

Adaptive Simulated Annealing: A Near-optimal Connection between Sampling and Counting

Daniel Štefankovič*

Santosh Vempala†

Eric Vigoda‡

Abstract

We present a near-optimal reduction from approximately counting the cardinality of a discrete set to approximately sampling elements of the set. An important application of our work is to approximating the partition function Z of a discrete system, such as the Ising model, matchings or colorings of a graph. The standard approach to estimating the partition function $Z(\beta^*)$ at some desired inverse temperature β^* is to define a sequence, which we call a cooling schedule, $\beta_0 = 0 < \beta_1 < \dots < \beta_\ell = \beta^*$ where $Z(0)$ is trivial to compute and the ratios $Z(\beta_{i+1})/Z(\beta_i)$ are easy to estimate by sampling from the distribution corresponding to $Z(\beta_i)$. Previous approaches required a cooling schedule of length $O^*(\ln A)$ where $A = Z(0)$, thereby ensuring that each ratio $Z(\beta_{i+1})/Z(\beta_i)$ is bounded. We present a cooling schedule of length $\ell = O^*(\sqrt{\ln A})$.

For well-studied problems such as estimating the partition function of the Ising model, or approximating the number of colorings or matchings of a graph, our cooling schedule is of length $O^*(\sqrt{n})$ and the total number of samples required is $O^*(n)$. This implies an overall savings of a factor of roughly n in the running time of the approximate counting algorithm compared to the previous best approach.

A similar improvement in the length of the cooling schedule was recently obtained by Lovász and Vempala in the context of estimating the volume of convex bodies. While our reduction is inspired by theirs, the discrete analogue of their result turns out to be significantly more difficult. Whereas a fixed schedule suffices in their setting, we prove that in the discrete setting we need an adaptive schedule, i. e., the schedule depends on Z . More precisely, we prove any non-adaptive cooling schedule has length at least $O^*(\ln A)$, and we present an algorithm to find an adaptive schedule of length $O^*(\sqrt{\ln A})$ and a nearly matching lower bound.

*Department of Computer Science, University of Rochester, Rochester, NY 14627. Email: stefanko@cs.rochester.edu

†College of Computing, Georgia Institute of Technology, Atlanta GA 30332. Email: vempala@cc.gatech.edu. Research supported in part by NSF grant CCF-0721503.

‡College of Computing, Georgia Institute of Technology, Atlanta GA 30332. Email: vigoda@cc.gatech.edu. Research supported in part by NSF grant CCF-0455666.

1 Introduction

This paper explores the intimate connection between counting and sampling problems. By counting problems, we refer to estimating the cardinality of a large set (or its weighted analogue), or in a continuous setting, an integral over a high-dimensional domain. The sampling problem refers to generating samples from a probability distribution over a large set. The well-known connection between counting and sampling (first studied in a general complexity-theoretic context by [13] and explored earlier in a more restricted setting by [1]) is the starting point for popular Markov chain Monte Carlo (MCMC) methods for many counting problems. Some notable examples from computer science are the problems of estimating the volume of a convex body [4, 16] and approximating the permanent of a non-negative matrix [12].

In statistical physics, a key computational task is estimating a partition function, which is an example of a counting problem. Evaluations of the partition function yield estimates of thermodynamic quantities of interest, such as the free energy and the specific heat. The corresponding sampling problem is to generate samples from the so-called Gibbs (or Boltzman) distribution. The analogue of the connection between sampling and counting in this area is multi-stage sampling [22].

We present an improved reduction from approximate counting to approximate sampling. These results improve the running time for many counting problems where efficient sampling schemes exist. We present our work in the general framework of partition functions from statistical physics. This framework captures many well-studied models from statistical physics, such as the Ising and Potts models, and also captures many natural combinatorial problems, such as colorings, independent sets, and matchings. For the purposes of this paper we define a (discrete) partition function as follows.

Definition 1.1. Let $n \geq 0$ be an integer. Let a_0, \dots, a_n be non-negative real numbers such that $a_0 \geq 1$. The function

$$Z(\beta) = \sum_{i=0}^n a_i e^{-i\beta}$$

is called a *partition function* of degree n . Let $A := Z(0)$.

This captures the standard notion of partition functions from statistical physics in the following manner. The quantity i corresponds to the possible values of the Hamiltonian. Then a_i is the number of configurations whose Hamiltonian equals i . For instance, in the (ferromagnetic) Ising model on a graph $G = (V, E)$, a configuration is an assignment of $+1$ and -1 spins to the vertices. The Hamiltonian of a configuration is the number of edges whose endpoints have different spins. The quantity β is referred to as the inverse temperature. The computational goal is to compute $Z(\beta)$ for some choice of $\beta \geq 0$. Note, when $\beta = 0$ the partition function is trivial since $Z(0) = \sum_{i=0}^n a_i = 2^{|V|}$. The condition $a_0 \geq 1$ is clearly satisfied, in fact, we have $a_0 = 2$ by considering the all $+1$ and the all -1 configurations.

The general notion of partition function also captures standard combinatorial counting problems as illustrated by the following example. Let Ω be the set of all k -labelings of a graph $G = (V, E)$ (i. e., labelings of the vertices of G by numbers $\{1, \dots, k\}$). Given a labeling σ , let its Hamiltonian $H(\sigma)$ be the number of edges in E that are monochromatic in σ . Let Ω_i denote the set of all k -labelings of G with $H(\sigma) = i$. Let $a_i = |\Omega_i|$. We would like to compute $Z(\infty) = a_0$, i. e., the number of valid k -colorings of G . Once again, the case $\beta = 0$ is trivial since we have $Z(0) = k^{|V|}$. The condition $a_0 \geq 1$ simply requires that there is at least one proper k -coloring.

The standard approach [22] to compute $Z(\beta)$ is to express it as a telescoping product of ratios of the partition function. Consider a set of configurations Ω which can be partitioned as $\Omega = \Omega_0 \cup \Omega_1 \cup \dots \cup \Omega_n$, where $|\Omega_i| = a_i$ for $0 \leq i \leq n$. Suppose that we have an algorithm which for any inverse temperature $\beta \geq 0$ generates a random configuration from the distribution μ_β over Ω where the probability of a configuration $\sigma \in \Omega$ is

$$\mu_\beta(\sigma) = \frac{e^{-\beta H(\sigma)}}{Z(\beta)}, \quad (1)$$

where $H(\sigma)$ is the Hamiltonian of the configuration defined as $H(\sigma) = i$ for $\sigma \in \Omega_i$. We now describe the standard approach estimating a partition function. To approximate $a_0 = Z(\infty)$, take $\beta_0 < \beta_1 < \dots < \beta_\ell$ with $\beta_0 = 0$ and $\beta_\ell = \infty$. Express $Z(\infty)$ as a telescoping product

$$Z(\infty) = Z(0) \frac{Z(\beta_1)}{Z(\beta_0)} \frac{Z(\beta_2)}{Z(\beta_1)} \dots \frac{Z(\beta_\ell)}{Z(\beta_{\ell-1})}. \quad (2)$$

The initial term $Z(0)$ is typically trivial to compute. It remains to estimate the ratios. In the general setting of Definition 1.1, for $X \sim \mu_\beta$, the random variable

$$W_{\beta, \beta'} := e^{(\beta - \beta')H(X)} \quad (3)$$

is an unbiased estimator for $Z(\beta')/Z(\beta)$. Indeed,

$$\mathbb{E}(W_{\beta, \beta'}) = \frac{1}{Z(\beta)} \sum_{\sigma \in \Omega} e^{-\beta H(\sigma)} \cdot e^{(\beta - \beta')H(\sigma)} = \frac{Z(\beta')}{Z(\beta)}. \quad (4)$$

Equation (4) is related to the single histogram, or reweighting methods in statistical physics [20, 6].

