# STATISTICAL ANALYSIS OF QUANTUM ANNEALING 

Yazhen Wang ${ }^{1}$, Shang $\mathrm{Wu}^{2}$ and Hongzhi Liu ${ }^{1}$<br>University of Wisconsin-Madison ${ }^{1}$ and Fudan University ${ }^{2}$

October 1, 2021
To appear in Statistica Sinica

Abstract: Quantum computation is based on quantum physics to build quantum devices for performing calculations and processing information. Although general-purpose quantum computers of large scale are still many years away, special-purpose quantum computers such as quantum annealers are being built with capabilities exceeding classical computers. These quantum annealers are special-purpose quantum computers created to realize quantum annealing. This paper explores quantum annealing and investigates its statistical properties. We establish a lower bound on the probability for quantum annealing to solve optimization problems. We examine physical devices and Monte Carlo simulations to implement quantum annealing and expand our understanding of the quantum annealing process.

Key words and phrases: Annealing, Combinatorial optimization, Hamiltonian, Ising model, Markov chain Monte Carlo, Quantum computation.

## 1. Introduction

As an optimization method, classical annealing is based on the analogy between the energy behavior of a complex physical system and the objective function of an optimization problem. Treating the objective function as the energy of the physical system, we convert the problem of minimizing the objective function into the problem of searching for minimum energy configurations (which are called ground states) of the physical system.

Simulated annealing (SA) is a well-known computer-based Monte Carlo simulation to mimic a system's behavior for finding its minimum energy configurations. The SA scheme is as follows. After identifying the objection function of the minimization problem with the energy of the physical system, we assign the physical system a temperature as an artificiallyintroduced control parameter. We select an initial temperature that is high relative to the system energy scale in order to induce thermal fluctuations and sample the energy configurations. Markov chain Monte Carlo (MCMC) simulations are used to carry out energy sampling and probabilistically explore the immense search space. As we decrease the temperature gradually from the initial value to zero, the system is driven to a state with the lowest energy value - namely, the minimum of the objective function - and thus we obtain a solution to the optimization problem. See Bertsimas and Tsitsiklis
(1993), Kirkpatrick et al. (1983), and Winker (2001) for more details.

Quantum annealing is the quantum analog of classical annealing. It is based on the process of a quantum physical system whose lowest energy provides a solution to the minimization problem under the study. A quantum state corresponding to the lowest energy is called a ground state of the system. Quantum annealing starts with a simple quantum system initialized in its ground state, and we then drive the simple system slowly towards the target complex system. According to the adiabatic quantum theorem (Farhi et al. 2000, 2001, 2002), when a quantum system is initialized in its ground state and then gradually evolves, it has a tendency to stay at a ground state. Hence, at the end of the quantum annealing process, measuring the state of the quantum system renders an answer to the original optimization problem with certain probability. More details can be found in Boixo et al. (2014), Brooke et al. (1999), Santoro et al. (2002), and Wang et al. (2016).

Both classical annealing and quantum annealing are powerful techniques to solve difficult optimization problems, whether they are utilized as tools by means of computer-based simulations or physical machines. The simulation approach applies 'escape' rules in Monte Carlo simulations to prevent the system from being trapped in local minima of an energy (or
objective) function, and eventually drives the system towards its lowest energy state with certain probability. The physical scheme utilizes a physical system or builds a device to engineer a physical system whose ground states represent the sought-after solution of an optimization problem.

The systems in both situations enable probabilistic exploration of their immense configuration spaces and ultimately 'freeze' in the global minima with certain probability. With enough repeated tries, each approach can find the global minimum and solve the optimization problem. The key difference between classical annealing and quantum annealing is thermal hopping used in classical annealing and quantum tunneling utilized in quantum annealing to escape from local minima and reach the global minimum. Quantum annealing has long been studied in quantum computation for building specialized quantum computers such as D-Wave annealers (Britton et al. (2012), Feynman (1982), Hu and Wang (2022), Martoňák et al. (2002), McGeoch (2014), Nielsen and Chuang (2010), Wang (2012), Wang et al. (2016), Wang and Song (2020), Wang (2021), and Wang and Liu (2022)). This paper, however, investigates quantum annealing from a statistical viewpoint. We provide a lower bound on the probability for the quantum annealing system to stay at a ground state at the end of the quantum annealing process, where the probability often refers to the suc-
cess probability of quantum annealing. We also discuss quantum annealing implementations and illustrate quantum tunneling through the lens of data augmentation.

The rest of the paper proceeds as follows. Section 2 reviews briefly the classical Ising model and SA. Section 3 explores quantum annealing. We establish its statistical properties and discusses its implementations by D-Wave devices and by MCMC based methods in the context of the Ising model. Section 4 features concluding remarks. We list topics for future research on quantum annealing. In particular, we point out that classical-computer-based MCMC simulations of quantum annealing can be used to provide some intuitive explanation of quantum tunneling in quantum annealing through a data augmentation connection. All proofs are relegated in Supplementary Appendix.

