The EA Ising Model via Parallel Tempering and D-Wave

Benjamin Krakoff

December 2, 2020

1 Edwards-Anderson Ising Model of a Weighted Graph

For a graph $G = (V, E)$ with variables $x_i \in \{\pm 1\}, i \in V$ and weights $J_{i,j}, (i, j) \in E$, define the hamiltonian

$$H(x) = \sum_{(i,j)\in E} J_{i,j} x_i x_j \tag{1}$$

Finding global minima for H is an NP-complete combinatorial optimization problem [1]. However, approximate solutions can be found efficiently by thinking of G as a physical system with sites i, spins x_i and couplings $J_{i,j}$. We investigate and compare parallel tempering and D-Wave quantum annealing as methods for finding approximate minima.

2 Parallel Tempering

2.1 Markov-Chain Monte Carlo with Single-Spin Updates

As an introduction parallel tempering and to set notation, we review a basic MCMC method for finding approximate minima. Thinking of (G, x, H) as a thermodynamic system at temperature T, the probability of the system being in a state x is given by the Boltzmann distribution

$$P(x; T) = \frac{e^{-H(x)/T}}{Z_T} \tag{2}$$

where $Z_T := \sum_x e^{-H(x)/T}$ is the partition function for the system. The first key observation is that P has maxima precisely at the minima of H, and the second is that we can construct a Markov chain with stationary distribution P without knowing Z_T, which is as difficult to compute as the minima of H

To construct the Markov chain, first initialize each $x_i = \pm 1$ with probability $1/2$. At each step, pick a site i at random, and let $\tilde{x}$ be the same state as x but with $\tilde{x}_i = -x_i$. Replace x with $\tilde{x}$ with probability

$$P_{x,\tilde{x}} = \min\{1, e^{(H(x)-H(\tilde{x}))/T}\} \tag{3}$$

This Markov chain is clearly ergodic, and satisfies the detailed balance equations.

$$\frac{P_{x,\tilde{x}}}{P_{\tilde{x},x}} = \frac{\min\{1, e^{(H(x)-H(\tilde{x}))/T}\}}{\min\{1, e^{(H(\tilde{x})-H(x))/T}\}} = \frac{P(\tilde{x},T)}{P(x,T)}$$

If you run this Markov process for long times, it will converge to the stationary distribution, and spend most of its time in the lowest energy states, in principle minimizing H. In practice, the system will typically get frustrated and stuck in local minima. By examining (2) and (3), the trade-off becomes apparent; at low temperatures, the probabilities concentrate on minimal energy solutions, but are likely to get stuck in local minima while at high temperatures, the system can more easily escape local minima but the probability distribution function is more spread out over phase space.

2.2 Parallel Tempering

Parallel tempering is a technique to leverage the benefits of MCMC methods at low and high temperatures. We follow the basic implementation suggested in [2].

Let's begin with the main idea of parallel tempering. Suppose we run two replicas $x^{(1)}$ and $x^{(2)}$ in parallel at temperatures $T_1 < T_2$ using the MCMC scheme outlined in 2.1. As the replicas explore phase space, $x^{(1)}$ is likely to have lower energy than $x^{(2)}$, but $x^{(2)}$ is able to more easily escape local minima. So as we let $x^{(1)}$ and $x^{(2)}$ walk around phase space, we let $x^{(1)}$ peek at the state of $x^{(2)}$, and if $x^{(2)}$ happens to have found a lower energy solution, $x^{(1)}$ can choose to swap states with $x^{(2)}$. Care must be taken to choose T_1 and T_2 sufficiently close so that the energy histograms of $x^{(1)}$ and $x^{(2)}$ overlap, and swaps have a non-negligible chance of occurring.