We approximate each fraction in the product (2) using the unbiased estimator $W_{\beta_i, \beta_{i+1}}$. Taking sufficiently many samples for each $W_{\beta_i, \beta_{i+1}}$ will give a good approximation of a_0 . The question we study in this paper is: how should one choose the inverse temperatures $\beta_0, \dots, \beta_\ell$ so as to minimize the number of samples needed to estimate (2)? A specific choice of $\beta_0, \dots, \beta_\ell$ is called a *cooling schedule*.

In the past, MCMC algorithms have used cooling schedules that ensure that each ratio in the telescoping product is bounded by a constant. Intuitively, this seems to be the best possible setting — a higher ratio in each phase requires more samples overall. For applications such as colorings or Ising model, requiring that each ratio is at most a constant implies that the length of the cooling schedule is at least $\Omega(n)$, since $Z(0)$ and $Z(\infty)$ typically differ by an exponential factor. All cooling schedules prior to our work were non-adaptive, i. e., the sequence depends only on the parameters n and A but not the structure of the partition function Z .

In the discrete setting, a trivial non-adaptive cooling schedule has length $O(n \ln A)$, and, recently, [2] presented an improved non-adaptive cooling schedule of length $O((\ln n) \ln A)$. The recent volume algorithm of [16, 17] uses a non-adaptive cooling schedule of length $O(\sqrt{n})$ to estimate the volume of a convex body in \mathbb{R}^n . The main idea for the short cooling schedule in the volume setting is that even though a ratio to be estimated is not bounded by a constant, the variance of the estimator is at most a constant times the square of its expectation. The proof of this relies heavily on the logconcavity of the function $\beta^n Z(\beta)$ where Z is the analogue of the partition function in their setting. The cooling schedule of [16, 17] was also useful in the setting of convex optimization [14].

The discrete setting presents a new challenge. As we show in this paper, there can be no short *non-adaptive* cooling schedule for discrete partition functions, i. e., any non-adaptive schedule has length $\Omega((\ln n) \ln A)$ in the worst case.

Our main result is that every partition function has an *adaptive schedule* of length roughly $\sqrt{\ln A}$, where $A = Z(0)$. (Note, $\sqrt{\ln A}$ is roughly \sqrt{n} in the examples we have been considering here). Further, *the schedule can be figured out efficiently on the fly*, with little overhead in the complexity. Lastly, this bound is nearly the best possible (up to logarithmic factors in the leading term).

The existence of a short schedule follows from an in-

interesting geometric fact: any convex function f can be approximated by a piecewise linear function g consisting of few pieces, see Figure 1 in Section 4 for an illustration.

For well-known problems such as counting colorings or matchings, and estimating the partition function of the Ising model, our results imply an improvement in the running time by a factor of n , since the complexity grows with the square of the schedule length; see Section 6 for a precise statement of the applications of our results.

In Section 2 we formalize the setup described in this introduction. The lower bound for non-adaptive schedules is formally stated as Lemma 3.1 in Section 3. The existence of a short cooling schedule is proved in Section 4, and formally stated in Theorem 4.1. The algorithm for constructing a short cooling schedule is presented in Section 5. Finally, in Section 6 we present applications of our improved cooling schedule.

Many of the proofs and details of the algorithms are omitted from this extended abstract. We encourage the interested reader to refer to the full version of the paper [21].

2 Chebyshev cooling schedules

Let $W := W_{\beta, \beta'}$ be the estimator defined by (3) whose expectation is a individual ratio in the telescoping product. As usual, we will use the squared coefficient of variance $\text{Var}(W)/\mathbb{E}(W)^2$ as a measure of the quality of the estimator W , namely to derive a bound on the number of samples needed for reliable estimation of $\mathbb{E}(W)$. We will also use the quantity $\mathbb{E}(W^2)/\mathbb{E}(W)^2 = 1 + \text{Var}(W)/\mathbb{E}(W)^2$.

The following lemma of Dyer and Frieze [3] is now well-known.

Theorem 2.1. *Let W_1, \dots, W_ℓ be independent random variables with $\mathbb{E}(W_i^2)/\mathbb{E}(W_i)^2 \leq B$ for $i \in [\ell]$. Let $\widehat{W} = W_1 \dots W_\ell$. Let S_i be the average of $16B\ell/\varepsilon^2$ independent random samples from W_i for $i \in [\ell]$. Let $\widehat{S} = S_1 S_2 \dots S_\ell$. Then*

$$\Pr\left((1 - \varepsilon)\mathbb{E}(\widehat{W}) \leq \widehat{S} \leq (1 + \varepsilon)\mathbb{E}(\widehat{W})\right) \geq 3/4.$$

It will be convenient to rewrite $\mathbb{E}(W^2)/\mathbb{E}(W)^2$ for $W := W_{\beta, \beta'}$ in terms of the partition function Z . We have

$$\mathbb{E}(W^2) = \frac{1}{Z(\beta)} \sum_{\sigma \in \Omega} e^{-\beta H(\sigma)} e^{2(\beta - \beta')H(\sigma)} = \frac{Z(2\beta' - \beta)}{Z(\beta)},$$

and hence

$$\frac{\mathbb{E}(W^2)}{\mathbb{E}(W)^2} = \frac{Z(2\beta' - \beta)Z(\beta)}{Z(\beta')^2}. \quad (5)$$

Equation (5) motivates the following definition.

Definition 2.2. Let $B > 0$ be a constant. Let Z be a partition function. Let $\beta_0, \dots, \beta_\ell$ be a sequence of inverse temperatures such that $0 = \beta_0 < \beta_1 < \dots < \beta_\ell = \infty$. The sequence is called a *B-Chebyshev cooling schedule* for Z if

$$\frac{Z(2\beta_{i+1} - \beta_i)Z(\beta_i)}{Z(\beta_{i+1})^2} \leq B,$$

for all $i = 0, \dots, \ell - 1$.

The following bound on the number of samples is an immediate consequence of Theorem 2.1.

Corollary 2.3. *Let Z be a partition function. Suppose that we are given a B-Chebyshev cooling schedule $\beta_0, \dots, \beta_\ell$ for Z . Then, using $16B\ell^2/\varepsilon^2$ samples in total, we can compute \widehat{S} such that*

$$\Pr\left((1 - \varepsilon)Z(\infty) \leq \widehat{S} \leq (1 + \varepsilon)Z(\infty)\right) \geq 3/4.$$

3 Lower bound for non-adaptive schedules

A cooling schedule will be called *non-adaptive* if it depends only on n and $A = Z(0)$ and assumes $Z(\infty) \geq 1$. Thus, such a schedule does not depend on the structure of the partition function.

The advantage of non-adaptive cooling schedules is that they do not need to be figured out on the fly. An example of a non-adaptive Chebyshev cooling schedule that works for any partition function of degree n , where $Z(0) = A$, is

$$0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n \ln A}{n}, \infty. \quad (6)$$

The idea behind the schedule (6) is that small changes in the inverse temperature result in small changes of the partition function.

