

Simulated Quantum Annealing Can Be Exponentially Faster than Classical Simulated Annealing

Elizabeth Crosson* Aram W. Harrow†

January 13, 2016

Abstract

Simulated Quantum Annealing (SQA) is a Markov Chain Monte-Carlo algorithm that samples the equilibrium thermal state of a Quantum Annealing (QA) Hamiltonian. In addition to simulating quantum systems, SQA has also been proposed as another physics-inspired classical algorithm for combinatorial optimization, alongside classical simulated annealing. However, in many cases it remains an open challenge to determine the performance of both QA and SQA.

One piece of evidence for the strength of QA over classical simulated annealing comes from an example by Farhi, Goldstone and Gutmann [14]. There a bit-symmetric cost function with a thin, high energy barrier was designed to show an exponential separation between classical simulated annealing, for which thermal fluctuations take exponential time to climb the barrier, and quantum annealing which passes through the barrier and reaches the global minimum in poly time, arguably by taking advantage of quantum tunneling. In this work we apply a comparison method to rigorously show that the Markov chain underlying SQA efficiently samples the target distribution and finds the global minimum of this spike cost function in polynomial time.

Our work provides evidence for the growing consensus that SQA inherits at least some of the advantages of tunneling in QA, and so QA is unlikely to achieve exponential speedups over classical computing solely by the use of quantum tunneling. Since we analyze only a particular model this evidence is not decisive. However, techniques applied here – including warm starts from the adiabatic path and the use of the quantum ground state probability distribution to understand the stationary distribution of SQA – may be valuable for future studies of the performance of SQA on cost functions for which QA is efficient.

1 Introduction

Classical algorithms are often useful but not provably so, with justifications for their success coming from a combination of empirical and heuristic evidence. For example, the simplex algorithm for linear programming was successful for decades before being proven to run in polynomial time, and for a long time was the most practical LP solver even while the ellipsoid algorithm was the only provably poly-time solver. Another example is MCMC (Markov chain Monte Carlo) which is used for applications in statistics, simulation, optimization and elsewhere, but almost never in regimes that are covered by formal proofs of correctness.

With quantum algorithms, there has been necessarily a greater emphasis on provable correctness. The present state of quantum computing technology does not yet allow us to test

*Caltech IQIM. crosson@caltech.edu

†MIT CTP. aram@mit.edu

large-scale quantum algorithms empirically, nor can we usually empirically determine whether a proposed quantum algorithm outperforms all classical algorithms on worst-case inputs. Nevertheless, heuristic quantum algorithms are likely to be important for practical problems, just as they have been throughout the history of classical computing.

A particularly compelling heuristic proposal for optimization problems is quantum annealing (QA), also known as quantum adiabatic optimization [20, 15]. (In this work we use the term “quantum annealing” to mean adiabatic optimization in thermal equilibrium at a low but non-zero temperature, though in some other contexts QA may be taken to include non-equilibrium thermal effects.) The idea of QA is to interpolate between a static problem-independent Hamiltonian such as $-\sum_i \sigma_x^i$ for which we can efficiently prepare the ground state, and a final Hamiltonian whose ground state yields the desired answer. If we want to minimize a function $f : \{0, 1\}^n \rightarrow \mathbb{R}$ then we can take this final Hamiltonian to be proportional to $\text{diag}(f)$. This can be thought of as a quantum version of classical simulated annealing (SA) with the diagonal terms playing the role of bias and the off-diagonal terms causing hopping. Like SA its performance is hard to make provable general statements about, but it is a promising general-purpose heuristic, and rigorous statements about its performance are known for many illustrative cases.

Intriguingly, QA has been shown to have an exponential asymptotic advantage over simulated annealing for certain cost functions [14]. Two examples are given in [14]: one in which the quantum algorithm could be said to be taking advantage of symmetry (“the bush of implications”), and another which models tunneling (“the spike”) that will be our primary focus here. The cost function for the spike is,

$$f(z) = \begin{cases} |z| & : |z| \neq n/4 \\ |z| + n^\alpha & : |z| = n/4 \end{cases}, \quad (1)$$

where $|z|$ is the Hamming weight of the string z and $\alpha > 0$ is a constant, independent of the system size n . The global minimum of f is the string with $|z| = 0$, but the spike creates a local minimum at $|z| = n/4 + 1$. It is straightforward to see why the spike presents a problem for a simulated annealing algorithm which only proposes single bit-flip moves: SA begins at high temperature and so the initial state is overwhelmingly likely to have Hamming weight near $n/2$, and as the temperature of the system is lowered the random walk will move to strings of lower Hamming weight until reaching the local minimum at $\frac{n}{4} + 1$, at which point it can only step on the spike with probability $e^{-\Omega(n^\alpha)}$, so with high probability SA will not find the global minimum in time less than $e^{\Omega(n^\alpha)}$. In contrast, for $\alpha < 1/2$ it can be shown that QA finds the global minimum with high probability in time $\mathcal{O}(n)$ [25], showing that an exponential separation in the performance of SA and QA is possible.

While the spike is clearly a toy problem and can be solved efficiently by classical algorithms that exploit its structure, an important aspect of both QA and SA is that a single, general implementation of these algorithms is meant to be useful for solving a large variety of different problems without knowledge of their structure. Moreover, the spike arguably demonstrates a general advantage of QA over SA in tunneling through thin, high barriers in the energy landscape.

On the other hand, the standard formulation of QA uses a stoquastic Hamiltonian (i.e. a local Hamiltonian with non-positive matrix elements in the computational basis), and computational models based on ground states or thermal states of such systems are believed to be less powerful than universal quantum computation. In addition to complexity theoretic evidence [7, 6], suggestive evidence for this belief is also provided by the quantum-to-classical mapping of Suzuki et al. [28, 27], which allows for equilibrium thermal states of a stoquastic Hamiltonian to be sampled using classical Markov chain Monte Carlo methods. These algorithms are known as Quantum Monte Carlo (QMC) methods, and despite the name, are algorithms for classical computers. While QMC for stoquastic Hamiltonians is always a well-defined algorithm, few general conditions are known to guarantee that it simulates their properties efficiently. A few cases

where the simulation can be made provably efficient are adiabatic evolution with frustration-free stoquastic Hamiltonians with a unique ground state [8] and ferromagnetic transverse Ising models in a large range of temperatures [5], but when these results are interpreted in terms of QA the corresponding cost functions do not exhibit prototypical features of hard optimization problems.

When QMC is applied to QA Hamiltonians the result is an algorithm called simulated quantum annealing (SQA). Although there are examples for which standard versions of SQA take exponentially longer than the quantum evolution being simulated [16], a the general challenge from SQA to QA remains: for any purported speedup of QA we should see whether it can also be achieved by SQA. Moreover, since SQA is a Markov chain based algorithm on a domain that can be interpreted as a classical spin system, and since SQA is designed to sample from the output of a quantum optimization procedure, SQA can be considered as yet-another physics-inspired classical optimization method in its own right, which can naturally be compared to SA.

The main result of this paper is that the standard version of SQA, which uses only single bit flip local moves and does not use any structure of the problem, finds the minimum of the cost function (1) in polynomial-time when $0 < \alpha < 1/2$. Thus SQA obtains an exponential speedup over SA for this particular problem. This result suggests that the benefit of adiabatic evolution in tunneling through barriers should not be thought of as an exclusively quantum advantage, since it can also be achieved by a general-purpose classical optimization algorithm.

Previous Work

There have been many past studies comparing the performance of SA and SQA using numerics [23, 1, 17] and more recently using analytical methods of physics such as the instanton approximation to tunneling [18]. Studies comparing QA to SQA have also begun to emerge since [2] found the success probabilities of SQA are highly correlated with the results of QA performed on D-Wave quantum hardware with hundreds of qubits, while the distribution of success probabilities for SA on the same set of instances bears little resemblance to that of QA and SQA. More recently, the performance of QA, SQA, and SA was empirically compared on an ensemble of spin glass instances with were designed to have tall, thin barriers [12], as a step towards understanding the kinds of instances for which QA has an advantage over SA. In that work QA and SQA were found to have roughly the same scaling with system size for that particular ensemble of instances, though it was also pointed out that the large constant overhead in SQA made it less competitive in the sense of wall-clock times using modern classical hardware.

Without access to quantum hardware, comparison of SQA and QA is either limited to small system sizes where QA Hamiltonians can be exactly diagonalized (≈ 30 qubits), or to models for which analytical solutions of the quantum system are known (such as the spike problem we study here, which has been the subject of recent analytic work [21, 4], and for which the performance of SQA has previously been studied numerically [10, 3], with findings that are consistent with the results proven in the present work).