## 2. The Ising model and simulated annealing

We describe the Ising model using a graph $\mathcal{G}$ whose site and edge sets are denoted by $\mathcal{V}$ and $\mathcal{E}$, respectively. Each site is associated with a random variable taking values in $\{+1,-1\}$, and each edge specifies the interaction (or coupling) between the random variables on the two sites connected by the edge. Denote by $d$ the total number of sites in $\mathcal{G}$. For example, we
may consider $\mathcal{G}$ as a lattice with $d$ sites. Denote a configuration or state by $s=\left(s_{1}, s_{2}, \ldots, s_{d}\right)$ namely, a $d$-dimensional vector with site variables $s_{i}= \pm 1$. The classical Ising model has the following Hamiltonian (or energy):

$$
\begin{equation*}
\mathbf{H}_{I}^{c}(s)=-\sum_{\langle i, j\rangle \in \mathcal{E}} J_{i j} s_{i} s_{j}-\sum_{i \in \mathcal{V}} h_{i} s_{i}, \tag{2.1}
\end{equation*}
$$

where $J_{i j}$ denotes the strength of the interaction between sites $i$ and $j$ associated with edge $\langle i, j\rangle$ in graph $\mathcal{G}$, and $h_{i}$ gives the strength of the external local fields imposed on site $i$. We refer to a set of fixed values $\left\{J_{i j}, h_{i}\right\}$ as one instance of the Ising model. The Boltzmann (or Gibbs) distribution specifies the probability of a given configuration $s$ as follows:

$$
\begin{equation*}
P_{\beta}(s)=\frac{e^{-\beta \mathbf{H}_{I}^{c}(s)}}{Z_{\beta}} \quad \text { and } \quad Z_{\beta}=\sum_{s} e^{-\beta \mathbf{H}_{I}^{c}(s)}, \tag{2.2}
\end{equation*}
$$

where $\beta=\frac{1}{k_{B} T}, T$ denotes the absolute temperature of the system, and $k_{B}$ is a generic physical constant called the Boltzmann constant. The normalization constant $Z_{\beta}$ refers to the partition function of the Boltzmann distribution.

We may represent a combinatorial optimization problem through the Ising model with its objective function corresponding to the Hamiltonian $\mathbf{H}_{I}^{c}(s)$. Minimizing the objective function is equivalent to finding a configu-
ration $s^{*}$ with the minimum energy - namely, $s^{*}$ minimizes the Hamiltonian $\mathbf{H}_{I}^{c}(s)$ over all $s$. We refer to the configuration $s^{*}$ as a ground state of the Ising model.

Combinatorial optimization is computationally very hard, since there are $2^{d}$ configurations in the search space, which has an exponential increase in the system size $d$. SA is often applied to solve such optimization problems. It involves a temperature $T=T(t)$ as a decreasing function of evolution time $t$. A relatively high initial temperature $T(0)$ is set to induce thermal fluctuations and facilitate the exploration of the large search space. As the temperature gradually decreases, MCMC simulations are used to sample configurations, and we eventually drive the system to a ground state, which renders a solution to the minimization problem.

## 3. Quantum annealing

### 3.1 Theoretical analysis

We describe a quantum system by its quantum state and the dynamic evolution of the state, where the quantum state is characterized by a unit vector in a complex vector space, and the dynamic evolution of the state is governed by a Hermitian matrix called quantum Hamiltonian. To specify a quantum Hamiltonian for the quantum system that drives quantum
annealing, we need to introduce some notations. Define

$$
I_{j}=\left(\begin{array}{cc}
1 & 0  \tag{3.3}\\
0 & 1
\end{array}\right), \quad \sigma_{j}^{x}=\left(\begin{array}{cc}
0 & 1 \\
1 & 0
\end{array}\right), \quad \text { and } \sigma_{j}^{z}=\left(\begin{array}{cc}
1 & 0 \\
0 & -1
\end{array}\right), \quad j=1, \ldots, d,
$$

where $\sigma_{j}^{x}$ and $\sigma_{j}^{z}$ are called Pauli matrices in the $x$ and $z$ axes, respectively. The matrices and site index $j$ are used to serve as the quantum counterparts of the values $\{+1,-1\}$ for the binary random variables associated with sites in the classical Ising model.