The length of the schedule (6) is $O(n \ln A)$. The following more efficient non-adaptive Chebyshev cooling schedule of length $O((\ln A) \ln n)$ is given in [2]:

$$0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{k}{n}, \frac{k\gamma}{n}, \frac{k\gamma^2}{n}, \dots, \frac{k\gamma^t}{n}, \infty, \quad (7)$$

where $k = \lceil \ln A \rceil$, $\gamma = 1 + \frac{1}{\ln A}$, and $t = \lceil (1 + \ln A) \ln n \rceil$.

Next we show that the schedule (7) is the best possible up to a constant factor. We will see later that *adaptive* cooling schedules can be much shorter.

Lemma 3.1. *Let $n \in \mathbb{Z}^+$, and $A, B \in \mathbb{R}^+$. Let $S = \beta_0, \beta_1, \dots, \beta_\ell$ be a non-adaptive B-Chebyshev cooling schedule which works for all partition functions of degree at most n with $Z(0) = A$, and $Z(\infty) \geq 1$. Assume $\beta_0 = 0$ and $\beta_\ell = \infty$. Then*

$$\ell \geq \ln(n/e) \left(\frac{\ln(A-1)}{\ln(4B)} - 1 \right).$$

The number of samples needed in Theorem 2.1 (and Corollary 2.3) is linear in B and hence, in view of Lemma 3.1, the optimal value of B is a constant. Our understanding of non-adaptive schedules is now complete up to a constant factor. In particular, the schedule (7) and Lemma 3.1 imply that the optimal non-adaptive schedule has length $\Theta((\ln A) \ln n)$.

We would like to have a similar understanding of adaptive cooling schedules. A reasonable conjecture is that the optimal adaptive schedule has length $\Theta(\sqrt{(\ln A) \ln n})$. We will present an adaptive schedule of length $O(\sqrt{\ln A}(\ln n) \ln \ln A)$, which comes reasonably close to our guess (in fact, in our applications we are only off by polylogarithmic factors).

4 Adaptive cooling schedules

In this section, we prove the existence of short adaptive cooling schedules for general partition functions. We now formally state the result (to simplify the exposition we will choose $B = e^2$, the construction works for any B).

Theorem 4.1. *Let Z be a partition function of degree n . Let $A = Z(0)$. Assume that $Z(\infty) \geq 1$. There exists an e^2 -Chebyshev cooling schedule S for Z whose length is at most $4(\ln \ln A)\sqrt{(\ln A) \ln n}$.*

It will be convenient to define $f(\beta) = \ln Z(\beta)$. Some useful properties of f are summarized in the next lemma.

Lemma 4.2. *Let $f(\beta) = \ln Z(\beta)$ where Z is a partition function of degree n . Then (a) f is decreasing, (b) f' is increasing (i. e., f is convex) (c) $f'(0) \geq -n$.*

Recall that an e^2 -Chebyshev cooling schedule for Z is a sequence of inverse temperatures $\beta_0, \beta_1, \dots, \beta_\ell$ such that $\beta_0 = 0, \beta_\ell = \infty$, and

$$\frac{Z(2\beta_{i+1} - \beta_i)Z(\beta_i)}{Z(\beta_{i+1})^2} \leq e^2. \quad (8)$$

Since (8) is invariant under scaling we can, without loss of generality, assume $Z(\infty) = 1$ (or equivalently $a_0 = 1$). Since we assumed $a_0 \geq 1$ the scaling will not increase $Z(0)$.

Let $f(\beta) = \ln Z(\beta)$, so that $f(0) = \ln A$, and $f(\infty) = 0$. The condition (8) is equivalent to

$$\frac{f(2\beta_{i+1} - \beta_i) + f(\beta_i)}{2} - f(\beta_{i+1}) \leq 1. \quad (9)$$

If we substitute $x = \beta_i$ and $y = 2\beta_{i+1} - \beta_i$, the condition can be rewritten as

$$f\left(\frac{x+y}{2}\right) \geq \frac{f(x) + f(y)}{2} - 1.$$

In words, f satisfies approximate concavity. The main idea of the proof is that we do not require this property to hold everywhere but only in a sparse subset of points which will correspond to the cooling schedule. A similar viewpoint is that we will show that f can be approximated by a piecewise linear function g with few pieces, see Figure 1 for an illustration. We form the segments of g in the following inductive, greedy manner. Let γ_i denote the endpoint of the last segment. We then set γ_{i+1} as the maximum value such that the midpoint m_i of the segment (γ_i, γ_{i+1}) satisfies (9) (for $\beta_i = \gamma_i, \beta_{i+1} = m_i$). We now formally state the lemma on the approximation of f by a piecewise linear function.

Lemma 4.3. *Let $f : [0, \gamma] \mapsto \mathbb{R}$ be a decreasing, convex function. There exists a sequence $\gamma_0 = 0 < \gamma_1 < \dots < \gamma_j = \gamma$ such that for all $i \in \{0, \dots, j-1\}$,*

$$f\left(\frac{\gamma_i + \gamma_{i+1}}{2}\right) \geq \frac{f(\gamma_i) + f(\gamma_{i+1})}{2} - 1, \quad (10)$$

and

$$j \leq 1 + \sqrt{(f(0) - f(\gamma)) \ln \frac{f'(0)}{f'(\gamma)}}.$$

Proof :

Let $\gamma_0 := 0$. Suppose that we already constructed the sequence up to γ_i . Let γ_{i+1} be the largest number from the interval $[\gamma_i, \gamma]$ such that (10) is satisfied. Let $m_i = (\gamma_i + \gamma_{i+1})/2$, let $\Delta_i = (\gamma_{i+1} - \gamma_i)/2$, and $K_i = f(\gamma_i) - f(\gamma_{i+1})$.

If $\gamma_{i+1} = \gamma$ then we are done constructing the sequence. Otherwise, by the maximality of γ_{i+1} , we have

$$f(m_i) = \frac{f(\gamma_i) + f(\gamma_{i+1})}{2} - 1. \quad (11)$$

Using the convexity of f and (11) we obtain

$$\begin{aligned} -f'(\gamma_i) &\geq \frac{f(\gamma_i) - f(m_i)}{\Delta} = \frac{K_i + 2}{2\Delta}, \quad \text{and} \\ -f'(\gamma_{i+1}) &\leq \frac{f(m_i) - \gamma_{i+1}}{\Delta} = \frac{K_i - 2}{2\Delta}. \end{aligned} \quad (12)$$

Combining the two equations from (12) we obtain

$$\frac{f'(\gamma_{i+1})}{f'(\gamma_i)} = \frac{-f'(\gamma_{i+1})}{-f'(\gamma_i)} \leq \frac{K_i - 2}{K_i + 2} = 1 - \frac{4}{K_i + 2}. \quad (13)$$

From the second part of (12) and the fact that f is decreasing we obtain $K_i \geq 2$. Hence we can estimate (13) as follows

$$\frac{f'(\gamma_{i+1})}{f'(\gamma_i)} \leq 1 - \frac{4}{K_i + 2} \leq 1 - \frac{1}{K_i} \leq e^{-1/K_i}. \quad (14)$$