Proof Outline

Our proof of the efficient convergence of SQA on the spike problem involves bounding the mixing time of the underlying Markov chain, and there is an interesting parallel between a method which was used to lower bound the QA spectral gap when $\alpha < 1/2$ [25]. There, a lower bound on the quantum gap can be found using a variational method with a trial wave function equal to the ground state of the system when no spike term is present (i.e. QA for the spikeless Hamming weight cost function $\tilde{f}(z) = |z|$). Similarly, we compare the spectral gap λ of the SQA Markov chain for the spike system with the spectral gap λ of the spikeless system (throughout the subsequent sections we use tildes to distinguish quantities belonging to the spikeless system). Without a spike term, the quantum Hamiltonian \tilde{H} is a tensor product

operator with no interactions between the qubits. This trivial system translates in SQA to a collection of Markov Chains with local moves and Metropolis transition probabilities acting on n non-interacting 1D classical ferromagnetic Ising models in a uniform magnetic field (which will become clear when the SQA Markov chain is described in detail in Section 2.2), and upper bounding the mixing time for this system is relatively straightforward.

Let π and $\tilde{\pi}$ be the stationary distribution of the SQA Markov chain with and without the spike. These stationary distributions are close in a sense, $\|\pi - \tilde{\pi}\|_1 < \text{poly}(n^{-1})$, but on the other hand there are exponentially many points $x \in \Omega$ for which the ratio $\pi(x)/\tilde{\pi}(x)$ is exponentially small. A review of existing comparison techniques in Section 3.1 concludes that none is quite suited to the present problem; indeed the review [13] states that there have been “relatively few successes in comparing chains with very different stationary distributions”. To overcome this we introduce a novel comparison method which involves partitioning the state space into “good” and “bad” sets of vertices, $\Omega = \Omega_G \cup \Omega_B$. In Section 3.2 we begin with a set of canonical paths yielding a bound $\tilde{\rho}$ on the congestion of the easy-to-analyze chain, and show that the paths which lie entirely within Ω_G can be used to construct an upper bound on the congestion ρ of the difficult-to-analyze chain, albeit within the set Ω_G of measure less than 1. This “most-paths comparison” method may be of independent interest, and so the exposition in Section 3.2 is given without dependence on the specific details of SQA.

The bound on the congestion derived in this way applies to the Markov chain P on the subset Ω_G , but since P is not restricted to Ω_G we are subsequently left with a sub-stochastic “leaky” random walk on the set Ω_G , with a quasi-stationary distribution equal to π within this subset. The adiabatic path is used to provide a “warm start” within Ω_G , and the mixing time to the quasi-stationary distribution is shown to be much faster than the time scale at which the walk leaves Ω_G , which is the basis for the statement that the convergence of SQA is efficient.

Finally, after the development of the preceding mixing machinery the remaining task is to bound the size of $\pi(\Omega_B)$. The SQA state space can be interpreted as a path (worldline) representation of the original quantum system, and the bad states which constitute Ω_B will be those for which the paths spend too much “time” on the location of the spike (i.e. strings with Hamming weight $n/4$). States that spend too much time on the spike are those for which $\pi(x)/\tilde{\pi}(x)$ is exponentially small, and naturally those are the ones we will need to exclude. In Section 4 we show that the mean spike time is proportional to the square of the ground state amplitude on the spike, while the m -th moment of the spike time distribution can also be bounded using the properties of the corresponding quantum system. Finally, we use the derived upper bound on the m -th moment of the spike time distribution to upper bound the probability of large deviations from the mean spike time, which yields an upper bound on $\pi(\Omega_B)$ that suffices to complete the proof.

2 Background

2.1 Quantum annealing

Quantum annealing associates a cost function $f : \{0,1\}^n \rightarrow \mathbb{R}$ with a Hamiltonian that is diagonal in the computational basis,

$$H_f = \sum_{z \in \{0,1\}^n} f(z) |z\rangle\langle z| \quad , \quad (2)$$

so that the ground state of H_f is a computational basis state corresponding to the bit string that minimizes f . To prepare the ground state of H_f the system is initialized in the ground

state of a uniform transverse field, which can be easily prepared,

$$H_0 = - \sum_{i=1}^n \sigma_i^x \quad , \quad |\psi_{\text{init}}\rangle = \frac{1}{\sqrt{2^n}} \sum_{z \in \{0,1\}^n} |z\rangle, \quad (3)$$

and then linearly interpolates between H_0 and H_f ,

$$H := H(s) = (1-s)H_0 + sH_f \quad , \quad (4)$$

where the adiabatic parameter s sweeps through the interval $0 \leq s \leq 1$. The total run time T of the algorithm depends on how quickly the adiabatic parameter is adjusted, which defines a time-dependent Hamiltonian $H(t) := H(s = t/T)$. At zero temperature the system evolves according to the Schrodinger equation, $i\partial_t|\psi(t)\rangle = H(t)\psi(t)$, and the adiabatic theorem ensures that the state $\psi(T)$ at the end of the evolution has a high overlap with the ground state of H_f as long as $T \geq \text{poly}(\Delta^{-1})$, where $\Delta = \min_s E_1(s) - E_0(s)$ is the minimum gap between the two lowest eigenvalues of $H(s)$ during the evolution.

More realistically the annealing will be performed at a finite inverse temperature β , which is taken to be sufficiently large so that the system will remain close to the ground state. The equilibrium thermal state of the system evolves with the adiabatic parameter,

$$\sigma = \frac{e^{-\beta H(s)}}{\mathcal{Z}} \quad , \quad \mathcal{Z} = \text{tr} e^{-\beta H(s)}. \quad (5)$$

Of course, thermal equilibrium is an idealization that is never achieved by real physical systems, however, the formulation (5) is frequently considered and it is convenient for our purposes since the simulated quantum annealing algorithm described in the next section will produce samples from the distribution $\Pi(z) = \langle z|\sigma|z\rangle$. The minimum gap Δ of the Hamiltonian (4) with the cost function (1) is constant when $0 \leq \alpha < 1/2$ [25], and together with facts about the density of states we will show that taking $\beta = n^\epsilon$ for a constant $\epsilon > 0$ suffices to make $\|\rho(s) - |\psi_0(s)\rangle\langle\psi_0(s)|\|_1 < 1/\text{poly}(n)$, and so this low-temperature thermal equilibrium version of QA efficiently produces a final state which can be sampled to obtain the minimum of f .

2.2 Simulated quantum annealing

Hamiltonians such as (4) which have all non-negative matrix elements in the computational basis are sometimes called “stoquastic”, a term which combines “quantum” with “stochastic” (in the sense of stochastic matrices). In principle such Hamiltonians are amenable to a variety of classical Markov chain based simulation algorithms, which are collectively known as quantum Monte Carlo (QMC) methods. Any QMC method applied to the QA Hamiltonian (4) can be considered to define a version of SQA. Here we considered QMC based on the path-integral representation of the thermal state (5), which is arguably the most commonly considered version of SQA.

The starting point of the method is to express the quantum partition function as a path integral over trajectories (finite sequences) of basis states. Since the Hamiltonian is stoquastic in the computational basis, these trajectories will have the form (x_1, \dots, x_L) , where $x_i \in \{0, 1\}^n$,

$$\mathcal{Z} = \text{tr} e^{-\beta H} = \text{tr} \prod_{i=1}^L e^{-\frac{\beta H}{L}} = \sum_{x_1, \dots, x_L} \prod_{i=1}^L \langle x_i | e^{-\frac{\beta H}{L}} | x_{i+1} \rangle, \quad (6)$$

where $x_{L+1} := x_1$ and L is to be chosen in the next step. The individual bit strings x_i in (x_1, \dots, x_L) are sometimes called “time slices.” When $\beta\|H\|/L$ is sufficiently small the Suzuki-Trotter approximation provides a way to split up the non-commuting terms in $e^{-\beta H/L}$ while only incurring a small error in the partition function. Define $A = -\beta s H_f$ and $B = -\beta(1-s)H_0$,

so that the partition function is $\mathcal{Z} = \text{tr} e^{A+B}$. Define the Suzuki-Trotter approximation to the partition function,

$$Z = \text{tr} \left[e^{\frac{A}{L}} e^{\frac{B}{L}} \right]^L. \quad (7)$$

According to Lemma 3 and the surrounding discussion of [5], taking $L = \Theta((\beta\|H\|)^{-3/2}\delta^{-1})$, with $\delta = 1/n$, suffices to achieve

$$(1 - \delta)\mathcal{Z} \leq Z \leq \mathcal{Z}(1 + \delta).$$

We will subsequently ignore the factor δ because it does not affect the convergence time of SQA, and creates only a negligible error in the distribution which SQA will sample from for the spike cost function (1). Expanding (7) as was done for (6),

$$Z = \sum_{x_1, \dots, x_L} \prod_{i=1}^L \langle x_i | e^{\frac{A}{L}} e^{\frac{B}{L}} | x_{i+1} \rangle \quad (8)$$