Quantum annealing has the following quantum Hamiltonian:

$$
\begin{align*}
\mathbf{H}_{Q A}(t) & =A(t) \mathbf{H}_{I}^{q}+B(t) \mathbf{H}_{X},  \tag{3.4}\\
\mathbf{H}_{I}^{q} & =-\sum_{\langle i, j\rangle \in \mathcal{E}} J_{i j} \boldsymbol{\sigma}_{i}^{z} \boldsymbol{\sigma}_{j}^{z}-\sum_{i \in \mathcal{V}} h_{i} \boldsymbol{\sigma}_{i}^{z}, \quad \mathbf{H}_{X}=-\sum_{i \in \mathcal{V}} \boldsymbol{\sigma}_{i}^{x}, \tag{3.5}
\end{align*}
$$

where $A(t)$ and $B(t)$ are time-dependent smooth functions controlling the annealing schedules, $\mathcal{G}$ is the graph specified in the definition of the classical Ising model with site set $\mathcal{V}$ and edge set $\mathcal{E}, J_{i j}$ represents the interaction between sites $i$ and $j$ associated with edge $\langle i, j\rangle \in \mathcal{E}$, and $h_{i}$ is the strength of the external local fields imposed on site $i \in \mathcal{V}$. Here we use the convention in the quantum literature that $\boldsymbol{\sigma}_{i}^{z} \boldsymbol{\sigma}_{j}^{z}$ denotes the tensor product of $\sigma_{i}^{z}$ and $\sigma_{j}^{z}$ along with identity matrices in such a way that

$$
\boldsymbol{\sigma}_{i}^{z} \boldsymbol{\sigma}_{j}^{z} \equiv I_{1} \otimes \cdots \otimes I_{i-1} \otimes \sigma_{i}^{z} \otimes I_{i+1} \otimes \cdots \otimes I_{j-1} \otimes \sigma_{j}^{z} \otimes I_{j+1} \otimes \cdots \otimes I_{d}
$$

similarly, $\boldsymbol{\sigma}_{i}^{x}$ and $\boldsymbol{\sigma}_{i}^{z}$ stand for the following tensor products of $d$ matrices of size 2 :

$$
\begin{aligned}
& \boldsymbol{\sigma}_{i}^{x} \equiv I_{1} \otimes \cdots \otimes I_{i-1} \otimes \sigma_{i}^{x} \otimes I_{i+1} \otimes \cdots \otimes I_{d} \\
& \boldsymbol{\sigma}_{i}^{z} \equiv I_{1} \otimes \cdots \otimes I_{i-1} \otimes \sigma_{i}^{z} \otimes I_{i+1} \otimes \cdots \otimes I_{d}
\end{aligned}
$$

and Pauli matrices $\sigma_{i}^{x}$ and $\sigma_{i}^{z}$ in the tensor products are defined in (3.3).
Suppose that quantum annealing starts at $t=0$ and ends at $t=t_{f}$, where $t_{f}$ refers to the annealing duration. The quantum annealing schedules $A(t)$ and $B(t)$ typically satisfy that $A\left(t_{f}\right)=B(0)=0, A(t)$ is decreasing, and $B(t)$ is increasing. By controlling the annealing schedules $A(t)$ and $B(t)$, we allow quantum annealing to realize the gradual move of the Hamiltonian from $\mathbf{H}_{Q A}(0)=A(0) \mathbf{H}_{X}$ to $\mathbf{H}_{Q A}\left(t_{f}\right)=B\left(t_{f}\right) \mathbf{H}_{I}^{q}$. As $A(0)$ and $B\left(t_{f}\right)$ are known scalars, $\mathbf{H}_{Q A}(t)$ shares the same eigenvectors as $\mathbf{H}_{X}$ at the initial time $t=0$ and as $\mathbf{H}_{I}^{q}$ at the final time $t_{f}$, and their corresponding eigenvalues differ by factors of $A(0)$ and $B\left(t_{f}\right)$, respectively. Note that for a quantum system, its lowest energy is equal to the smallest eigenvalue of its Hamiltonian, with its ground state(s) being the eigenvector(s) corresponding to the smallest eigenvalue. Hamiltonian $\mathbf{H}_{X}=-\sum_{i=1} \boldsymbol{\sigma}_{i}^{x}$ is a simple Hermitian matrix with explicit expressions for its smallest eigenvalue and the corresponding eigenvector, and the quantum system governed by the Hamiltonian $\mathbf{H}_{X}$ can be easily prepared in its ground state.

The quantum annealing procedure is as follows. We begin with an initial quantum system prepared in its ground state; and the selected annealing schedules $A(t)$ and $B(t)$ allow us to engineer the quantum system to gradually move from $\mathbf{H}_{Q A}(0)=A(0) \mathbf{H}_{X}$ towards $\mathbf{H}_{Q A}=B\left(t_{f}\right) \mathbf{H}_{I}^{q}$. Therefore, the quantum annealing evolution driven by $\mathbf{H}_{Q A}(t)$ essentially evolves the quantum system from the initial system $\mathbf{H}_{X}$ initialized at its ground state to the final system $\mathbf{H}_{I}^{q}$. The quantum adiabatic theorem indicates that if the quantum system is initially started in its ground state, the system has the tendency to remain in the ground states of the instantaneous Hamiltonian during the Hamiltonian evolution. Thus, at the end of the quantum annealing evolution, we measure the system to find its lowest energy of $\mathbf{H}_{I}^{q}$ if the quantum system is in its ground state.