Since f is decreasing, we have

$$\sum_{i=0}^{j-2} K_i \leq f(0) - f(\gamma). \quad (15)$$

Now we combine (14) for all $i \in \{0, \dots, j-2\}$ (we use the fact that f' is increasing).

$$\sum_{i=0}^{j-2} \frac{1}{K_i} \leq \ln \frac{f'(0)}{f'(\gamma)}. \quad (16)$$

Applying Cauchy-Schwarz inequality on (15) and (16) we obtain

$$(j-1)^2 \leq (f(0) - f(\gamma)) \ln \frac{f'(0)}{f'(\gamma)}.$$

■

The construction immediately yields a natural cooling schedule. A schedule ending at $\beta_k = \gamma_i$, can now be extended by $\beta_{k+1} = m_i$ where m_i is the midpoint of the segment (γ_i, γ_{i+1}) . Moreover, we can then set β_{k+2} as the midpoint of (m_i, γ_{i+1}) . We continue in this geometric manner for at most $\ln \ln A$ steps, after which we can set the next inverse temperature in our schedule to γ_{i+1} . Then we continue on the next segment. It then follows that the length ℓ of the cooling schedule satisfies $\ell \leq j \ln \ln A$ where j is the length of the sequence from Lemma 4.3. We now present the proof of the Theorem 4.1.

Proof of Theorem 4.1:

Let γ be such that $f(\gamma) = 1$. We describe a sequence $\beta_0 = 0 < \beta_1 < \dots < \beta_\ell = \gamma$ satisfying (9). Note that since $f(\gamma) = 1$, we can take $\beta_{\ell+1} = \infty$ and the sequence will still satisfy (9) (and thus we get a complete e^2 -Chebyshev cooling schedule for Z). We have

$$Z(\gamma) = \exp(f(\gamma)) = \sum_{i=0}^n a_i e^{-i\gamma} = e,$$

and, hence, (using $a_0 = 1$)

$$-Z'(\gamma) = \sum_{i=0}^n i a_i e^{-i\gamma} \geq e - 1.$$

Thus

$$\begin{aligned} -f'(\gamma) &= -\ln Z(\gamma) = \frac{-Z'(\gamma)}{Z(\gamma)} = \\ &= \frac{\sum_{i=0}^n i a_i e^{-i\gamma}}{\sum_{i=0}^n a_i e^{-i\gamma}} \geq \frac{e-1}{e}. \end{aligned}$$

By Lemma 4.3, there exists a sequence of $\gamma_0 = 0 < \gamma_1 < \dots < \gamma_j = \gamma$ of length

$$j \leq 1 + \sqrt{(\ln A) \ln \frac{ne}{e-1}} \quad (17)$$

such that (10) is satisfied.

Now we show how to add $\lceil \ln \ln A \rceil$ inverse temperatures between each pair γ_i and γ_{i+1} to obtain our cooling

schedule. For notational convenience we show this only for $\gamma_0 = 0$ and γ_1 .

Note that (10) implies that (9) is satisfied for $\beta_0 = 0$ and $\beta_1 = \gamma_1/2$. We now show that

$$0, (1/2)\gamma_1, (3/4)\gamma_1, (7/8)\gamma_1, \dots, (1 - 2^{-\lceil \ln \ln A \rceil})\gamma_1, \gamma_1$$

is an e^2 -Chebyshev cooling schedule. Let

$$g(x) = f\left(\frac{\gamma_1 + x}{2}\right) - \frac{f(x) + f(\gamma_1)}{2}.$$

Note that by (11) we have $g(0) = -1$. We have

$$g'(x) = \frac{1}{2} \left(f'\left(\frac{\gamma_1 + x}{2}\right) - f'(x) \right).$$

Thus

$$\text{if } x \leq \gamma_1 \text{ we have } g'(x) \geq 0, \quad (18)$$

and, hence,

$$g(x) \geq g(0) = -1.$$

Plugging in $x = (1 - 2^{-t})\gamma_1$ we conclude

$$f((1 - 2^{-t-1})\gamma_1) \geq \frac{f((1 - 2^{-t})\gamma_1) + f(\gamma_1)}{2} - 1. \quad (19)$$

From (11) and (19) it follows that the sequence

$$0, (1/2)\gamma_1, (3/4)\gamma_1, (7/8)\gamma_1, \dots, (1 - 2^{-t})\gamma_1, \gamma_1 \quad (20)$$

satisfies (9). We will now show that we can truncate the sequence at $t = \lceil \ln \ln A \rceil$ and take the next step to γ_1 .

By the convexity of f

$$f((1 - 2^{-t-1})\gamma_1) \leq \frac{f((2 - 2^{-t})\gamma_1) + f(\gamma_1)}{2},$$

and hence

$$f((1 - 2^{-t-1})\gamma_1) - f(\gamma_1) \leq \frac{f((1 - 2^{-t})\gamma_1) - f(\gamma_1)}{2}. \quad (21)$$

The equation (21) states that the distance of $f((1 - 2^{-t})\gamma_1)$ from $f(\gamma_1)$ halves in each step. Recall that $f(\gamma_1/2) - f(\gamma_1) \leq f(0) \leq \ln A$ and, hence, for $t := \lceil \ln \ln A \rceil$ we have

$$f((1 - 2^{-t})\gamma_1) - f(\gamma_1) \leq 1.$$

This completes the construction of the cooling schedule. The length of the schedule is $\leq jt$. Plugging in (17) yields the theorem. ■

The optimal Chebyshev cooling schedule can be obtained in a greedy manner. In particular, starting with $\beta_0 = 0$, and then from β_i , choosing the maximum β_{i+1} for which (9) is satisfied. The reason why the greedy strategy works is that if we can step from β to β' , then for any $\gamma \in [\beta, \beta']$ we can step from γ to β' (i.e., having large inverse temperature can not hurt us). The last fact follows from the convexity of f (or alternatively from (18)).

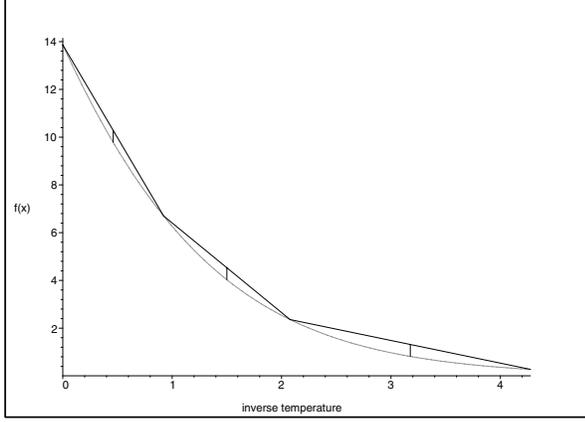


Figure 1. The light curve is $f(x) = \ln Z(x)$ for the partition function $Z(x) = (1 + \exp(-x))^{20}$. The dark curve is a piecewise linear function g consisting of 3 pieces which approximates f . In particular, $g \geq f$ and the midpoint of each piece is close to the average of the endpoints (specifically, (8) holds).

Corollary 4.4. *Let Z be a partition function of degree n . Let $A = Z(0)$. Assume that $Z(\infty) \geq 1$. Suppose that $\beta_0 < \dots < \beta_\ell$ is a cooling schedule for Z . Then the number of indices i for which*

$$\frac{Z(2\beta_{i+1} - \beta_i)Z(\beta_i)}{Z(\beta_{i+1})^2} \geq e^2$$

is at most $4(\ln \ln A)\sqrt{(\ln A)\ln n}$.