$$= \sum_{x_1, \dots, x_L} e^{-\frac{\beta s}{L} \sum_{i=1}^L f(x_i)} \prod_{k=1}^L \langle x_k | e^{\frac{B}{L}} | x_{k+1} \rangle \quad (9)$$

$$= \sum_{x_1, \dots, x_L} e^{-\frac{\beta s}{L} \sum_{i=1}^L f(x_i)} \prod_{j=1}^n \prod_{k=1}^L \langle x_{j,k} | e^{\omega \sigma_j^x} | x_{j,k+1} \rangle, \quad (10)$$

where $\omega = \beta(1-s)/L$. Using the identity $\exp(\omega \sigma^x) = \cosh(\omega)I + \sinh(\omega)\sigma^x$, the individual factors of the product in (10) become

$$\langle x_{j,k} | e^{\omega \sigma_j^x} | x_{j,k+1} \rangle = \cosh(\omega) [1_{x_{j,k}=x_{j,k+1}} + \tanh(\omega)1_{x_{j,k} \neq x_{j,k+1}}], \quad (11)$$

and after suppressing the uninteresting multiplicative factor of $\cosh(\omega)^{nL}$, the partition function is expressed as

$$Z = \sum_{x_1, \dots, x_L} e^{-\frac{\beta s}{L} \sum_{i=1}^L f(x_i)} \prod_{j=1}^n \prod_{k=1}^L [1_{x_{j,k}=x_{j,k+1}} + \tanh(\omega)1_{x_{j,k} \neq x_{j,k+1}}], \quad (12)$$

and so Z can be viewed as the normalizing constant of a probability distribution,

$$\pi(x_1, \dots, x_L) = \frac{1}{Z} e^{-\frac{\beta s}{L} \sum_{i=1}^L f(x_i)} \prod_{j,k=1}^{n,L} [1_{x_{j,k}=x_{j,k+1}} + \tanh(\omega)1_{x_{j,k} \neq x_{j,k+1}}] \quad (13)$$

$$= \frac{1}{Z} e^{-\frac{\beta s}{L} \sum_{i=1}^L f(x_i)} \prod_{j=1}^n \phi(\bar{x}_j) \quad (14)$$

where $\bar{x}_j = [x_{j,1}, \dots, x_{j,L}]$ is called “the worldline of the j -th qubit”, and $\phi(\bar{x}_j) = \tanh(\omega)^{|\{k: x_{j,k} \neq x_{j,k+1}\}|}$ counts the number of consecutive bits which disagree in that worldline.

Performing a similar calculation for $\Pi(x) = \langle x | \sigma | x \rangle$ (where $\sigma = e^{-\beta H} / \mathcal{Z}$) one finds it can be expressed as the marginal of π on the first time slice,

$$\Pi(x) = \sum_{x_2, \dots, x_L} \pi(x, x_2, \dots, x_L) \quad (15)$$

From this point SQA proceeds by discretizing the adiabatic path and using the Markov chain Monte Carlo method to sample from π at various values of the adiabatic parameter s_1, \dots, s_{\max} . The state space of this discrete-time Markov chain is $\Omega = \{0, 1\}^{n \times L}$, and the Markov chain

consist of local moves with Metropolis transition probabilities. If $x, x' \in \Omega$ differ by a single bit, the transition probability from x to x' is

$$P(x, x') = \frac{1}{nL} \min \left\{ 1, \frac{\pi(x')}{\pi(x)} \right\}, \quad (16)$$

and otherwise the transition probability is zero. We interpret this as randomly choosing one of the nL bits to consider flipping. We then flip the bit with probability $\min(1, \pi'(x)/\pi(x))$. Note that π is supported on all of Ω when $s < 1$, while at $s = 1$ the state space becomes disconnected under the local move transitions described above. This limitation is not important for our application, however, since sampling from Π when $s_{\max} = 1 - \frac{1}{n}$ suffices to find the true minimum of f . (In practical applications it is important to avoid this “critical (polynomial) slowing down” using non-local Markov Chain moves to equilibrate the system when $s \approx 1$.)

At each value of the adiabatic parameter, the run-time of SQA is determined by the mixing time of the Markov chain described above. This quantity can be defined in terms of the total variation distance from π to the distribution P_x^t obtained by running the chain for t steps starting from x ,

$$d_x(t) = \max_{A \subseteq \Omega} |P_x^t(A) - \pi(A)| = \frac{1}{2} \sum_{x' \in \Omega} |P_x^t(x') - \pi(x')|, \quad (17)$$

with the mixing time $\tau(\epsilon)$ being the worst-case time needed to be within variation distance ϵ of the stationary distribution,

$$\tau(\epsilon) = \max_{x \in \Omega} \min_t \{t : d_x(t) \leq \epsilon \ \forall t \geq t'\}. \quad (18)$$

A standard way to bound the mixing time is to relate it to the spectral gap λ of the transition matrix P [22],

$$\tau(\epsilon) \leq \lambda^{-1} \log \left(\frac{1}{\epsilon \pi_{\min}} \right). \quad (19)$$

These bounds are worst-case in the sense that they handle any starting vertex $x \in \Omega$. In the next section of our paper, we will develop slightly different methods to deal with the fact that our Markov chain may not mix efficiently from some starting vertices.

3 Incomplete sets of canonical paths

In this section we discuss how to show rapid mixing even in the presence of a small number of bad vertices. While we will freely make assumptions specific to our particular problem we will introduce some techniques that apply more generally to the analysis of Markov chains, and may be of use elsewhere. The notion of “good” and “bad” sets in Markov chains has been used before [19, 9], but always (to our knowledge) in a setting where a separate argument shows the bad set can always be quickly escaped. By contrast we model the bad set quite pessimistically and can assume that the walker gets absorbed (or equivalently, trapped for an exponential amount of time) upon hitting a bad vertex. Despite this we will show that the overall algorithm works with high probability.

3.1 Review of Markov chain comparison methods

This subsection reviews some standard facts from section 13.5 of the book by Levin, Peres and Wilmer [22]. Let P, \tilde{P} be reversible Markov chains (i.e. the transition probability from $x \rightarrow y$ is $P(x, y)$) with stationary distributions $\pi, \tilde{\pi}$. Define $Q(x, y) = \pi(x)P(x, y)$ and likewise for \tilde{Q} . Define the inner product $\langle f, g \rangle_\pi := \sum_x \pi(x) f(x) g(x)$ and the Dirichlet form

$$\mathcal{E}(f, g) = \langle \mathcal{E}f, g \rangle_\pi = \langle (I - P)f, g \rangle_\pi. \quad (20)$$

Lemma 13.11 of [22] states that

$$\mathcal{E}(f) := \mathcal{E}(f, f) = \frac{1}{2} \sum_{x, y \in \Omega} [f(x) - f(y)]^2 Q(x, y). \quad (21)$$

This can be used to define the gap (cf. Remark 13.13 of [22])

$$\lambda = \min_{\substack{f \in \mathbb{R}^\Omega \\ f \perp \pi \mathbf{1}, \|f\|_2=1}} \mathcal{E}(f) = \min_{\substack{f \in \mathbb{R}^\Omega \\ \text{Var}_\pi(f) \neq 0}} \frac{\mathcal{E}(f)}{\text{Var}_\pi(f)}. \quad (22)$$

To estimate the gap we will use various comparison methods. Given $x, y \in \Omega$ let \mathcal{P}_{xy} be the set of simple paths from x to y , and suppose that ν_{xy} is a measure over this set. Then we can define a congestion ratio:

$$\rho := \max_{e \in E} \rho(e) \quad (23)$$

$$\rho(e) := \frac{1}{Q(e)} \sum_{x, y \in \Omega} \tilde{Q}(x, y) \sum_{\Gamma: e \in \Gamma \in \mathcal{P}_{xy}} \nu_{xy}(\Gamma) |\Gamma| \quad (24)$$

This last sum ranges over all paths Γ in \mathcal{P}_{xy} that contain the edge e and thus is a measure of the total load on edge e . Then Corollary 13.26 of [22] states that

$$\tilde{\lambda} \leq \left[\max_{x \in \Omega} \frac{\pi(x)}{\tilde{\pi}(x)} \right] \rho \lambda. \quad (25)$$

If ν_{xy} has all its weight on a single flow γ_{xy} then we can slightly simplify (24) to

$$\rho(e) = \frac{1}{Q(e)} \sum_{\substack{x, y \in \Omega \\ e \in \gamma_{xy}}} \tilde{Q}(x, y) |\gamma_{xy}|. \quad (26)$$