Here is how we implemented parallel tempering. Run replicas $x^{(1)}, x^{(2)}, \ldots, x^{(n)}$ at temperatures $T_1 < T_2 < \ldots < T_{n-1} < T_n$, where the T_i's are evenly spaced and sufficiently close together so that the energy histograms of $x^{(i)}$ and $x^{(i+1)}$ overlap. Run each replica for m steps, where m is the size of the system. Pick an i at random, and swap the state of $x^{(i)}$ with the state of $x^{(i+1)}$ with probability $\min\{1, e^{(1/T_{i+1}-1/T_i)(H(x^{(i+1)})-H(x^{(i)}))}\}$. The swap probability is chosen to satisfy detailed balance for the joint probability distribution

$$P(x^{(1)}, T_1, \ldots, x^{(n)}, T_n) = \prod_{1 \le i \le n} P(x^{(i)}, T_i) \propto e^{-\sum_i H(x^{(i)})/T_i}$$

3 D-Wave Systems

The D-Wave quantum annealer is another tool for finding approximate minima of H. The D-Wave annealer consists of a large Pegasus graph, where the vertices are qubits with spins ± 1 and the edges are couplers, whose weights, or coupling strengths, corresponding to the $J_{i,j}$ can be set to any value. The precise structure of the Pegasus graph and more details can be found in [3], but the relevant fact for our purposes is that D-Wave systems minimizes hamiltonians associated to subgraphs of this particular sparse graph. To solve problems for an arbitrary graph G, one needs a minor embedding of G into the Pegasus graph.

3.1 Minor Embedding

Let's start with an illustrative example. Suppose we have a device that can minimize any hamiltonian on a loop with 4 vertices, and we want to minimize a hamiltonian on a loop with 3 vertices. We can write our hamiltonian explicitly as

$$H(x) = J_{1,2}x_1x_2 + J_{2,3}x_2x_3 + J_{1,3}x_1x_3$$

We ask the device to minimize

$$\hat{H}(x) = J_{1,2}x_1x_2 + J_{2,3}x_2x_3 + J_{1,3}x_1x_4 - Cx_3x_4$$

where $C >> 1$.
In effect, the 3^{rd} vertex is replaced with vertices 3 and 4 connected by an edge, and the weight of this edge is sufficiently large so that any minimal state for $\hat{H}$ must have $x_3 = x_4$, and so we can recover the minimal state for H by taking the first 3 coordinates of the solution for $\hat{H}$.

More generally, a minor embedding of G_1 in G_2 is a map $\phi : V(G_1) \to 2^{V(G_2)}$ so that

- For each $v \in V(G_1)$, $\phi(v)$ is connected.
- For $v, w \in V(G_1)$, $\phi(v)$ and $\phi(w)$ are disjoint.
- If (v, w) is an edge in G_1, there is at least one edge between $\phi(v)$ and $\phi(w)$.

To minimize a hamiltonian on an arbitrary graph, D-Wave systems first finds a minor embedding ϕ into the Pegasus graph, ensures vertices within $\phi(v)$ are very strongly coupled, and then assigns edges between $\phi(u)$ and $\phi(v)$ weight $J_{u,v}$. The general problem of finding a minor embedding of G_1 in G_2 is NP-hard, but when G is sparse and has many fewer vertices than the Pegasus graph D-Wave supplies heuristic algorithms that work well in practice [4]. When G is not sparse, D-Wave has computed minor embeddings of complete graphs which can be used to minor-embed G by just forgetting unneeded edges.