5 An adaptive cooling algorithm

The main theorem of the previous section proves the existence of a short adaptive cooling schedule, whereas in Section 3 we proved any non-adaptive cooling schedule is much longer. In this section, we present an adaptive algorithm to find a short cooling schedule. We state the main result before describing the details of the algorithm. The algorithm has access to a sampling oracle, which on input β produces a random sample from the distribution μ_β , defined by (1) (or a distribution sufficiently close to μ_β).

Theorem 5.1. *Let Z be a partition function. Assume that we have access to an (approximate) sampling oracle from μ_β for any inverse temperature β . Let $\delta' > 0$. With probability at least $1 - \delta'$, the following algorithm outputs a B -Chebyshev cooling schedule for Z (with $B = 3 \cdot 10^6$), where the length of the schedule is at most $\ell \leq 38\sqrt{\ln A}(\ln n)\ln \ln A$. The algorithm uses at most*

$Q \leq 10^7(\ln A)((\ln n) + \ln \ln A)^5 \ln \frac{1}{\delta'}$ samples from the μ_β -oracles. The samples output by the oracles have to be from a distribution μ'_β which is within variation distance $\leq \delta'/(2Q)$ from μ_β .

Combining Theorem 5.1 with Corollary 2.3 we obtain.

Corollary 5.2. *Let Z be a partition function. Let $\varepsilon > 0$ be the desired precision. Suppose that we are given access to oracles which sample from the distribution within variation distance $\varepsilon^2/(10^8(\ln A)((\ln n) + \ln \ln A)^5)$ from μ_β for any inverse temperature β . Then, using $\frac{10^{10}}{\varepsilon^2}(\ln A)((\ln n) + \ln \ln A)^5$ samples in total, we can obtain a random variable \hat{S} such that*

$$P((1 - \varepsilon)Z(\infty) \leq \hat{S} \leq (1 + \varepsilon)Z(\infty)) \geq 3/4.$$

5.1 High-level Algorithm Description

We begin by presenting the high-level idea of our algorithm. Ideally we would like to find a sequence $\beta_0 = 0 < \beta_1 < \dots < \beta_\ell = \infty$ such that, for some constants $1 < c_1 < c_2$, for all i , the random variable $W := W_{\beta_i, \beta_{i+1}}$ satisfies

$$c_1 \leq \frac{\mathbb{E}(W^2)}{\mathbb{E}(W)^2} \leq c_2. \quad (22)$$

The upper bound in (22) is necessary so that Chebyshev's inequality guarantees that few samples of W are required to obtain a close estimate of the ratio $Z(\beta_{i+1})/Z(\beta_i)$. On the other side, the lower bound would imply that the length of the cooling schedule is close to optimal. We will guarantee the upper bound for every pair of inverse temperatures, but we will only obtain the lower bound for a sizable fraction of the pairs. Then, using Corollary 4.4, we will argue that the schedule is short.

During the course of the algorithm we will try to find the next inverse temperature β_{i+1} so that (22) is satisfied. For this we will need to estimate $u = u(\beta_i, \beta_{i+1}) := \mathbb{E}(W^2)/\mathbb{E}(W)^2$. We already have an expression for u , given by equation (5):

$$u = \frac{\mathbb{E}(W^2)}{\mathbb{E}(W)^2} = \frac{Z(2\beta_{i+1} - \beta_i)Z(\beta_i)}{Z(\beta_{i+1})^2} = \frac{Z(2\beta_{i+1} - \beta_i)}{Z(\beta_{i+1})} \frac{Z(\beta_i)}{Z(\beta_{i+1})}.$$

Hence, to estimate u it suffices to estimate the ratios $Z(2\beta_{i+1} - \beta_i)/Z(\beta_i)$ and $Z(\beta_i)/Z(\beta_{i+1})$. Recall that the goal of estimating u was to show that W is an efficient estimator of $Z(\beta_{i+1})/Z(\beta_i)$. Now it seems that to estimate u we already need a good estimator for W (with roles of β_i, β_{i+1} switched). An important component of our algorithm, which allows us to escape from this circular loop, is a rough estimator for u which bypasses W .

Recall, the Hamiltonian H takes values in $\{0, 1, \dots, n\}$. For the purposes of estimating u it will suffice to know the Hamiltonian within some relative accuracy. Thus, we partition $\{0, 1, \dots, n\}$ into (discrete) intervals of roughly equivalent values of the Hamiltonian. Since we need relative accuracy the size of the interval is smaller for smaller values of the Hamiltonian (specifically, value i is an interval of size about $i/\sqrt{\ln A}$). We let P denote the set of intervals. We construct P inductively, starting with interval $[0, 0]$. Suppose that $\{0, \dots, b-1\}$ is already partitioned. Let $w := \lfloor b/\sqrt{\ln A} \rfloor$, add the interval $[b, b+w]$ to P , and continue inductively on $\{b+w+1, \dots, n\}$. Note, the initial $\sqrt{\ln A}$ intervals are of size 1 (i.e., contain one natural number), and have width 0. We have the following explicit upper bound on the number of intervals in P .

Lemma 5.3. $|P| \leq 4\sqrt{\ln A} \ln n$.

The rough estimator for u needs an interval $I = [b, c] \subseteq \{1, \dots, n\}$ which contributes a significant portion to $Z(\beta)$ for all $\beta \in [\beta_i, 2\beta_{i+1} - \beta_i]$. Let $h = 1/(8|P|)$.

Definition 5.4. Let Z be a partition function. Let $\beta \in \mathbb{R}^+$ be an inverse temperature. Let $I = [b, c] \subseteq \{0, \dots, n\}$ be an interval. For $h \in (0, 1)$, we say that I is h -heavy for β , if for X chosen from μ_β , we have $\Pr(H(X) \in I) \geq h$.

Thus, if we generate a random sample from μ_β we have a significant probability that the sample is in the interval I .

Lemma 5.5. Given an inverse temperature β , using $s = \lceil (8/h) \ln \frac{1}{\delta} \rceil$ samples from μ_β we can find an h -heavy interval $I \in P$. The failure probability of the procedure is at most $\delta|P|$.

The key observation is that if an interval I is heavy for inverse temperatures β_1 and β_2 , then by generating samples from μ_{β_1} and μ_{β_2} , and looking at the proportion of samples whose Hamiltonian falls into interval I , we can roughly estimate $Z(\beta_2)/Z(\beta_1)$.

Lemma 5.6. Let Z be a partition function. Let $I = [b, c] \subseteq \{0, \dots, n\}$ be an interval. Let $\delta \in (0, 1]$. Suppose that I is h -heavy for inverse temperatures $\beta_1, \beta_2 \in \mathbb{R}^+$. Assume that

$$|\beta_1 - \beta_2| \cdot (c - b) \leq 1.$$

For $k = 1, 2$ we define the following. Let $X_k \sim \mu_{\beta_k}$ and let Y_k be the indicator function for the event $H(X_k) \in I$. Let $s = \lceil (8/h) \ln \frac{1}{\delta} \rceil$. Let U_k be the average of s independent samples from Y_k . Let

$$\text{EST}(I, \beta_1, \beta_2) := \frac{U_1}{U_2} \exp(b(\beta_1 - \beta_2)).$$

With probability at least $1 - 4\delta$ we have

$$\frac{Z(\beta_2)}{4eZ(\beta_1)} \leq \text{EST}(I, \beta_1, \beta_2) \leq \frac{4eZ(\beta_2)}{Z(\beta_1)}.$$

Thus, if an interval I is heavy for an interval of inverse temperatures $B = [\beta_i, \beta^*]$, then we can find a $\beta_{i+1} \in B' = [\beta_i, (\beta_i + \beta^*)/2]$ satisfying (22) (making an optimal move in some sense) or determine there is no such $\beta_{i+1} \in B'$.