Another variant is when we compare a chain \tilde{Q} with the complete graph for which the $x \rightarrow y$ transition probability is simply $\tilde{\pi}(y)$. The paths used for this purpose are labelled $\{\tilde{\gamma}_{xy}\}$. The corresponding congestion is

$$\tilde{\rho}(e) := \frac{1}{\tilde{Q}(e)} \sum_{\substack{x, y \in \Omega \\ e \in \tilde{\gamma}_{xy}}} \tilde{\pi}(x) \tilde{\pi}(y) |\tilde{\gamma}_{xy}|. \quad (27)$$

Since the complete graph has gap 1 and the same stationary distribution, we have

$$\tilde{\lambda} \geq \frac{1}{\tilde{\rho}}. \quad (28)$$

3.2 Most-paths comparison

In what follows we should think of \tilde{P} as an easy-to-analyze random walk (later we will take it to be n copies of Metropolis dynamics with local moves for a 1D Ising model of length $L = n^{O(1)}$) and P to be a walk such that $|\ln(P(x, y)/\tilde{P}(x, y))| \leq O(n^\alpha/L)$ but with $\tilde{\pi}$ potentially far from π . Specifically we will assume that

$$e^{-n^\alpha} \leq \frac{\pi(x)}{\tilde{\pi}(x)} \leq a = O(1), \quad (29)$$

(think of these bounds as generally achieved) and that $\|\tilde{\pi} - \pi\|_1 \leq O(n^\alpha/L)$. While this may seem close, note that the size of the perturbation is large relative to the spectral gap ($O(1/nL)$)

so that conventional stability theorems do not apply [26]. The fact that $\pi(x)/\tilde{\pi}(x)$ can be exponentially small also rules out direct application of (25).

We make one further assumption on the relation between π and $\tilde{\pi}$; namely that the ratio between them is usually not too small. Specifically let $\Omega_\theta \subset \Omega$ be the set of points x for which $\tilde{\pi}(x)/\pi(x) > \theta$ for some $\theta > 1$ that we will choose later. Then we will assume that $\tilde{\pi}(\Omega_\theta)$ is small. (By definition $\pi(\Omega_\theta) < \theta^{-1}\tilde{\pi}(\Omega_\theta)$ as well.)

Suppose further that \tilde{P} has a set of paths $\{\tilde{\gamma}_{xy}\}$ for which the congestion (defined in (27)) is $\tilde{\rho}$. These paths may lead to far larger congestion for P if they involve routing flow over edges e where $Q(e) \ll \tilde{Q}(e)$. However, our first claim is that most of these paths should avoid the bad set Ω_θ . First observe that for any e the definition of $\tilde{\rho}$ implies that

$$\sum_{x,y:\gamma_{xy} \ni e} \tilde{\pi}(x)\tilde{\pi}(y) \leq \tilde{\rho}\tilde{Q}(e). \quad (30)$$

Define E_θ to be the set of edges incident upon Ω_θ . Reversibility implies that

$$\frac{1}{2}\tilde{\pi}(\Omega_\theta) \leq \tilde{Q}(E_\theta) \leq \tilde{\pi}(\Omega_\theta). \quad (31)$$

(The inequalities are because an edge may be counted once or twice depending on whether both end points are in Ω_θ .) Additionally for $e = (v, w) \notin E_\theta$ we have

$$\tilde{Q}(e) = \tilde{\pi}(v)\tilde{P}(v, w) \leq \theta^{-1}\pi(v)\tilde{P}(v, w) \leq 2\theta^{-1}\pi(v)P(v, w) = \frac{2}{\theta}Q(e), \quad (32)$$

where we have used 2 as an overestimate for $e^{n^\alpha/L}$.

Let $C := \Omega^2$, $C_B := \{(x, y) \in C : \gamma_{xy} \cap E_\theta \neq \emptyset\}$ and $C_G := C - C_B$. Now we sum (30) over all $e \in E_\theta$ to obtain

$$\sum_{(x,y) \in C_B} \tilde{\pi}(x)\tilde{\pi}(y)|\tilde{\gamma}_{xy}| \leq \tilde{\rho}\tilde{Q}(E_\theta) = \tilde{\rho}\tilde{\pi}(\Omega_\theta). \quad (33)$$

We conclude that not many of the $\tilde{\gamma}_{xy}$ go through any edges in E_θ .

Now we define a second partition of Ω into good and bad vertices. Let $C_B(x) := \{y : (x, y) \in C_B\}$ and define $C_G(x)$ similarly. Then define

$$\Omega_B := \{x : \tilde{\pi}(C_B(x)) \geq 1/3a\}, \quad (34)$$

and $\Omega_G = \Omega - \Omega_B$. From (33) we have $\tilde{\pi}(\Omega_B) \leq 3a\tilde{\rho}\tilde{\pi}(\Omega_\theta)$.

Using first (29) and the assumption that $\pi(\Omega_\theta)$ is sufficiently small we have

$$\pi(\Omega_G) = 1 - \pi(\Omega_B) \geq 1 - 3a^2\tilde{\rho}\tilde{\pi}(\Omega_\theta) \geq \frac{11}{12}. \quad (35)$$

Thus Ω_G is a large-measure set that is mostly well connected. Indeed $\forall x \in \Omega_G$, $\pi(C_G(x)) \geq 2/3$.

We can now define canonical flows between all pairs $x, y \in \Omega_G$. Observe that

$$\pi(C_G(x) \cap C_G(y) \cap \Omega_G) \geq 1/4. \quad (36)$$

This means that even if x, y are not directly connected to each other by a path that avoids E_θ , they are still indirectly connected via pairs of paths that route through a large-measure ($\geq 1/4$) set. We will see that this allows us to construct canonical flows that are not too much worse than our original canonical paths.

Observe next that the conditional distribution $\pi_{xy} := \pi|_{C_G(x) \cap C_G(y) \cap \Omega_G}$ satisfies $\pi_{xy} \leq 4\pi$. We will construct our flow ν_{xy} by choosing a random $r \sim \pi_{xy}$ and concatenating the paths $\tilde{\gamma}_{xr}$

and $\tilde{\gamma}_{ry}$. The load on edge e from these flows is 0 if $e \notin E_\theta$, or if $e \in E_\theta$ can be bounded as

$$\sum_{x,y \in \Omega_G} \pi(x)\pi(y) \sum_r \pi_{xy}(r)(1_{e \in \tilde{\gamma}_{xr}} + 1_{e \in \tilde{\gamma}_{ry}})(|\tilde{\gamma}_{xr}| + |\tilde{\gamma}_{ry}|) \quad (37)$$

$$\leq 4 \sum_{x,y \in \Omega_G} \pi(x)\pi(y) \sum_r \pi_{xy}(r) 1_{e \in \tilde{\gamma}_{xr}} |\tilde{\gamma}_{xr}| \quad \text{by symmetry} \quad (38)$$

$$\leq 16 \sum_{x,y \in \Omega_G} \pi(x)\pi(y) \sum_{r \in C_G(x) \cap \Omega_G} \pi(r) 1_{e \in \tilde{\gamma}_{xr}} |\tilde{\gamma}_{xr}| \quad \text{since } \pi_{xy} \leq 4\pi \quad (39)$$

$$= 16 \sum_{x \in \Omega_G} \pi(x) \sum_{r \in C_G(x) \cap \Omega_G} \pi(r) 1_{e \in \tilde{\gamma}_{xr}} |\tilde{\gamma}_{xr}| \quad \text{since } \pi \text{ is normalized} \quad (40)$$

$$\leq 16 \sum_{(x,r) \in C_G} \pi(x)\pi(r) 1_{e \in \tilde{\gamma}_{xr}} |\tilde{\gamma}_{xr}| \quad \text{passing to a superset} \quad (41)$$

$$\leq 16a^2 \sum_{(x,r) \in C_G} \tilde{\pi}(x)\tilde{\pi}(r) 1_{e \in \tilde{\gamma}_{xr}} |\tilde{\gamma}_{xr}| \quad \text{from (29)} \quad (42)$$

$$\leq 16a^2 \tilde{\rho} \tilde{Q}(e) \quad \text{from (30)} \quad (43)$$

$$\leq 32\theta^{-1} a^2 \tilde{\rho} Q(e) \quad \text{from (32)} \quad (44)$$

We conclude that our set of canonical flows achieves congestion

$$\rho \leq 32\theta^{-1} a^2 \tilde{\rho}, \quad (45)$$

albeit on a set of measure less than one, namely Ω_G . In the next section we will discuss the implications of this for mixing.

