Using Cluster Expansion for Approximate Counting on the Hard-Core Model in Sub-Quadratic Time.

Brian Tan, Ewan Davies
Department of Computer Science, Colorado State University

Motivation

Graphical models arise in a number of important applications including physics and statistics. The **hard-core model** is a discrete model of gas particles as independent sets in a graph: sets of vertices which are not joined by any edges. The model can be very intricate, and we investigate computational problems associated with understanding its macroscopic behavior. The key problem is to approximate the **partition function** in polynomial time, which is related to the classic #P-complete problem of counting the number of independent sets in a graph.

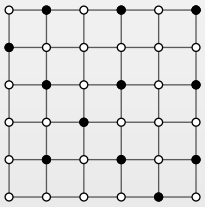


Figure 1: An independent set in a lattice graph. The shaded vertices represent locations in a crystal that are currently occupied by gas particles.

Introduction

The hard-core partition function on a graph G is a weighted count of independent sets in G:

$$Z_G(\lambda) = \sum_{I \in \mathcal{I}(G)} \lambda^{|I|} = \sum_{k \geq 0} \lambda^k i_k(G)$$

where $i_k(G)$ is the number of independent sets of size k in G and $\lambda \in \mathbb{C}$. We focus on graphs of maximum degree Δ and real values of the parameter λ such that $\lambda = O(1/\Delta^{1+k})$, for some $k > 0$.

The **correlation decay** algorithm for approximating the partition function due to Weitz [1] uses a self avoiding walk (SAW) tree to calculate **occupation ratios**, which can then be used to approximate the partition function. The hard-core model corresponds to a probability distribution on independent sets such that

$$\Pr_{G,\lambda}[I] = \frac{\lambda^{|I|}}{Z_G(\lambda)}$$

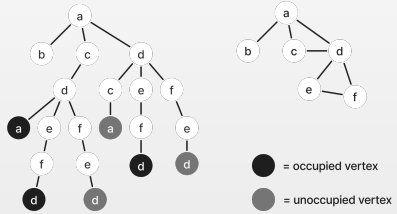


Figure 2: Self-Avoiding Walk Tree

The main computation propagates occupation ratios from leaves to the root using the recurrence

$$R_v = \lambda \prod_{i=1}^{d} \frac{1}{1 + R_i}$$

where $R_1, \ldots R_d$ are the ratios of the children of v. An occupied vertex has ratio ∞ and an unoccupied vertex has ratio 0.

Using the full SAW tree gives an exact answer in exponential time but for small enough λ,truncating the SAW tree gives an approximate answer in polynomial time. We can approximate $Z_G(\lambda)$ to constant accuracy in time $\tilde{O}(n^{1+1/k})$ by truncating the SAW tree at depth

$$\ell = \frac{\log n}{k \log \Delta} + O(1).$$

Anand et al [2] showed that you can improve Weitz's results by introducing randomness. They halve ℓ and get a runtime of $\tilde{O}(n^{1+1/(2k)})$. To compensate for the smaller SAW tree, they use a Recursive Marginal Sampler (RMS) to sample the ratios $R_i \in \{0, \infty\}$ of the leaves of the truncated tree with the correct probabilities, avoiding the error associated with Weitz' method on such a small tree.

The RMS is a simple, randomized graph traversal algorithm similar to BFS/DFS, but the proofs that (1) it gives a sample with the correct probability and (2) terminates quickly with high probability are quite subtle and difficult.

Results

Our goal is to match the performance of Anand et al [2], but remove the Recursive Marginal Sampler which is difficult to analyze.

The **cluster expansion** (CE) shows that the marginal probability $\Pr[u \in I]$ of a vertex u can be computed using an infinite series whose terms correspond to the graph structure around u. Truncating the series gives an approximation of the marginal. Using these probabilities for the leaves of the SAW tree, we can replace the RMS with simple simulation of Bernoulli random variables. Provided the total variation error is small enough, there is a coupling of the simulated variables with the RMS that will agree with high probability.

Our algorithm replaces the elegant but intricate RMS with a fixed calculation using cluster expansion followed by simulation of Bernoulli random variables to achieve the same goal in the same running time.

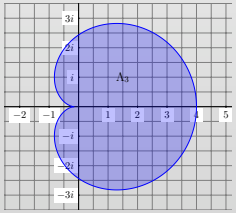
Algorithm

```
Input: graph G with n vertices
Output: approximation of Z_G(λ)

Set G_1 = G
For i = 1,…,n:
    Build SAW tree T_i for v_i in G_i
    For each leaf u of T_i:
        Use CE to approximate Pr[u ∈ I]
        Sample R_u ∈ {0,∞} using Pr[u ∈ I]
    Compute R_{v_i} propagating ratios in T_i
    G_{i+1} = G_i − v_i
Return ∏_{i=1}^{n} (1 + R_{v_i})
```

Conclusion

This preliminary investigation shows that cluster expansion can replace the RMS with similar performance. Our ultimate goal is to remove the use of randomness entirely in this sub-quadratic algorithmic problem, and we hope that this work will lead to further breakthroughs. Progress may require understanding the complex zeros of $Z_G(\lambda)$, illustrated below.



References

[1] *Dror Weitz. 2006. Counting independent sets up to the tree threshold. In Proceedings of the thirty-eighth annual ACM symposium on Theory of Computing (STOC '06). Association for Computing Machinery, New York, NY, USA, 140–149. https://doi.org/10.1145/1132516.1132538*

[2] *Konrad Anand, Weiming Feng, Graham Freifeld, Heng Guo, Jiaheng Wang: "Approximate Counting for Spin Systems in Sub-Quadratic Time", 2023; arXiv:2306.14867.*