If such a $\beta_{i+1} \in B$ satisfying (22) exists, then we can set β_{i+1} as the next temperature and continue the algorithm by inductively considering the interval $[\beta_{i+1}, \beta^*]$, in which I is still heavy. We call the intermediate inverse temperature β_{i+1} an ‘‘optimal’’ step since the number of such temperatures in our cooling schedule is upper bounded by Corollary 4.4.

In the case that no such β_{i+1} exists, we construct a sequence of inverse temperatures that goes from β_i to β^* where the upper bound in (22) holds for this sequence. We will show that $O(\ln \ln A)$ intermediate inverse temperatures are sufficient to go from β_i to β^* (the construction is analogous to the sequence (20) in the proof of Theorem 4.1). We refer to these $O(\ln \ln A)$ intermediate inverse temperatures as ‘‘interval’’ steps since they are used to finish off an interval and are not optimal in the sense of Corollary 4.4. An important fact is that for an interval I , the set of β 's where I is heavy is itself an interval.

Lemma 5.7. Let Z be a partition function. Let $I = [b, c] \subseteq \{0, \dots, n\}$ be an interval. Let $h \in (0, 1]$. The set of inverse temperatures for which I is h -heavy forms an interval (possibly empty).

Hence, once we reach β^* we will be done with this interval I and will not need to consider it again. Therefore, the number of interval steps is at most $O(|P| \ln \ln A)$.

Finally, our algorithm will find a cooling schedule whose length is at most

$$O\left((\ln \ln A)\sqrt{(\ln A) \ln n} + \sqrt{\ln A}(\ln n) \ln \ln A\right), \quad (23)$$

where the first term comes from Corollary 4.4 and the second term is $O(|P| \ln \ln A)$.

To simplify the high-level exposition of the algorithm we glossed over a technical aspect of the algorithm. The interval B might be too long so that the estimator of u is too rough (since the range for which the rough estimator of Lemma 5.6 works is bounded by the reciprocal of the width of the interval I). Therefore if B is too long, we truncate the interval B into a subinterval $B_1 = [\beta_i, \beta_i + 1/w]$ where $w = b - c$ is the width of interval $I = [b, c]$. We first consider the subinterval B_1 . If we find an optimal step within B_1 , then we add this inverse temperature β_{i+1} as the next step in the cooling schedule and continue the algorithm by inductively considering the interval $[\beta_{i+1}, \beta^*]$. Alternatively if there is no optimal step within B_1 we finish off the subinterval B_1 using $O(\ln \ln A)$ steps. We refer to these moves as ‘‘long’’ steps, since they no longer finish off the interval I , and after finishing B_1 we continue the algorithm by inductively considering the interval $[\beta_i + 1/w, \beta^*]$.

Long steps will be analyzed by a separate argument, and their number will be smaller than (23). Thus, in the detailed description of the algorithm we will have three kinds of steps: optimal steps, interval steps, and long steps.

5.2 Detailed Algorithm Description

We now give a detailed description of our algorithm for constructing the cooling schedule. Let δ' be the desired final error probability of our algorithm. We will use the algorithms implicitly described in Lemmas 5.5 and 5.6 with parameters $\delta = \frac{\delta'}{1600(\ln n)^2(\ln A)^2}$ and $s = \lceil (8/h) \ln \frac{1}{\delta} \rceil$. Certain technical details of the algorithm are omitted, as well as the analysis of the running time and proof of correctness. We encourage the interested reader to refer to [21].

The algorithm will keep a set Bad of banned intervals which is initially empty.

Note it suffices to have the penultimate β in the sequence be $\beta_{i-1} = \ln A$, since we can then set $\beta_i = \infty$. The algorithm for constructing the sequence works inductively. Thus, consider some starting β_0 .

1. We first find an interval I that is h -heavy at β_0 and is not banned. By generating s samples from the distribution μ_{β_0} and taking the most frequently seen interval, we will successfully find an h -heavy interval with high probability.
2. Let w denote the width of I , i. e., $w = c - b$ where $I = [b, c]$. Our rough estimator (given by Lemma 5.6) only applies for $\beta_1 \leq \beta_0 + 1/w$ (by convention $1/0 = \infty$). Moreover, since we only need to reach a final inverse temperature of $\ln A$, let $L = \min\{\beta_0 + 1/w, \ln A\}$. Now we concentrate on constructing a cooling schedule within $(\beta_0, L]$.
3. We do binary search in the interval $[\beta_0, L]$ to find the maximum β^* such that β^* is h -heavy. We can use binary search because, by Lemma 5.7, the set of inverse temperatures for which an interval is heavy is an interval in \mathbb{R}^+ .
4. We now check if there is an “optimal” move within the interval $B' = (\beta_0, (\beta_0 + \beta^*)/2]$. We want to find the maximum $\beta \in B'$ satisfying (22) for $u(\beta_0, \beta)$, or determine no such β exists. Let $c_1 = e^2$ and $c_2 = 3 \cdot 10^6$ for (22). To find such a β , we do a binary search and apply Lemma 5.6 to estimate the ratios $Z(2\beta - \beta_0)/Z(\beta)$ and $Z(\beta_0)/Z(\beta)$. Note for $\beta \in B'$ we have $2\beta - \beta_0 \in [\beta_0, \beta^*]$, hence, the interval I is h -heavy at inverse temperatures β_0, β and $2\beta - \beta_0$ and Lemma 5.6 applies.
 - (a) If such a $\beta \in B'$ exists, then we set β as the next inverse temperature and we repeat the algorithm

starting from β . We refer to these steps as “optimal” moves.

- (b) If no such β exists, then we can reach the end of the interval B as follows. There are two cases, either the interval was too wide for the application of Lemma 5.6, or the interval I stops being heavy too soon. More precisely, either:
 - i. If $\beta^* = L$, then we set $(\beta_0 + \beta^*)/2$ as the next inverse temperature. Moreover, if $\beta^* < \ln A$ we continue the algorithm starting from β^* ; whereas if $\beta^* = \ln A$ we are done. We refer to these steps as “long” moves.
 - ii. Otherwise, we add the following inverse temperatures to our schedule:

$$\beta_0 + \frac{1}{2}\gamma, \beta_0 + \frac{3}{4}\gamma, \beta_0 + \frac{7}{8}\gamma, \dots, \\ \beta_0 + (1 - 2^{-t})\gamma, \beta_0 + \gamma,$$

where $\gamma = \beta^* - \beta_0$ and $t = \lceil \ln \ln A \rceil$. We add the interval I to the set of banned intervals Bad and continue the algorithm starting from β^* . We refer to these steps as “interval” moves since the interval I will not be used by the algorithm again.