3.3 Leaky random walks

For this section we will not need all of the assumptions used in the previous section. Assume merely that Ω is partitioned into good and bad sets Ω_G, Ω_B and define the “good-only” walk P_G to be

$$P_G(x, y) = P(x, y) 1_{x \in \Omega_G} 1_{y \in \Omega_G}.$$

Since P_G is substochastic (nonnegative entries and rows sum to ≤ 1) we can think of it as corresponding to a random process in which a walker at x moves to y with probability $P_G(x, y)$ and is deleted with probability $1 - \sum_{y \in \Omega} P_G(x, y) = \sum_{y \in \Omega_B} P(x, y)$. If we define Π_G to be the projector onto \mathbb{R}^{Ω_G} then we can also define P_G as

$$P_G = \Pi_G P \Pi_G. \quad (46)$$

Assuming that P is ergodic, $\lim_{t \rightarrow \infty} P_G^t(x, \cdot) = 0$ for any starting point $x \in \Omega$. However there may be an intermediate range of times, $t_{\text{mix}} \leq t \leq t_{\text{fail}}$, for which $\|P_G^t(x, \cdot) - \pi\|_1 < \delta$ for some desired δ , if the initial point $x \in \Omega_G$ is a *warm start* taken from $x \sim \mu$ with $\|\pi - \mu\|_1$ sufficiently small. Formalizing the preceding statement will be the goal of this section.

A standard fact about Markov chain mixing (Thm 12.3 of [22]) is that for all $x \in \Omega$,

$$\|P^t(x, \cdot) - \pi\|_1 \leq \pi_{\min}^{-1} e^{-\lambda t}. \quad (47)$$

To bound the gap of P_G observe that Thm 13.23 of [22] can be applied even to substochastic matrices. Here it implies that

$$\Pi_G \mathcal{E} \Pi_G \succeq_\pi \rho^{-1} \Pi_G. \quad (48)$$

with ρ from (45), and with \succeq_π denoting the semidefinite ordering with respect to the $\langle \cdot, \cdot \rangle_\pi$ inner product (i.e. $A \succeq_\pi 0$ means $\langle f, Af \rangle_\pi \geq 0$). Eq. (48) does not directly give bounds on the gap

of P . Indeed we could in principle have $P(x, x) = 1$ for all $x \in \Omega_B$ in which case P would have many eigenvalues equal to 1. Thus without additional information we cannot say anything more about P .

Instead we will analyze a slightly different Markov chain. We will define a “filled-in” chain P_F by adding nonnegative numbers to the entries of P_G in such a way that P_F is stochastic and has π as a stationary distribution. This does not uniquely specify P_F and indeed P also satisfies these conditions. We will instead choose P_F in a way that guarantees fast mixing. Specifically with probability $1 - \sum_{y \in \Omega} P_G(x, y)$ we will forget our current location x and jump according to some “fill-in” measure φ . For this to result in π being the stationary distribution we must have $\varphi = \pi(I - P_G)$. Defining the column vector $\mathbf{1} = (1, 1, \dots, 1)^T$, we have

$$P_F = P_G + \mathbf{1}\varphi \quad (49)$$

$$\pi P_F = \pi(P_G + \mathbf{1}\varphi) = \pi(P_G + \mathbf{1}\pi(I - P_G)) = \pi. \quad (50)$$

Now that P_F is a proper Markov chain we will bound its Dirichlet form. Recall that in (22) we can assume that $f \perp_{\pi} \mathbf{1}$ meaning that $\sum_x \pi(x)f(x) = 0$. Note as well that

$$\mathcal{E}_F = I - P_F = I - P_G - \mathbf{1}\pi(I - P_G) = (I - \mathbf{1}\pi)(I - P_G). \quad (51)$$

Define D_{π} to have π along the diagonal and zeros elsewhere. Then

$$\langle \mathcal{E}_F f, f \rangle = \langle (I - \mathbf{1}\pi)(I - P_G)f, f \rangle \quad (52a)$$

$$= f^T (I - P_G^T)(I - \pi^T \mathbf{1}^T) D_{\pi} f \quad (52b)$$

$$= f^T (I - P_G^T) D_{\pi} f \quad \text{since } f \perp_{\pi} \mathbf{1} \quad (52c)$$

$$= f^T (I - \Pi) D_{\pi} f + f^T (\Pi - P_G^T) D_{\pi} f \quad (52d)$$

$$\geq f^T (I - \Pi) D_{\pi} f + \rho^{-1} f^T \Pi D_{\pi} f \quad \text{using (48)} \quad (52e)$$

$$\geq \rho^{-1} f^T D_{\pi} f \quad \text{since } \rho \geq 1 \quad (52f)$$

$$= \rho^{-1} \text{Var}_{\pi}[f] \quad \text{since } f \perp_{\pi} \mathbf{1} \quad (52g)$$

We conclude that P_F mixes rapidly, while differing from P_G only in the events where leakage occurs. More precisely we can define a coupling between two processes: the first a walk evolving according to P_F and the second a walk in which x moves to y with probability $P_G(x, y)$ and to a special state $*$ with probability $1 - \sum_y P_G(x, y)$. The state $*$ is absorbing, meaning that when the second walker is at state $*$ it stays there. Conditioned on the second walker not being in state $*$ the two walkers will be at the same location. Additionally the probability of the second walker ending in $*$ equals the probability that the first walker ever passed through Ω_B . If we start with a probability distribution μ and take t steps then the probability of ending in $*$ is $\mu(P_F^t - P_G^t)\mathbf{1}$. This quantity also equals the variational distance between the two walkers' probability distributions.

Suppose that we start with an M -warm distribution μ , by which we mean that $\mu(x) \leq M\pi(x)$ for all x . Note that this property is strictly preserved by P_F and P_G since

$$\mu P_G^t \leq \mu P_F^t \leq M\pi P_F^t = M\pi. \quad (53)$$

In this case the probability that a path x_1, \dots, x_t (with $x_1 \sim \mu$ and $x_i \sim P_F(x_{i-1})$ for $i > 1$) passes through Ω_B is

$$\leq \sum_{i=1}^t \Pr[x_i \in \Omega_B] \leq Mt\pi(\Omega_B), \quad (54)$$

where we have used first the union bound and then (53). By the above arguments we have

$$\|\mu(P_F^t - P_G^t)\|_1 \leq Mt\pi(\Omega_B). \quad (55)$$

Using the spectral gap of P_F (from (52)) and the mixing bound in (47) we conclude that

$$\|\pi - \mu P_G^t\|_1 \leq Mt\pi(\Omega_B) + \pi_{\min}^{-1} e^{-t/\rho} \quad (56)$$

We see that the leakage probability increases linearly with t while the usual distance to stationarity decreases exponentially with t . In many cases this leaves a wide range of t in which the RHS of (56) can be small.

4 Efficient convergence of SQA for the spike cost function

In this section we apply the method from Section 3.2 with the easy-to-analyze chain $(\tilde{\pi}, \tilde{P})$ taken to be the SQA Markov chain for the system without the spike, and (π, P) equal to the chain with the spike. The subset Ω_B will be shown to satisfy $\tilde{\pi}(\Omega_B) \leq \mathcal{O}(n^{-c})$ for a constant c that we will choose so that we can prove a walker beginning in Ω_G is likely to mix before it hits a point in Ω_B .

Congestion of the spikeless chain. Recall that Ω_B is defined in terms of a set of canonical paths $\{\gamma\}$ on Ω with congestion $\tilde{\rho}$ for the spikeless chain, together with a subset Ω_θ of points which are excluded from paths in $\{\gamma\}$ to obtain a new set of paths with congestion $\rho \leq \mathcal{O}(\theta^{-1}\tilde{\rho})$ for the chain with the spike, within the subset Ω_G . The chain for the spikeless distribution $\tilde{\pi}$ corresponds to a discrete-time Glauber dynamics on a collection of n non-interacting 1D ferromagnetic Ising models with (with 1-local fields), each of length L , with a spin-spin coupling such that each broken bond in x lowers $\tilde{\pi}(x)$ by a factor of $\Theta(\tanh(\omega))$.