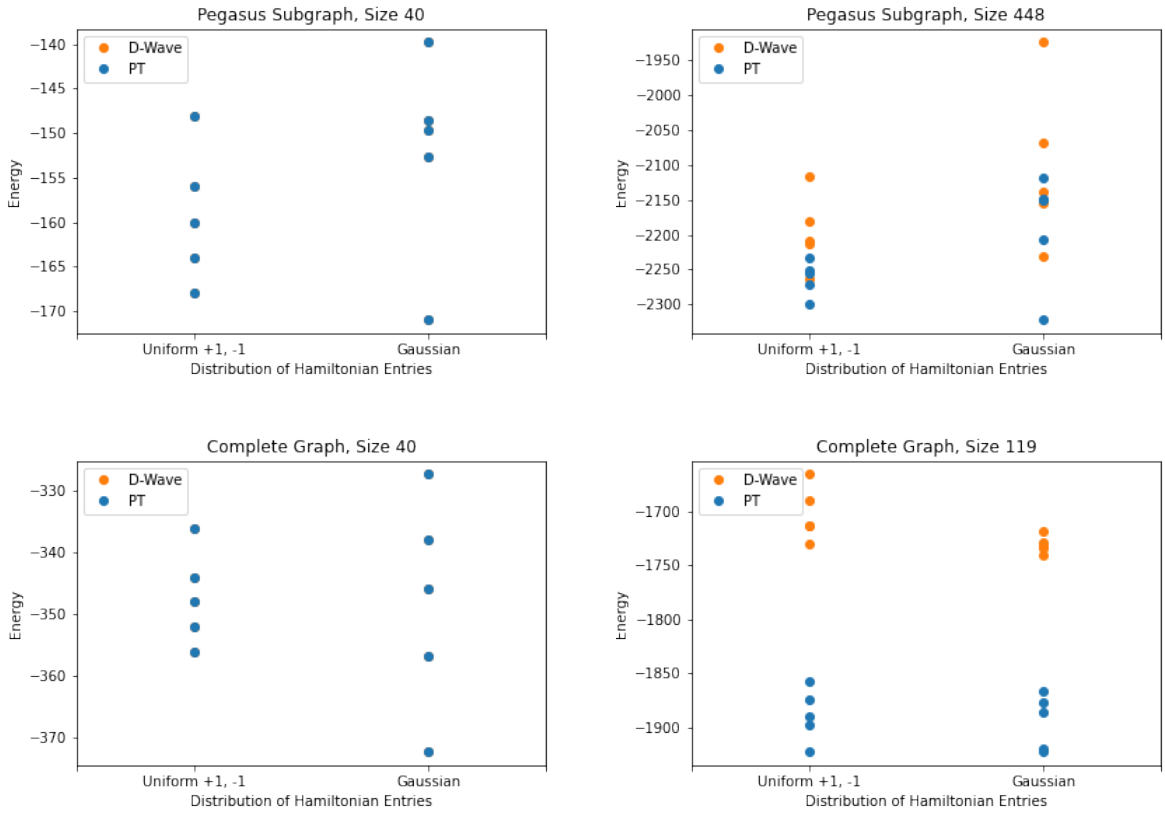
4 Comparison

Performance is compared across several families of graphs and ensembles. We consider subgraphs of the Pegasus graph of size 40 and 448 along with complete graphs of size 40 and 119. For each graph we generate 10 hamiltonians, 5 by choosing $J_{i,j} = \pm 1$ uniformly at random and 5 by choosing $J_{i,j} \sim \mathcal{N}(0, 1)$ and then let D-Wave and parallel tempering attempt to find the global minimum for each for approximately the same amount of time. Both solvers had identical performance on small graphs, but on larger graphs parallel tempering performed better for all graphs and ensembles of J_{ij}. For large graphs we also compute the discrepancy, defined as

$$\min\{(x_D + x_P)/2, (x_D - x_P)/2\}/n$$

where x_D and x_P are the solutions produced by D-Wave and parallel tempering, respectively and n is the number of vertices. The discrepancy computes at how many nodes the

solutions x_D and x_P differ taking into account the symmetry $H(x) = H(-x)$. The source code for parallel tempering and data can be found at `https://github.com/benkrakoff/Parallel-Tempering`. The source code for interacting with D-Wave can be found at `https://github.com/benkrakoff/DWave_Tester`.



	Pegasus, Uniform	Pegasus, Gaussian	Complete, Uniform	Complete, Gaussian
Discrepancy	.446	.386	.074	.106

References

[1] F. Barahona, "On the computational complexity of Ising spin glass models," *Journal of Physics A Mathematical General*, vol. 15, pp. 3241–3253, Oct. 1982.

[2] D. J. Earl and M. W. Deem, "Parallel tempering: Theory, applications, and new perspectives," *Phys. Chem. Chem. Phys.*, vol. 7, pp. 3910–3916, 2005.

[3] N. Dattani, S. Szalay, and N. Chancellor, "Pegasus: The second connectivity graph for large-scale quantum annealing hardware," 2019.

[4] J. Cai, W. G. Macready, and A. Roy, "A practical heuristic for finding graph minors," 2014.