For the quantum Hamiltonian $\mathbf{H}_{I}^{q}$ defined in (3.5), its lowest energy is equal to its smallest eigenvalue. $\mathbf{H}_{I}^{q}$ involves only commuting diagonal matrices $\boldsymbol{\sigma}_{i}^{z}$, and its eigenvalues are equal to its diagonal entries, which in turn are exactly all the $2^{d}$ values of the classical Hamiltonian $\mathbf{H}_{I}^{c}(\mathbf{s})$ in (2.1) corresponding to the $2^{d}$ configurations ordered lexicographically. We have the following theorem to describe the relationship between the classical Hamiltonian $\mathbf{H}_{I}^{c}(\mathbf{s})$ and the quantum Hamiltonian $\mathbf{H}_{I}^{q}$.

Theorem 1. The eigenvalues of the quantum Hamiltonian $\mathbf{H}_{I}^{q}$ in (3.5) are
given by the $2^{d}$ values of the classical Hamiltonian $\mathbf{H}_{I}^{c}(\mathbf{s})$ in 2.1) evaluated at $2^{d}$ configurations $\mathbf{s} \in\{+1,-1\}^{d}$. In particular, the minimum of $\mathbf{H}_{I}^{c}(\mathbf{s})$ over $\mathbf{s} \in\{+1,-1\}^{d}$ is equal to the smallest eigenvalue of $\mathbf{H}_{I}^{q}$.

Theorem 1 shows that finding the minimal energy of the classical Ising model described by $\mathbf{H}_{I}^{c}$ is mathematically identical to finding the minimal energy of the quantum Hamiltonian $\mathbf{H}_{I}^{q}$. Thus, at the end of the quantum annealing process, measuring the quantum system renders a solution to the combinatorial minimization problem with the objective function $\mathbf{H}_{I}^{c}(s)$. Like SA, each run of quantum annealing can produce a solution to the optimization problem with some probability, and running quantum annealing many times enables us to solve the optimization problem.

According to the quantum adiabatic theorem (Aharonov et al. (2007), Born and Fock (1928), McGeoch (2014), Morita and Nishimori (2008) and Wang et al. (2015)), for appropriately chosen $A(t)$ and $B(t)$, we have that with some probability, the quantum annealing driven by (3.4) can find the global minimum of $\mathbf{H}_{I}^{c}(\mathbf{s})$ in (2.1) and solve the minimization problem at the final annealing time $t_{f}$. We therefore present the following theorem to provide a probability bound on successfully solving the optimization problem at the final annealing time $t_{f}$ using quantum annealing.

Theorem 2. Suppose that the quantum system associated with quantum
annealing is driven by $\mathbf{H}_{Q A}(t)$ as defined in (3.4). Then the probability that the lowest energy of $\mathbf{H}_{I}^{c}$ in (2.1) is obtained by measuring the system at the end of quantum annealing is bounded from below by
$\max \left\{\left[\left(1-\int_{0}^{1} \| \frac{d}{d u}\left(\left|v_{1}(u)\right\rangle, \cdots,\left|v_{r}(u)\right\rangle\right) \| d u\right)_{+}\right]^{2},\left\{v_{j}\right\}_{1 \leq j \leq r}, 1 \leq r \leq \zeta\right\}$,
where $\zeta$ denotes the number of ground states for the quantum Hamiltonian $\mathbf{H}_{I}^{q}$ in (3.5), $(x)_{+}$denotes the positive part of $x$-namely, $(x)_{+}$is equal to $x$ if $x \geq 0$ and zero otherwise, and $\left(\left|v_{1}(u)\right\rangle, \cdots,\left|v_{r}(u)\right\rangle\right)$ is a matrix formed by $r$ column vectors $\left|v_{1}(u)\right\rangle, \cdots,\left|v_{r}(u)\right\rangle$ that are defined as follows. Denote by $\xi_{1}(u) \leq \xi_{2}(u) \leq \cdots \leq \xi_{2^{d}}(u)$ the $2^{d}$ instantaneous eigenvalues of $\mathbf{H}_{Q A}\left(u t_{f}\right)$ listed in an increasing order along with the corresponding $2^{d}$ normalized eigenvectors $v_{1}(u), v_{2}(u), \cdots, v_{2^{d}}(u)$, where for any eigenvalue with a multiplicity greater than 1, the eigenvalue is repeated in the list with the number of repetitions equal to its multiplicity, and the multiple eigenvectors corresponding to the same eigenvalue are ordered as a group-that is, their positions in the list are exchangeable, and the maximum in (3.6) is taken over $1 \leq r \leq \zeta$ and possible group orderings of $v_{j}(u)$.