For the canonical paths $\{\gamma\}$ we apply the standard hypercube bit-fixing paths to each of the worldlines, one at a time: to go from $x = (x_1, \dots, x_L)$ to $y = (y_1, \dots, y_L)$, we first change the bit $x_{1,j}$ to $y_{1,j}$ for consecutive $j = 1, \dots, L$, and then repeat these steps to turn each x_i into y_i . Evaluating the congestion for these paths in the usual way with an encoding function yields $\tilde{\rho} \leq \mathcal{O}(\coth(\omega)nL)$, which is maximized when $\omega = \omega_{\min}$,

$$\tilde{\rho} \leq \mathcal{O}(\coth(\omega_{\min})nL) \leq \mathcal{O}(\omega_{\min}^{-1}nL) \leq \mathcal{O}(\beta^{-1}nL^2) \quad (57)$$

Ω_θ and the spike time distribution. The states in Ω_θ which will be excluded from the set of paths $\{\gamma\}$ are those which have $|x_i| = n/4$ for too many i . Define $1_S : \{0, 1\}^n \rightarrow \{0, 1\}$ to be the indicator function for the spike i.e. $1_S(z) = 1$ if $|z| = n/4$, and $1_S(z)$ is zero otherwise. The spike time for $x \in \Omega$ is defined to be

$$\text{ST}(x) = \sum_{i=1}^L 1_S(x_i). \quad (58)$$

Let $\epsilon = \frac{1}{2} - \alpha$, and define

$$\Omega_\theta = \left\{ x \in \Omega : \text{ST}(x) \geq \frac{L}{n^{\frac{1}{2}(1-\epsilon)}} \right\}. \quad (59)$$

Set $\beta = n^{\epsilon/2}$ so that every $x \in \Omega_\theta$ satisfies

$$\frac{\tilde{\pi}(x)}{\pi(x)} \geq \left(\frac{Z}{\tilde{Z}} \right) \exp \left[-\frac{\beta n^\alpha}{L} \cdot \left(\min_{x \in \Omega_\theta} \text{ST}(x) \right) \right] = \mathcal{O}(1), \quad (60)$$

which shows $\theta = \mathcal{O}(1)$ in the congestion bound (45). The remainder of the section will be devoted to computing the m -th moment of the random variable $\text{ST} \sim \tilde{\pi}$, with $m = c/\epsilon$, in order to show,

$$\Pr \left[\text{ST} \geq \frac{L}{n^{\frac{1}{2}(1-\epsilon)}} \right]_{\tilde{\pi}} \leq \mathcal{O}(n^{-c}), \quad (61)$$

which is equivalent to the statement $\tilde{\pi}(\Omega_\theta) \leq \mathcal{O}(n^{-c})$.

To calculate the moments $\langle \text{ST}^m \rangle_{\tilde{\pi}}$ we will relate them to expectation values of the spikeless quantum system, and use the fact that the latter is exactly solvable because the qubits are non-interacting. Recall that the quantum expectation value of an operator A can be expressed as a derivative of the partition function,

$$\langle A \rangle = \frac{1}{\mathcal{Z}} \text{tr} [Ae^{-\beta H}] = \frac{1}{\mathcal{Z}} \frac{\partial}{\partial \lambda} \text{tr} [e^{-\beta H + \lambda A}] |_{\lambda=0} = \frac{1}{\mathcal{Z}(0)} \frac{\partial \mathcal{Z}(\lambda)}{\partial \lambda} |_{\lambda=0}. \quad (62)$$

Passing from $\mathcal{Z}(\lambda)$ to the corresponding Suzuki-Trotter approximation $Z(\lambda)$ in the above expression changes the value by at most a multiplicative factor of $(1 \pm \mathcal{O}(L^{-1}))$ (a proof of this fact will appear in a future work [11]). Let $\{|k\rangle : k = 0, \dots, n\}$ be a basis of states for the symmetric subspace which are labeled by Hamming weight, and let $S = |n/4\rangle\langle n/4|$. Since the observable S is diagonal in the computational basis we can include the term λS into the diagonal part of the Hamiltonian for the quantum-to-classical mapping and compute,

$$\langle S \rangle_{\tilde{\sigma}} = \frac{1}{Z} \frac{\partial}{\partial \lambda} \left[\sum_{x_1, \dots, x_L} \exp \left(\sum_{i=1}^L -\frac{\beta \tilde{f}(x_i)}{L} + \frac{\lambda(x_i |S| x_i)}{L} \right) \prod_{j=1}^n \phi(\bar{x}_j) \right] |_{\lambda=0} \quad (63)$$

$$= \frac{1}{Z} \sum_{x_1, \dots, x_L} \left[\frac{1}{L} \sum_{i=1}^L 1_S(x_i) \right] e^{-\frac{\beta}{L} \sum_{i=1}^L \tilde{f}(x_i)} \prod_{j=1}^n \phi(\bar{x}_j) \quad (64)$$

$$= \sum_{x \in \Omega} L^{-1} \text{ST}(x) \tilde{\pi}(x) = L^{-1} \langle \text{ST} \rangle_{\tilde{\pi}} \quad (65)$$

Since the low-temperature thermal state and the ground state have a large overlap, the expectation value (65) can be computed within the ground state. To compute the error caused by this replacement we will examine the density of states of the spikeless system. First add a constant shift to the Hamiltonian so that $\tilde{H}|\tilde{\psi}_0\rangle = 0$. Next, let $|\tilde{\psi}_1\rangle, \dots, |\tilde{\psi}_n\rangle$ denote the excited eigenstates of \tilde{H} . Define $\Delta := 2\sqrt{(1-s)^2 + s^2}$ and observe that $|\tilde{\psi}_k\rangle$ is an eigenstate of \tilde{H} with eigenvalue $k\Delta$, and that the degeneracy of the k -th energy level is $\binom{n}{k}$.

$$\|\tilde{\sigma} - |\tilde{\psi}_0\rangle\langle\tilde{\psi}_0|\|_1 \leq \sum_{k=1}^n e^{-\beta\Delta k} \binom{n}{k} = \mathcal{O}(ne^{-n^\epsilon}) \quad (66)$$

and since $\epsilon > 0$ is a constant this error will be subleading.

Since the spikeless system is non-interacting the ground state can be written explicitly, and the ground state probability distribution on the Hamming weights is a binomial distribution [25]. It follows that $\langle S \rangle_{\tilde{\sigma}}$ is less than the central binomial coefficient,

$$\langle \text{ST} \rangle_{\tilde{\pi}} < L \left(\frac{1}{\sqrt{n}} + \mathcal{O}(n^{-1}) \right). \quad (67)$$

To obtain (61) we will use the moment inequality,

$$\Pr[\text{ST} \geq b]_{\tilde{\pi}} \leq \frac{\langle \text{ST}^m \rangle_{\tilde{\pi}}}{b^m}, \quad (68)$$

with $b = Ln^{-\frac{1}{2}(1-\epsilon)}$. By definition we have,

$$\langle \text{ST}^m \rangle_{\tilde{\pi}} = \sum_{z_1, \dots, z_L} \tilde{\pi}(z_1, \dots, z_L) \left(\sum_{t=1}^L 1_S(z_t) \right)^m \quad (69)$$

$$= \sum_{z_1, \dots, z_L} \tilde{\pi}(z_1, \dots, z_L) \sum_{t_1, \dots, t_m}^L 1_S(z_{t_1}) \dots 1_S(z_{t_m}) \quad (70)$$

$$= \sum_{t_1, \dots, t_m}^L \langle 1_S(z_{t_1}) \dots 1_S(z_{t_m}) \rangle_{\tilde{\pi}}. \quad (71)$$

To compute these m -point correlation functions we return to the quantum description (this generates a multiplicative error of size $1 \pm \mathcal{O}(L^{-1})$, which will make a subleading contribution to the m -th moment and thus will be ignored),

$$\langle 1_S(z_{t_1}) \dots 1_S(z_{t_m}) \rangle_{\tilde{\pi}} = \langle e^{-\tau_1 H} S e^{-(\tau_2 - \tau_1) H} S \dots e^{-(\tau_m - \tau_{m-1}) H} S e^{-(\beta - \tau_m) H} \rangle_{\tilde{\sigma}} \quad (72)$$

where $\tau_i = \beta t_i / L$ and $S = |n/4\rangle\langle n/4|$. Once again we replace the low-temperature thermal state with the ground state and incur a subleading error as in (66). Therefore,

$$\langle \text{ST}^m \rangle_{\tilde{\pi}} = \sum_{t_1, \dots, t_m}^L \langle \tilde{\psi}_0 | e^{-\tau_1 H} S e^{-(\tau_2 - \tau_1) H} S \dots e^{-(\tau_m - \tau_{m-1}) H} S e^{-(\beta - \tau_m) H} | \tilde{\psi}_0 \rangle \quad (73)$$

Since the ground state, the Hamiltonian, and the operator S are all bit-symmetric, the expectation can be evaluated in the symmetric subspace. Expanding each of the terms in (73) in the basis of symmetric energy eigenstates $\{|\tilde{\psi}_k\rangle\}$,

$$\langle \text{ST}^m \rangle_{\tilde{\pi}} = \sum_{\substack{k_1, \dots, k_m \\ t_1, \dots, t_m}} e^{-(\tau_2 - \tau_1) \Delta k_1} \dots e^{-(\tau_m - \tau_{m-1}) \Delta k_1} \langle \tilde{\psi}_0 | S | \tilde{\psi}_{k_1} \rangle \langle \tilde{\psi}_{k_2} | S | \tilde{\psi}_{k_3} \rangle \dots \langle \tilde{\psi}_{k_m} | S | \tilde{\psi}_0 \rangle \quad (74)$$

States with higher energy will contribute less to the sum over all times t_1, \dots, t_m in (73), because the exponentials in (74) decay more quickly. For $k_i > 0$, the sum over t_i can be truncated whenever $\tau_i - \tau_{i-1} \gg 1/k_i \Delta$.