6 Applications

We detail several specific applications of our work: matchings, Ising model, colorings and independent sets. To simplify the comparison of our results with previous work and since we have not optimized polylogarithmic factors in our work, we use $O^*(\cdot)$ notation which hides polylogarithmic terms and the dependence on ϵ . Our cooling schedule results in a savings of a factor of $O^*(n)$ in the running time for all of the approximate counting problems considered here.

6.1 Matchings

Jerrum and Sinclair [11] presented a Markov chain for sampling a random matching of an arbitrary input graph $G = (V, E)$. They proved the chain has relaxation time $\tau_2 = O(nm)$, where n is the number of vertices and m is the number of edges of G (see [9] for the claimed upper bound). Our work yields a cooling schedule of length $\ell = O(\sqrt{n} \log^4 n)$. To be precise, this requires what we refer to as a “reversible” cooling schedule to utilize the notion of “warm-starts” (this is carried out in detail in [21]). The previous best schedule was presented by [2] which had length $O(n \log^2 n)$. Thus, we save a factor of $O^*(n)$ in the running time for approximately counting the number of matchings of G .

Corollary 6.1. For any $G = (V, E)$, for all $\varepsilon > 0$, let $\mathcal{M}(G)$ denote the set of matchings of G . We can compute an estimate EST such that: $\text{EST}(1 - \varepsilon) \leq |\mathcal{M}(G)| \leq \text{EST}(1 + \varepsilon)$ with probability $\geq 3/4$ in time $O(n^2 m \varepsilon^{-2} \log^7 n) = O^*(n^2 m)$.

Recall, the error probability $3/4$ can be replaced by $1 - \delta$, for any $\delta > 0$, at the expense of an extra factor of $O(\log(1/\delta))$ in the running time.

6.2 Spin Systems

Spin systems are a general class of statistical physics models where our results apply. We refer the reader to [18, 24] for an introduction to spin systems. The examples we highlight here are well-studied examples of spin systems. Recall, the mixing time of a Markov chain is the number of transitions (from the worst initial state) to reach within variation distance $\leq \delta$ of the stationary distribution, where $0 < \delta < 1$. The following results follow in a standard way from the stated mixing time result combined with Corollary 5.2.

Colorings: For a graph $G = (V, E)$ with maximum degree Δ we are interested in approximating the number of k -colorings of G . Here, we are coloring the vertices using a palette of k colors so that adjacent vertices receive different colors. This problem is also known as the zero-temperature (thus $\beta = \infty$) anti-ferromagnetic Potts model. The simple single-site update Markov chain known as the Glauber dynamics is ergodic with unique stationary distribution uniform over all k -colorings whenever $k \geq \Delta + 2$. There are various regions where fast convergence of the Glauber dynamics is known, we refer the interested reader to [7] for a survey. For concreteness we consider the result of Jerrum [10] who proved that the Glauber dynamics has mixing time $O(kn \log(n/\delta))$ whenever $k > 2\Delta$. Moreover, his proof easily extends to any non-zero temperature. Since $A = k^n$, using Corollary 5.2 we obtain the following result.

Corollary 6.2. For all $k > 0$, any graph $G = (V, E)$ with maximum degree Δ , let $\Omega(G)$ denote the set of k -colorings of G . For all $\varepsilon > 0$, whenever $k > 2\Delta$, we can compute an estimate EST such that $\text{EST}(1 - \varepsilon) \leq |\Omega(G)| \leq \text{EST}(1 + \varepsilon)$ with probability $\geq 3/4$ in time $O(kn^2 \varepsilon^{-2} \log^6 n) = O^*(n^2)$.

In comparison, the previous bound [2] required $O^*(n^3)$ time (and Jerrum [10] required $O^*(nm^2)$ time).

Ising model: There are extensive results on sampling from the Gibbs distribution and approximating the partition function of the (ferromagnetic) Ising model. We refer the reader to [18] for background and a survey of results. We consider a particularly well-known result. For the Ising model on an $\sqrt{n} \times \sqrt{n}$ 2-dimensional grid, Martinelli and

Olivieri [19] proved that the Glauber dynamics (i. e., single-site update Markov chain) has mixing time $O(n \log(n/\delta))$ for all $\beta > \beta_c$ where β_c is the critical point for the phase transition between uniqueness and non-uniqueness of the infinite-volume Gibbs measure. In this setting, we have $A = 2^n$ and, hence, we obtain the following result.

Corollary 6.3. For the Ising model on a $\sqrt{n} \times \sqrt{n}$ 2-dimensional grid, let $Z(\beta)$ denote the partition function at inverse temperature $\beta > 0$. For all $\varepsilon > 0$, for all $\beta > \beta_c$, we can compute an estimate EST such that $\text{EST}(1 - \varepsilon) \leq Z(\beta) \leq \text{EST}(1 + \varepsilon)$ with probability $\geq 3/4$ in time $O(n^2 \varepsilon^{-2} \log^6 n) = O^*(n^2)$.

Independent Sets: Given a fugacity $\lambda > 0$ and a graph $G = (V, E)$ with maximum degree Δ , we are interested in computing

$$Z_G(\lambda) = \sum_{\sigma \in \Omega} \lambda^{|\sigma|},$$

where Ω is the set of independent sets of G . This is known as hard-core lattice gas model. In [23, 5], it was proved that the Glauber dynamics for sampling from the distribution corresponding to $Z_G(\lambda)$ has $O(n \log(n/\delta))$ mixing time whenever $\lambda < 2/(\Delta - 2)$. As a consequence, we obtain the following result.

Corollary 6.4. For any graph $G = (V, E)$ with maximum degree Δ , for all $\varepsilon > 0$, for any $\lambda < 2/(\Delta - 2)$, we can compute an estimate EST such that: $\text{EST}(1 - \varepsilon) \leq Z_G(\lambda) \leq \text{EST}(1 + \varepsilon)$ with probability $\geq 3/4$ in time $O(n^2 \varepsilon^{-2} \log^6 n) = O^*(n^2)$.

Note, Weitz [25] has an alternative approach for this problem. His approach approximates $Z_G(\lambda)$ directly (without using sampling) and holds for a larger range of λ (though Δ is required to be constant).

7 Lower bound for adaptive cooling

Lemma 7.1. Let $n \geq 1$. Consider the following partition function of degree n :

$$Z(\beta) = (1 + e^{-\beta})^n.$$

Any B -Chebyshev cooling schedule for $Z(\beta)$ has length at least $\sqrt{n/(20 \ln B)}$.

Proof :

Let $f(\beta) = \ln Z(\beta) = n \ln(1 + e^{-\beta})$. If the current inverse temperature is $\beta_i =: \beta$, the next inverse temperature $\beta_{i+1} =: \beta + x$ has to satisfy

$$f(\beta) + f(\beta + 2x) - 2f(\beta + x) \leq \ln B.$$

Later we will show that for any $\beta \in [0, 1]$ and $x \in [0, 1]$ we have

$$f(\beta) + f(\beta + 2x) - 2f(\beta + x) \geq \frac{n}{20} x^2. \quad (24)$$

From (24) it follows that for $\beta \leq 1$ the inverse temperature increases by at most

$$x \leq \sqrt{\frac{20 \ln B}{n}},$$

and, hence, the length of the schedule is at least $\sqrt{n/(20 \ln B)}$.