Furthermore, assume that $\lambda_{1}(u)-\lambda_{0}(u)$ is bounded below from zero uniformly over $u \in[0,1]$, where $\lambda_{0}(u)$ and $\lambda_{1}(u)$ denote the smallest and second smallest instantaneous eigenvalues of $\mathbf{H}_{Q A}\left(u t_{f}\right)$, respectively. Then
the probability that the quantum annealing procedure can find the lowest energy of $\mathbf{H}_{I}^{c}$ in (2.1) is bounded from below by

$$
\begin{equation*}
1-2^{d} \zeta \max _{u \in[0,1]}\left\{\frac{1}{\lambda_{1}(u)-\lambda_{0}(u)}\left\|\frac{d \mathbf{H}_{Q A}\left(u t_{f}\right)}{d u}\right\|\right\}^{2} \tag{3.7}
\end{equation*}
$$

where $\|\cdot\|$ denotes the matrix spectral norm.

The ground state success probability for quantum annealing is usually derived under the unique ground state condition, in the asymptotic sense that we obtain some expressions or bounds for the leading terms of the ground state success probability by taking $t_{f}$ to infinity (Aharonov et al. (2007), Born and Fock (1928), McGeoch (2014) and Morita and Nishimori (2008)). The probability lower bounds in (3.6) and (3.7) are for finite $t_{f}$ without the unique ground state restriction. The results established in Theorems 1 and 2 together with the existing asymptotic results provide the theoretical foundation that makes it possible for the quantum annealing process, driven by (3.4), to find the global minimum of $\mathbf{H}_{I}^{c}(\mathbf{s})$ and solve the minimization problem with a certain probability.

### 3.2 Implementation via $D$-Wave machines

D-Wave machines are commercially available computing hardware devices that have been built via superconducting technology to implement quantum annealing. They are analog computers made specifically to handle
combinatorial optimization linked to the classical Ising model, and their quantum processor chips are based on associated graphs that specify quantum annealing with possible adjustments to the standard $20 \mu s$ duration and annealing schedules, as described in Sections 2 and 3. Since 2011, five generations of D-Wave computing machines have been developed, with the number of graph sites equal to $128,512,1152,2048$, and 5640 , respectively, where the graphs used were the Chimera graph for the first four generations and the Pegasus graph for the last generation. D-Wave machines have been applied to solve combinatorial optimization problems in research studies and real applications. While there is no quantum speedup found in D-Wave machines, it has been demonstrated that quantum annealing can be much faster than classical annealing for solving certain optimization problems. In particular, specific examples are provided in Denchev et al. (2015) and Farhi et al. (2002) that illustrate the advantage of quantum annealing over classical annealing. More details can be found in Boixo et al. (2014), Boixo et al. (2015 a, b), Dattani et al. (2019), Denchev et al. (2015), Farhi et al. (2002), Hastings (2020), Hen et al. (2015), Johnson et al. (2011), Katzgraber et al. (2014), Lanting et al. (2014), O'Gorman et al. (2014), Perdomo-Ortiz et al. (2012), Perdomo-Ortiz et al. (2014), Rieffel et al. (2014), and Wang et al. (2016).

### 3.3 Implementation via path-integral and MCMC simulations

Various methods have been established via asymptotic expansion to approximately implement quantum annealing using MCMC simulations on classical computers. The main idea behind these approaches is the path-integral formulation with the Trotter formula (see Kato (1978), Suzuki (1976), and Trotter (1959)) to show that the quantum system driven by quantum annealing Hamiltonian $\mathbf{H}_{Q A}(t)$ in (3.4) is asymptotically equivalent to a classical anisotropic Ising model. For simplicity, consider the case of $h_{i}=0$. The classical anisotropic Ising model with temperature $\tau T$ has the following Hamiltonian:

$$
\begin{equation*}
\mathbf{H}_{a I}^{c}(\mathbf{s})=-\sum_{l=1}^{\tau}\left[B(t) \sum_{(i, j) \in \mathcal{E}(\mathcal{G})} J_{i j} s_{i l} s_{j l}+J(t) \sum_{j \in \mathcal{V}(\mathcal{G})} s_{j l} s_{j, l+1}\right], \tag{3.8}
\end{equation*}
$$

where $\tau$ is an integer, $s_{i l}$ are random variables taking values in $\{+1,-1\}$, $J_{i j}$ are the regular couplings along the direction of the original Ising model, and $l$ is the index for an extra new direction that is often referred to as the imaginary-time direction with

$$
J(t)=-\frac{\tau T}{2} \ln \left[\tanh \left(\frac{A(t)}{\tau T}\right)\right]
$$

