to establish a direct connection with the MED. Any fixed point of the iteration equations for messages m yields a lower bound to the free energy of the system. Moreover, as in [8,11,12], this iterative procedure can be used to evaluate other quantities such as correlation functions, and extends other geometries.

State reconstruction and probabilistically checkable proofs (PCP).—Given a global quantum state ρ, such that SSA is saturated for the given choice of Markov shields, we can reconstruct the global state from the

local state. Using the structure theorem of [19], we have $\log(\rho_{\{<k+1\}}) = \log(\rho_{\{<k\}}) + \log(\rho_{k\cup\mathcal{M}_k} - \log(\rho_{\mathcal{M}_k})$. Iterating this procedure allows us to reconstruct the global state from the local state. In the supplementary material [20], we extend this idea and show that any state saturating SSA on a tree graph is the thermal state of a Hamiltonian that is the sum of local, commuting terms. This procedure may help address the structure of topologically ordered states, since many lattice models with topological order saturate SSA with an appropriate choice of shields [21] (see [20]).

Deciding whether the ground-state energy of a classical Hamiltonian on N particles is 0 or greater than $N\epsilon$ for some positive constant ϵ is a very difficult problem. In general, it is NP complete, by the famous PCP theorem [22]. The analogous decision problem for a quantum Hamiltonian [23] is in QMA [24], but it is not known to be QMA-complete (this is the quantum PCP conjecture). While this question concerns zero temperature, it is equivalent to determining whether the free energy becomes negative within a constant accuracy ϵ at temperature $T < \epsilon/\log d$ where d is the number of levels of each particle. It is easy to verify if a set of operators $\{A_k, B_k\}$ are a solution to Eqs. (3) and (4), so the problem of lower bounding the free energy of a quantum system using the Markov entropy decomposition is in NP. Thus, one way to disprove the quantum PCP conjecture would be to rigorously evaluate tightness of this bound. Showing that a constant accuracy can be achieved with a shield of size $n \in \mathcal{O}(\log N)$ would suffice.

Multipatch MED.—We now discuss a possible extension of our method. Let F_M^1 and F_M^2 denote the Markov free-energy formulas obtained from two different neighborhoods in our procedure. Clearly, $F_M^{\max} = \max_k F_M^k$ is a lower bound to the free energy. The convex function

$$F_{\mathrm{MED}}^{1,2}(T) \equiv \min_{\{\rho_X\}\in\tilde{\Omega}} \max_k F_M^k(\{\rho_X\})$$

is an even better lower bound. That is, instead of minimizing F_M^1 and F_M^2 separately, we minimize their maximum, subject to the constraint that the reduced density matrices used to compute the two formulas are locally consistent with one another.

In particular, the shapes of $\mathcal{M}_1$ and $\mathcal{M}_2$ can be chosen to capture correlations on different length scales of the system. For instance, $\mathcal{M}_1$ could consist of $n = 6$ consecutive sites while $\mathcal{M}_2$ could consist of three pairs of sites, each separated by some distance, i.e., $\mathcal{M}_2 = \{1, 2, 4, 5, 8, 9\}$. Clearly, $\mathcal{M}_1$ captures the short range correlations while $\mathcal{M}_2$ captures the long-range correlations. The free-energy formula obtained by the combination of both regions is forced to assign reduced density matrices compatible with both type of correlations.

Discussion.—MED is on the one hand a possible numerical tool for studying the thermodynamics of quantum systems in a more accurate way than is possible using exact diagonalization. On the other hand, it provides a theoretical basis for the quantum belief propagation procedure developed previously to study disordered quantum systems; while we focused in translationally invariant systems above, we can apply the procedure more generally, e.g., to quantum spin glasses [25], treating each reduced density matrix ρ_X as an independent variable. Finally, it offers a physics-inspired procedure that may help tackle outstanding problems in quantum computational complexity.

We thank Roger Melko for providing the QMC data. Computational resources were provided by RQCHP. D. P. receives financial support from NSERC and FQRNT.

[1] S. Östlund and S. Rommer, Phys. Rev. Lett. **75**, 3537 (1995).
[2] J. Dukelsky, M. Martín-Delgado, T. Nishino, and G. Sierra, Europhys. Lett. **43**, 457 (1998).
[3] F. Verstraete, D. Porras, and J. I. Cirac, Phys. Rev. Lett. **93**, 227205 (2004).
[4] G. Vidal, Phys. Rev. Lett. **99**, 220405 (2007).
[5] F. Verstraete and J. I. Cirac, arXiv:cond-mat/0407066.
[6] Y.-Y. Shi, L.-M. Duan, and G. Vidal, Phys. Rev. A **74**, 022320 (2006).
[7] E. Lieb and M. Ruskai, J. Math. Phys. (N.Y.) **14**, 1938 (1973).
[8] M. Leifer and D. Poulin, Ann. Phys. (Leipzig) **323**, 1899 (2008).
[9] H. Bethe, Proc. R. Soc. A **150**, 552 (1935).
[10] R. Kikuchi, Phys. Rev. **81**, 988 (1951).
[11] M. B. Hastings, Phys. Rev. B **76**, 201102(R) (2007).
[12] D. Poulin and E. Bilgin, Phys. Rev. A **77**, 052318 (2008).
[13] C. Laumann, A. Scardicchio, and S. L. Sondhi, Phys. Rev. B **78**, 134424 (2008).
[14] J. S. Yedidia, in *Advanced Mean Field Methods: Theory and Practice* (MIT Press, Cambridge, Massachusetts, 2001), p. 21.
[15] A. Globerson and T. Jaakkola, in *11th Int. Conf. on Art. Intel. and Stat.* (2007) http://www.stat.umn.edu/~aistat/proceedings/start.htm.
[16] Y.-K. Liu, *Proc. RANDOM* (Springer-Verlag, Berlin, 2006), p. 438.
[17] That conditional entropy is convex also follows from SSA.
[18] S. Chakravarty, B. I. Halperin, and D. R. Nelson, Phys. Rev. B **39**, 2344 (1989).
[19] P. Hayden, R. Jozsa, D. Petz, and A. Winter, Commun. Math. Phys. **246**, 359 (2004).
[20] See supplementary material at http://link.aps.org/supplemental/10.1103/PhysRevLett.106.080403.
[21] M. Levin and X.-G. Wen, Phys. Rev. Lett. **96**, 110405 (2006).
[22] S. Arora, C. Lund, R. Motwani, and M. Szegedy, J. ACM **45**, 501 (1998).
[23] D. Aharonov, I. Arad, Z. Landau, and U. Varizani, in *Proc. of the 41st Ann. ACM Symp. on Theo. of Comp.* (2009), p. 417.
[24] A. Kitaev, in *Classical and Quantum Computation*, edited by A. Kitaev, A. H. Shen, and M. N. Vyalyi (2002).
[25] T. K. Kopeć and K. D. Usadel, Phys. Rev. Lett. **78**, 1988 (1997).