Since the ground state wave function is a binomial distribution the mean spike time will only be large when the peak of the distribution is near the spike, which only happens when the adiabatic parameter s is in the interval $s^* - \mathcal{O}(n^{-1/2}) < s < s^* + \mathcal{O}(n^{1/2})$ with $s^* = (\sqrt{3}-1)/2$. In this interval the spikeless eigenstates satisfy $\langle \tilde{\psi}_i | n/4 \rangle \leq |\langle \tilde{\psi}_0 | n/4 \rangle| \leq \mathcal{O}(n^{-1/2})$ for all i , because the ground state wave function is centered on the spike and the excited state wave functions have a greater spread, which can be seen from the explicit form of the spikeless eigenfunctions given in [21]. We can define $g_i = t_i - t_{i-1}$ and relabel the sum of $t_1, \dots, t_m = 0, \dots, L$ by a sum over the g_i . For the purpose of obtaining an upper bound on the m -th moment we relax the constraint $\sum_i g_i = L$, and instead sum over the full range $g_i = 1, \dots, L$ for each i . Using these facts we can upper bound (74) by

$$\langle \text{ST}^m \rangle_{\tilde{\pi}} \leq n^{-m/2} \sum_{\substack{k_1, \dots, k_m \\ g_1, \dots, g_m}} e^{-g_1 \Delta k_1} \dots e^{-g_m \Delta k_m} \quad (75)$$

We will now organize the terms of (75) according to the number ℓ of excited energies $\tilde{E}_{k_i} > 0$ they contain. There are $\binom{m}{\ell}$ terms of (75) that contain ℓ eigenstates above the ground state, and for each ℓ we must sum over the $g_{a_1}, \dots, g_{a_\ell}$ for which the corresponding $k_{a_1}, \dots, k_{a_\ell}$ are

non-zero. Now (75) becomes

$$\langle \text{ST}^m \rangle_{\tilde{\pi}} \leq n^{-m/2} \sum_{\ell=1}^m \binom{m}{\ell} \frac{L^{m-\ell}}{(m-\ell)!} \sum_{\substack{g_{a_1}, \dots, g_{a_\ell} \\ k_{a_1}, \dots, k_{a_\ell}}} \prod_{i=1}^{\ell} e^{-g_{a_i} \Delta k_{a_i}} \quad (76)$$

where the factor of $L^{m-\ell}/(m-\ell)!$ results from performing the sum over the $m-\ell$ of the g_i which have $k_i = 0$. Now we sum over $g_{a_1}, \dots, g_{a_\ell}$ using the fact that $\sum_{g=1}^L e^{-gk} \leq k^{-1}$,

$$\langle \text{ST}^m \rangle_{\tilde{\pi}} \leq L^m n^{-m/2} \sum_{\ell=0}^m \binom{m}{\ell} (\beta \Delta)^{-\ell} \sum_{k_{a_1}, \dots, k_{a_\ell}} \prod_{i=1}^{\ell} \frac{1}{k_i}. \quad (77)$$

Using $\binom{m}{\ell} \leq m^\ell$ and $\sum_{k=1}^n \leq \log(n) + 1$, at last this becomes

$$\langle \text{ST}^m \rangle_{\tilde{\pi}} \leq L^m n^{-m/2} \sum_{\ell=0}^m \left(\frac{\log(n) + 1}{m\beta\Delta} \right)^\ell \quad (78)$$

Since m and Δ are constant and $\beta = n^{\epsilon/2}$ with fixed $\epsilon > 0$ the terms with inverse powers of β are subleading and so $\langle \text{ST}^m \rangle_{\tilde{\pi}} \leq \mathcal{O}(L^m n^{-m/2})$. Finally, applying (68) with $m = c/\epsilon$ yields the desired result (61).

Adiabatic schedule. Here we show that a discretization of the adiabatic with an inverse polynomial step size is sufficient to fulfill the statement we need for the warm starts in Section 3.3. We will take the largest value of the adiabatic parameter to be $s_{\max} = 1 - n^{-1}$ so that $\|\psi_0(1) - |\psi_0(s_{\max})\rangle\|_1 \leq \text{poly}(n^{-1})$, and the global minimum of the cost function can be obtained by sampling from Π at $s = s_{\max}$ with essentially the same probability as it would be obtained by sampling from the ground state probability distribution.

We will sample from π at several values of the adiabatic parameter s_1, \dots, s_{\max} . Define $s_i = s_0 - i\Delta s$, $\omega_i = \beta(1 - s_i)/L$, $\Delta\omega = \beta\Delta s/L$, and let π^i be the stationary distribution (14) when the adiabatic parameter is s_i . At each stage we simulate the Markov chain (16) for sufficiently many steps to achieve a variational distance to the stationary distribution of $\exp(-n^{\Omega(1)})$. These errors then add up to a negligible amount.

To choose a step size Δs satisfying the warm start condition $\pi_{i+1} \leq 2\pi_i$ the following calculation makes use of the form (14) of the stationary distribution as well as the form (5) of the partition function,

$$\frac{\pi^{i+1}(x)}{\pi^i(x)} = \left(\frac{Z^i}{Z^{i+1}} \right) e^{-\frac{\beta\Delta s}{L} \sum_{i=1}^L f(x_i)} \prod_{j,k=1}^{n,L} \frac{[1_{x_{j,k}=x_{j,k+1}} + \tanh(\omega_i - \Delta\omega) 1_{x_{j,k} \neq x_{j,k+1}}]}{[1_{x_{j,k}=x_{j,k+1}} + \tanh(\omega_i) 1_{x_{j,k} \neq x_{j,k+1}}]} \quad (79)$$

$$\geq e^{-2\beta\Delta s(\|H_0\| + \|H_f\|)} \left[\frac{\tanh(\omega_i - \Delta\omega)}{\tanh(\omega_i)} \right]^{nL} \quad (80)$$

$$\geq (1 - 2\beta\Delta s(\|H_0\| + \|H_f\|)) \left[1 + \Delta\omega \max_{w \in (\omega_{i+1}, \omega_{i-1})} (\coth(w) - \tanh(w)) \right]^{nL} \quad (81)$$

$$\geq (1 - 2\beta\Delta s(\|H_0\| + \|H_f\|)) [1 + \coth(\omega_{i+1}) \cdot \Delta\omega]^{nL}, \quad (82)$$

and so taking $\Delta s = [\coth(\omega_{\min})(nL)^2]^{-1} = \Theta(L^3 n^2 [\beta(1 - s_{\max})]^{-1})$ will satisfy $\pi^i(x) \leq 2\pi^{i+1}(x)$, and since $s_{\max} = 1 - n^{-1}$ and $\beta = n^{\epsilon/2}$ this becomes $\Delta s = \Theta\left(n^{-\left(\frac{19}{2} - \frac{13\epsilon}{4}\right)}\right)$.

An alternative approach to lower bounding the necessary step size Δs would be to replace the thermal state with the ground state and apply Lemma 5.1 from [8], which would yield $\Delta s = 1/\text{poly}(n)$ since the spectral gap of the spike Hamiltonian is constant for $\alpha < 1/2$.

Quasi-stationary mixing. In this section we will bound the overall run time of SQA applied to the spike cost function. First we need to show that taking c in (61) to be a sufficiently large constant will allow the leaky walk for the spike system sample from Ω_G according to the quasi-stationary distribution π for an expected time of n^q , for any desired constant q , before it is eventually likely to escape into Ω_B .

Inserting (57) and (60) into (45) yields the bound $\rho \leq \mathcal{O}(\beta^{-1}nL^2) = \mathcal{O}(n^{6+\epsilon})$. Since $\log \pi_{\min}^{-1} \leq \mathcal{O}(\beta n + nL \log(1 - \frac{1}{n}))$ this implies a mixing time of $t_{\text{mix}} \leq \mathcal{O}(n^{\frac{17}{2} + \frac{7}{4}\epsilon})$ within Ω_G for the SQA spike chain at each value s_i of the adiabatic path.

Meanwhile, from (35) together with (61) we have $\pi(\Omega_B) \leq \Theta(\tilde{\rho}\tilde{\pi}(\Omega_\theta)) = \Theta(n^{6+\epsilon-c})$. At each step s_i of the adiabatic path, after time $t \geq t_{\text{mix}}$ the leaky random walk mixes to within a distance $\mathcal{O}(t\pi(\Omega_B))$ so by (56) it suffices to take $c = 6 + \epsilon + q + \log(1/\delta)$ in order for the leaky walk to be with distance δ to the stationary distribution π for times $t_{\text{mix}} \leq t \leq \Theta(n^q)$.