as the coupling along the imaginary-time direction. Let $\mathbf{s}_{l}=\left\{s_{i l}, i=\right.$ $1, \ldots, b\}, l=1, \ldots, \tau$. We call $\mathbf{s}_{l}$ the $l$-th Trotter slice. Similar to the SA case, standard MCMC techniques are used to carry out simulations of the classical anisotropic Ising model with Hamiltonian $\mathbf{H}_{a I}^{c}$ on classical computers. With the Trotter slices $s_{i l}, i=1, \ldots, d, l=1, \ldots, \tau$, generated from the MCMC simulations we use the majority rule to produce each site value $s_{i}=$ sign of the sum $s_{i 1}+\cdots+s_{i \tau}$ and form a configuration $s=\left\{s_{i}, i=1, \ldots, d\right\}$ as a solution to minimize $\mathbf{H}_{I}^{c}(\cdot)$ defined in (2.1). The classical-computer-based simulation approach refers to simulated quantum annealing (SQA). Figure 1 illustrates a lattice structure for a classical Ising model and a Trotter slice structure for its corresponding classical anisotropic Ising model. As the Trotter slice graph is much more complex than the lattice, MCMC simulations of the two Ising models with different structures evidently indicate that SQA is much slower than SA. Consequently SQA is often used for benchmarking in quantum computation and for gaining some insightful understanding of quantum annealing. See Hu and Wang (2021), Martoňák et al. (2002), Morita and Nishimori (2008), and Wang et al. (2016) and the references therein for more details.

## (a) Lattice Structure


(b) Trotter Slices


Figure 1: Plots of lattice structures for a classical Ising system and its corresponding classical anisotropic Ising system, where (a) and (b) illustrate a lattice as a simple graph for the Ising system and its corresponding graph with four Trotter slices for the anisotropic Ising system, respectively.

### 3.4 Discussion of theoretical analysis and practical implementation

Quantum annealing was proposed as a potential quantum means to solve optimization problems. The theoretical results establish the foundation for quantum annealing, and implementations by physical quantum annealers and path-integral based MCMC simulations demonstrate its practical usefulness. It is interesting but challenging to accurately connect the theoretical results with practical performances of the implementations by physical or simulation means, because quantum annealers are noisy, and MCMC simulations are approximation methods based on asymptotics.

Specifically, theoretical results indicates certain probabilities for quantum annealing to find ground states, and quantum annealers and SQA practically confirm some success probabilities to obtain ground states by quantum annealing. The asymptotic justification derives positive probabilities for quantum annealing to find ground states by taking annealing time $t_{f}$ to go to infinity. The probability lower bound in (3.7) for finite $t_{f}$ is given by $1-2^{d} \zeta \aleph$, where

$$
\begin{equation*}
\aleph=\max _{u \in[0,1]}\left\{\frac{1}{\lambda_{1}(u)-\lambda_{0}(u)}\left\|\frac{d \mathbf{H}_{Q A}\left(u t_{f}\right)}{d u}\right\|\right\}^{2} . \tag{3.9}
\end{equation*}
$$

3.4 Discussion of theoretical analysis and practical implementation

Note that

$$
\begin{equation*}
\frac{d \mathbf{H}_{Q A}\left(u t_{f}\right)}{d u}=\frac{d A\left(u t_{f}\right)}{d u} \mathbf{H}_{X}+\frac{d B\left(u t_{f}\right)}{d u} \mathbf{H}_{I}^{q}, \tag{3.10}
\end{equation*}
$$

which depends on $u$ only through the derivatives of annealing schedules $A(t)$ and $B(t)$. By choosing appropriate $A(t)$ and $B(t)$, we can ensure that the probability lower bound in 3.7 is positive and thus guarantee that quantum annealing can find the lowest energy of $\mathbf{H}_{I}^{c}$ with some probability. Indeed, from (3.9) and (3.10), we obtain

$$
\begin{align*}
\aleph \leq & \max _{u \in[0,1]}\left\{\left[\lambda_{1}(u)-\lambda_{0}(u)\right]^{-2}\left[\left|\frac{d A\left(u t_{f}\right)}{d u}\right|\left\|\mathbf{H}_{X}\right\|+\left|\frac{d B\left(u t_{f}\right)}{d u}\right|\left\|\mathbf{H}_{I}^{q}\right\|\right]^{2}\right\} \\
\leq & \left\{\min _{u \in[0,1]}\left[\lambda_{1}(u)-\lambda_{0}(u)\right]\right\}^{-2}\left[\left\|\mathbf{H}_{X}\right\|+\left\|\mathbf{H}_{I}^{q}\right\|\right]^{2} \\
& \max _{u \in[0,1]}\left\{\left|\frac{d A\left(u t_{f}\right)}{d u}\right| \bigvee\left|\frac{d B\left(u t_{f}\right)}{d u}\right|\right\}^{2}, \tag{3.11}
\end{align*}
$$