It remains to show (24). Let

$$g(x, \beta) := \frac{f(\beta) + f(\beta + 2x) - 2f(\beta + x)}{2n}.$$

We have

$$\frac{\partial}{\partial x} g(x, \beta) = \frac{e^{-\beta-x}}{1 + e^{-\beta-x}} - \frac{e^{-\beta-2x}}{1 + e^{-\beta-2x}}.$$

We will show

$$\frac{e^{-\beta-x}}{1 + e^{-\beta-x}} - \frac{e^{-\beta-2x}}{1 + e^{-\beta-2x}} \geq x/20, \quad (25)$$

which will imply (24) (by integration over x).

Let $C := e^{-\beta}$ and $y := 1 - e^{-x}$. Note that $C \in [1/e, 1]$, $y \in [0, 1 - 1/e]$, and $x = -\ln(1 - y)$. For $y \in [0, 1 - 1/e]$ we have $-\ln(1 - y) \leq y + y^2$ and hence it is enough to show

$$\frac{C(1 - y)}{1 + C(1 - y)} - \frac{C(1 - y)^2}{1 + C(1 - y)^2} \geq \frac{1}{20}(y + y^2). \quad (26)$$

Multiplying both sides by the numerators we obtain that (26) is equivalent to

$$P(y, C) := y(y + 1)(y - 1)^3 C^2 - (y^4 - 2y^3 + 19y^2 - 18y)C - (y^2 + y) \geq 0.$$

The polynomial $y(y + 1)(y - 1)^3$ is negative for our range of y and hence for any fixed y , the minimum of $P(y, C)$ over $C \in [1/3, 1]$ occurs either at $C = 1$ or at $C = 1/3$ (we only need to show positivity of $P(y, C)$ for $C \in [1/e, 1]$, but for numerical convenience we show it for a larger interval). We have

$$p(y, 1) = y^5 - 3y^4 + 2y^3 - 18y^2 + 16y, \quad (27)$$

and

$$9p(y, 1/3) = y^5 - 5y^4 + 6y^3 - 64y^2 + 44y. \quad (28)$$

Both (27) and (28) are non-negative for our range of y (as is readily seen by the method of Sturm sequences). This finishes the proof of (25), which in turn implies (24). ■

8 Discussion

An immediate question is whether these results extend to estimating the permanent of a 0/1 matrix. Our current adaptive scheme works assuming a sampling subroutine that can produce samples at any given temperature (at least from a warm start). The permanent algorithm of [12] also requires a set of $n^2 + 1$ weights to produce samples from a given temperature. These weights are computed from $n^2 + 1$ partition functions and it appears that a schedule of length $\Omega(n)$ is necessary if one considers all $n^2 + 1$ partition functions simultaneously. In fact, this is the case for the standard bad example of a chain of boxes (or a chain of hexagons as illustrated in Figure 2 of [12]).

References

- [1] László Babai, Monte-Carlo algorithms in graph isomorphism testing, Université tde Montréal Technical Report, DMS 79-10, 1979 (42), see also <http://people.cs.uchicago.edu/laci/lasvegas79.pdf>.
- [2] I. Bezáková, D. Štefankovič, V. Vazirani, and E. Vigoda, Accelerating simulated annealing for combinatorial counting. In *Proceedings of the 17th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 900–907, 2006.
- [3] M. E. Dyer and A. Frieze. Computing the volume of a convex body: a case where randomness provably helps. In *Proceedings of AMS Symposium on Probabilistic Combinatorics and Its Applications*, 123–170, 1991.
- [4] M.E. Dyer, A.M. Frieze, and R. Kannan, A random polynomial time algorithm for approximating the volume of convex bodies. *Journal of the ACM*, 38(1):1–17, 1991.
- [5] M. E. Dyer and C. Greenhill, On Markov Chains for Independent Sets. *J. Algorithms*, 35(1): 17–49, 2000.
- [6] A. M. Ferrenberg and R. H. Swendsen, *Physical Review Letters*, 61, 2635–2638, 1988.
- [7] A. Frieze and E. Vigoda, A survey on the use of Markov chains to randomly sample colorings. *Combinatorics, Complexity and Chance*, Oxford University Press, 53–71, 2007.
- [8] S. Janson, T. Łuczak, and A. Ruciński, *Random Graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization, 2000.

- [9] M. Jerrum, *Counting, sampling and integrating: algorithms and complexity*. Lectures in Mathematics, Birkhäuser Verlag, 2003.
- [10] M. Jerrum, A very simple algorithm for estimating the number of k -colorings of a low-degree graph. *Random Structures and Algorithms*, 7(2):157–165, 1995.
- [11] M. Jerrum and A. Sinclair, Approximating the permanent. *SIAM Journal on Computing*, 18:1149–1178, 1989.
- [12] M. Jerrum, A. Sinclair, and E. Vigoda, A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries. *Journal of the ACM*, 51(4):671–697, 2004.
- [13] M. Jerrum, L. Valiant, and V. Vazirani, Random generation of combinatorial structures from a uniform distribution. *Theoretical Computer Science*, 43(2-3):169–188, 1986.
- [14] A. Kalai and S. Vempala, Simulated Annealing for Convex Optimization. *Mathematics of Operations Research*, 31(2), 2006, 253–266.
- [15] R. Kannan, L. Lovász, and M. Simonovits, Random walks and an $O^*(n^5)$ volume algorithm for convex bodies. *Random Structures and Algorithms* 11, 1–50, 1997.
- [16] L. Lovász and S. Vempala, Simulated annealing in convex bodies and an $O^*(n^4)$ volume algorithm. *Journal of Computer and System Sciences*, 72(2):292–417, 2006.
- [17] L. Lovász and S. Vempala, Fast algorithms for log-concave functions: sampling, rounding, integration and optimization. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 57–68, 2006.
- [18] F. Martinelli, Relaxation times of Markov chains in statistical mechanics and combinatorial structures, *Encyclopedia of Mathematical Sciences*, Vol. 110, Springer, 2003.
- [19] F. Martinelli and E. Olivieri, Approach to equilibrium of Glauber dynamics in the one phase region I: The attractive case, *Communications in Mathematical Physics*, 161:447–486, 1994.
- [20] Z. W. Salsburg, J.D. Jacobson, W. Fickett, and W. W. Wood, Application of the Monte Carlo method to the lattice-gas model, *The Journal of Chemical Physics*, 30(1):65–72, 1959.
- [21] D. Štefankovič, S. Vempala, and E. Vigoda, Adaptive Simulated Annealing: A Near-optimal Connection between Sampling and Counting. Available from arXiv at: <http://arxiv.org/abs/cs.DS/0612058>
- [22] J. P. Valleau and D. N. Card, Monte Carlo Estimation of the Free Energy by Multistage Sampling, *The Journal of Chemical Physics*, 57(12):5457–5462, 1972.
- [23] E. Vigoda, A note on the Glauber dynamics for sampling independent sets. *Electronic Journal of Combinatorics*, 8(1), 2001.
- [24] D. Weitz, Mixing in time and space for discrete spin systems, Ph.D. thesis, U.C. Berkeley, May 2004. Available from <http://dimacs.rutgers.edu/~dror/thesis/thesis.pdf>
- [25] D. Weitz, Counting independent sets up to the tree threshold. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC)*, 140–149, 2006.