For an explicit upper bound on the run time of the entire SQA process, we make take $q = 1$ since only one sample needs to be drawn at each s_i for a warm start, and the total run time is $\mathcal{O}(t_{\text{mix}}\Delta s^{-1}) = \mathcal{O}(n^{18-6\epsilon})$.

5 Discussion

Our proof does not bound the convergence time for SQA $\alpha > 1/2$, although QA does work here (with a runtime $n^{O(\alpha)}$) [21]. We conjecture that SQA will be efficient for $\alpha > 1/2$ as well, though this will require extensions of the present techniques. The approach used in this work is also suggestive of a more general connection between the quantum spectral gap of Hamming symmetric barrier problems (including barriers of various shapes and widths) and the corresponding performance of SQA, which we sketch here.

Assume we are near the critical value s^* of the adiabatic parameter at which the system tunnels through the barrier i.e. for $s < s^*$ the ground state probability mass is concentrated on one side of the barrier, while for $s > s^*$ the opposite occurs, and for $s = s^*$ the probability mass on both sides of the barrier is $\mathcal{O}(1)$. While a general understanding of what barriers admit such a tunneling description has not yet been found, there are many examples for which this is known to occur [14, 21, 24]. Assume that the QA spectral gap at s^* is $\Delta = 1/\text{poly}(n)$, so that QA will be able to pass through the barrier efficiently.

We expect the first excited state to have a node at some location $|k\rangle$ inside the barrier region, and some properties of the ground state wave function near $|k\rangle$ can be inferred from the spectral gap. Since the spectral gap is at least $1/\text{poly}(n)$, we know that the ground state wave function cannot be too small inside of the barrier, or else there would be a balanced-weight cut in the ground state that could be used to construct an orthogonal state with low energy. This benefits the comparison approach because the ground state amplitudes inside the barrier need to be at least $1/\text{poly}(n)$ in order for a $1/\text{poly}(n)$ fraction of the canonical paths to transfer from the spikeless system to the system with a barrier.

On the other hand, the amplitudes inside the barrier cannot become too large because the barrier is a classically forbidden region, and because an upper bound on Δ will also upper bound the amplitudes near $|k\rangle$ (by the same argument using a balanced-weight cut together with the assumption that the first excited state has a node at $|k\rangle$). The fact that the total probability mass inside the barrier is not too large could be useful for deriving the necessary upper bounds on $\pi(\Omega_\theta)$ that are used in the comparison approach. The primary ingredient that would be needed to make this sketch rigorous is a better understanding of the wave functions and spectral gap of the quantum system with a barrier, and depending on the outcome of this understanding the comparison approach for analyzing SQA may require modifications as well.

While relatively few rigorous facts are known about the performance of SQA, it remains in practice a successful and widely used algorithm. This strikes us as an area where theorists should work to catch up with current practice.

Acknowledgments

We thank Dave Bacon, Wim van Dam and Alistair Sinclair for helpful conversations. Elizabeth Crosson gratefully acknowledges funding provided by the Institute for Quantum Information and Matter, an NSF Physics Frontiers Center (NSF Grant PHY-1125565) with support of the Gordon and Betty Moore Foundation (GBMF-12500028), and is also grateful for support received while completing a portion of this work at the MIT Center for Theoretical Physics with funding from NSF grant number CCF-1111382. Aram Harrow was funded by NSF grants CCF-1111382 and CCF-1452616 and ARO contract W911NF-12-1-0486.

References

- [1] D. Battaglia, G. E. Santoro, and E. Tosatti. Optimization by quantum annealing: Lessons from hard 3-SAT cases. *Phys. Rev. E*, 2005, [arXiv:cond-mat/0502468](#).
- [2] S. Boixo, T. F. Rønnow, S. V. Isakov, Z. Wang, D. Wecker, D. A. Lidar, J. M. Martinis, and M. Troyer. Quantum annealing with more than one hundred qubits. *Nature Phys.*, page 218, 2014, [arXiv:1304.4595](#).
- [3] L. T. Brady and W. van Dam. Quantum Monte Carlo simulations of tunneling in quantum adiabatic optimization, 2015, [arXiv:1509.02562](#).
- [4] L. T. Brady and W. van Dam. Spectral gap analysis for efficient tunneling in quantum adiabatic optimization, 2016, [arXiv:1601.01720](#).
- [5] S. Bravyi. Monte Carlo simulation of stoquastic Hamiltonians, 2014, [arXiv:1402.2295](#).
- [6] S. Bravyi, A. J. Bessen, and B. M. Terhal. Merlin-Arthur games and stoquastic complexity, 2006, [arXiv:quant-ph/0611021](#).
- [7] S. Bravyi, D. P. DiVincenzo, R. I. Oliveira, and B. M. Terhal. The complexity of stoquastic local Hamiltonian problems. *Quant. Inf. Comp.*, 8(5):0361–0385, 2006, [arXiv:quant-ph/0606140](#).
- [8] S. Bravyi and B. M. Terhal. Complexity of stoquastic frustration-free Hamiltonians. *SIAM J. Comput.*, 39(4):1462–1485, 2009, [arXiv:0806.1746](#).
- [9] P. Caputo, E. Lubetzky, F. Martinelli, A. Sly, and F. L. Toninelli. Dynamics of $(2 + 1)$ -dimensional SOS surfaces above a wall: Slow mixing induced by entropic repulsion. *Ann. Probab.*, 42(4):1516–1589, 07 2014, [arXiv:1205.6884](#).
- [10] E. Crosson and M. Deng. Tunneling through high energy barriers in simulated quantum annealing, 2014, [arXiv:1401.7320](#).
- [11] E. Crosson and A. Harrow. Rapidly mixing Monte Carlo for 1D stoquastic Hamiltonians, 2016. In preparation.
- [12] V. S. Denchev, S. Boixo, S. V. Isakov, N. Ding, R. Babbush, V. Smelyanskiy, J. Martinis, and H. Neven. What is the computational value of finite range tunneling?, 2015, [arXiv:1512.02206](#).
- [13] M. Dyer, L. A. Goldberg, M. Jerrum, and R. Martin. Markov chain comparison. *Probability Surveys*, 3:89–111, 2006, [arXiv:math/0410331](#).
- [14] E. Farhi, J. Goldstone, and S. Gutmann. Quantum adiabatic evolution algorithms versus simulated annealing, 2002, [arXiv:quant-ph/0201031](#).
- [15] E. Farhi, J. Goldstone, S. Gutmann, and M. Sipser. Quantum computation by adiabatic evolution, 2000, [arXiv:quant-ph/0001106](#).
- [16] M. B. Hastings. Obstructions to classically simulating the quantum adiabatic algorithm. *Quantum Information & Computation*, 13(11-12):1038–1076, 2013, [arXiv:1302.5733](#).

- [17] E. Inack and S. Pilati. Simulated quantum annealing of double-well and multiwell potentials. *Physical Review E*, 92(5):053304, 2015, [arXiv:1510.04650](#).
- [18] S. V. Isakov, G. Mazzola, V. N. Smelyanskiy, Z. Jiang, S. Boixo, H. Neven, and M. Troyer. Understanding quantum tunneling through quantum Monte Carlo simulations, 2015, [arXiv:1510.08057](#).
- [19] M. Jerrum, A. Sinclair, and E. Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *J. ACM*, 51(4):671–697, 2004.
- [20] T. Kadowaki and H. Nishimori. Quantum annealing in the transverse Ising model. *Phys. Rev. E*, 58:5355–5363, Nov 1998, [arXiv:cond-mat/9804280](#).
- [21] L. Kong and E. Crosson. The performance of the quantum adiabatic algorithm on spike Hamiltonians, 2015, [arXiv:1511.06991](#).
- [22] D. Levin, Y. Peres, and E. Wilmer. *Markov Chains and Mixing Times*. American Mathematical Soc., 2008.
- [23] R. Martoňák, G. E. Santoro, and E. Tosatti. Quantum annealing by the path-integral Monte Carlo method: The two-dimensional random Ising model. *Phys. Rev. B*, 66:094203, Sep 2002.
- [24] S. Muthukrishnan, T. Albash, and D. A. Lidar. Tunneling and speedup in quantum optimization for permutation-symmetric problems, 2015, [arXiv:1511.03910](#).
- [25] B. W. Reichardt. The quantum adiabatic optimization algorithm and local minima. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 502–510. ACM, 2004.
- [26] G. O. Roberts, J. S. Rosenthal, and P. O. Schwartz. Convergence properties of perturbed Markov chains. *J. Appl. Probab.*, 35(1):1–11, 03 1998.
- [27] M. Suzuki. Quantum statistical Monte Carlo methods and applications to spin systems. *Journal of Statistical Physics*, 43(5-6):883–909, 1986.
- [28] M. Suzuki, S. Miyashita, and A. Kuroda. Monte Carlo simulation of quantum spin systems. *Prog. Theor. Phys.*, 58(5):1377–1387, 1977.