where $\left\|\mathbf{H}_{X}\right\|$ and $\left\|\mathbf{H}_{I}^{q}\right\|$ are the spectral norms of $\mathbf{H}_{X}$ and $\mathbf{H}_{I}^{q}$, and $\vee$ stands for the maximum. For a given quantum annealing setup, we have fixed $d$, $\zeta,\left\|\mathbf{H}_{X}\right\|,\left\|\mathbf{H}_{I}^{q}\right\|$, and the specified minimum of $\lambda_{1}(u)-\lambda_{0}(u)$ over $u \in[0,1]$. Hence, (3.11) indicates that it is theoretically possible to choose schedule functions $A(\cdot)$ and $B(\cdot)$ with small enough absolute derivatives in order to make $\aleph<1 /\left[\zeta 2^{d}\right]$, and thus produce a positive probability lower bound $1-2^{d} \zeta \aleph$. More precisely, if the schedule functions $A(\cdot)$ and $B(\cdot)$ satisfy that for $u \in[0,1]$,

$$
\left|\frac{d A\left(u t_{f}\right)}{d u}\right| \bigvee\left|\frac{d B\left(u t_{f}\right)}{d u}\right|<\frac{\min _{u \in[0,1]}\left[\lambda_{1}(u)-\lambda_{0}(u)\right]}{2^{d / 2} \sqrt{\zeta}\left[\left\|\mathbf{H}_{X}\right\|+\left\|\mathbf{H}_{I}^{q}\right\|\right]}
$$

then $1-2^{d} \zeta \aleph>0$.

## 4. Concluding remarks

This paper explores quantum annealing along with classical annealing for solving combinatorial optimization problems. We establish a probability lower bound for quantum annealing to find a solution to an optimization problem. We discuss implementations of quantum annealing via D-Wave physical devices and path-integral based MCMC simulations. Our study provides a theoretical foundation for quantum annealing to solve combinatorial optimization.

### 4.1 Future research topics in quantum annealing

Quantum annealing plays an important role in quantum computation and is relatively new in statistics. This paper raises statistical issues in the study of quantum annealing and leaves open problems for the future study. For example, we may investigate sharpness of the lower bounds in Theorem 2 and their relationship with asymptotic results on the ground state success probability as well as their practical implications; we may study quantum processes related to the unitary dynamic evolution of quantum annealing and classical processes associated with MCMC implementations of quantum
annealing as well as the statistical relationship of the two types of processes; we may explore the performance of D-Wave machines-in particular the noise impact on solving optimization problems. As a case in point, below we explore a key difference between thermal fluctuations in classical annealing and quantum tunneling in quantum annealing and propose combining the data augmentation scheme and the Monte Carlo implementation approach to intuitively illustrate quantum tunneling from a statistical perspective.

### 4.2 Quantum tunneling and data augmentation

Unlike classical annealing, which utilizes thermal fluctuations to make the Ising system jump from state to state over intermediate energy barriers and search for the desired lowest-energy state, quantum annealing relies on quantum-mechanical fluctuations instead of thermal jumps for state transitions. The two terms $\mathbf{H}_{I}^{q}$ and $\mathbf{H}_{X}$ of $\mathbf{H}_{Q A}$ in (3.4) and (3.5) are noncommutable matrices and represent the potential and kinetic energies of the underlying quantum system, respectively, and the move from $\mathbf{H}_{X}$ to $\mathbf{H}_{I}^{q}$ through $\mathbf{H}_{Q A}$ during the quantum annealing process can be physically accomplished by engineering magnetic fields to induce quantum fluctuations via quantum tunneling. Quantum tunneling refers to the quantum phenomenon that particles tunnel through a barrier in an impossible con-
dition under classical physics. It permits the annealing process to search for distinct states by traveling directly through energy barriers, instead of hopping over them thermally in the classical annealing case. Quantum tunneling can be explained by the Heisenberg uncertainty principle and the wave-particle duality of matter in quantum physics but can not adequately explained by classical physics (see Sakurai and Napolitano (2010) and Shankar (1994)).

Sections 3.2 and 3.3 indicate that quantum annealing may be implemented by physical devices or MCMC simulations. D-Wave physical devices aim to realize the unitary dynamic evolution of quantum annealing by natural Schrödingier dynamics, and path-integral based MCMC simulations intend to approximate the dynamic evolution of quantum annealing by artificial time evolutions of Monte Carlo dynamics. The two approaches may be connected through some statistical sampling distributions for the quantum probability model associated with the unitary dynamic evolution in annealing and the statistical model linked to SQA. In particular, while it is hard to explain quantum tunneling in physical implementation of quantum annealing, we use SQA to provide an intuitive statistical illustration of tunneling. It is noted that the key difference between the classical Ising Hamiltonians $\mathbf{H}_{I}^{c}(s)$ in 2.1) and $\mathbf{H}_{a I}^{c}(\mathbf{s})$ in (3.8) is the extra imaginary-
time direction in $\mathbf{H}_{a I}^{c}(\mathbf{s})$. The well-known data augmentation scheme can be used to accommodate the extra direction as follows. We may introduce an augmented random variable to represent the imaginary-time direction and describe SQA in an augmented search space. The tunneling, which is hard to explain in the original search space, may have an intuitive explanation in the larger search space. For example, Figure 2 illustrates that tunneling is a possible way to travel through a barrier in two dimensions; however, in the augmented three dimensions, there is a natural route to bypass the barrier. The representation of SQA via data augmentation may offer an intuitive explanation for the tunneling effect in quantum annealing.
(a) Quantum Tunneling

(b) Front and Back Views of Tunneling in the Augmented Space


Figure 2: Plots of quantum tunneling, where (a) and (b) illustrate quantum tunneling in quantum annealing for getting over an energy barrier and a corresponding possible tunneling effect in the augmented space, respectively.

## 5. Acknowledgements

The research of Yazhen Wang was supported in part by NSF grants DMS1707605 and DMS-1913149. The authors thank Co-Editor Xiaotong Shen,
an associate editor, and two anonymous referees for comments and suggestions, which led to significant improvements in both substance and the presentation of the paper.

## References

Albash, T., Rønnow, T. F., Troyer, M. and Lidar, D. A. (2014). Reexamining classical and quantum models for the D-Wave One processor: the role of excited states and ground state degeneracy. arXiv:1409.3827v1.

Arute, F., Arya, K., Babbush, R. et al. (2019). Quantum supremacy using a programmable superconducting processor. Nature 574, 505-510. https://doi.org/10.1038/s41586-019-1666-5.

Aspuru-Guzik, A., Dutoi, A. D., Love, P. J. and Head-Gordon, M. (2005). Simulated quantum computation of molecular energies. Science 309, 17041707.

Bertsimas, D. and Tsitsiklis, J. (1993). Simulated annealing. Statistical Science 8, 10-15.

Blanes, S., Casas, F., Oteo, J. A. and Ros, J. (2009). The Magnus expansion and some of its applications. Physics Reports 470, 151-238.

Boixo, S., Albash, T., Spedalieri, F. M., Chancellor, N. and Lidar, D. A. (2014). Experimental signature of programmable quantum annealing. Nature Communications 4, 2067.

Boixo, S., Rønnow, T. F., Isakov, S. V., Wang, Z., Wecker, D., Lidar, D. A.,

Martinis, J. M. and Troyer, M. (2014). Evidence for quantum annealing with more than one hundred qubits. Nature Physics 10, 218-224.

DOI: 10.1038/NPHYS2900.
Boixo, S., Smelyanskiy, V. N., Shabani, A., Isakov, S. V., Dykman, M., Denchev, V. S., Amin, M., Smirnov, A., Mohseni, M. and Neven, H. (2015 a). Computational role of collective tunneling in a quantum annealer. arXiv:1411.4036v2.

Boixo, S., Smelyanskiy, V. N., Shabani, A., Isakov, S. V., Dykman, M., Denchev, V. S., Amin, M., Smirnov, A., Mohseni, M. and Neven, H. (2015 b). Computational role of multiqubit tunneling in a quantum annealer. arXiv:1502.05754v1.

Britton, J. W., Sawyer, B.C., Keith, A., Wang, C.-C.J., Freericks, J. K., Uys, H., Biercuk, M. J. and Bollinger, J. J. (2012). Engineered 2D Ising interactions on a trapped-ion quantum simulator with hundreds of spins. Nature 484, 489-492.

Brooke, J., Bitko, D., Rosenbaum, T. F. and Aeppli, G. (1999). Quantum annealing of a disordered magnet. Science 284, 779-781.

Clarke, J. and Wilhelm, F. (2008). Superconducting quantum bits. Nature 453, 1031-1042.

Dattani, N., Szalay, S. and Chancellor, N. (2019). Pegasus: The second connectivity graph for large-scale quantum annealing hardware. arXiv:1901.07636.

Denchev, V., Boixo, S., Isakov, S., Ding, N., Babbush, R., Smelyanskiy, V.,

Martinis, J. and Neven, H. (2015). What is the computational value of finite range tunneling? Physical Review X 6, 031015.

Farhi, E., Goldstone, J., Gutmann, S., Lapan, J., Lundgren, A. and D. Preda, D. (2001). A quantum adiabatic evolution algorithm applied to random instances of an NP-complete problem. Science 292, 472-476.

Farhi, E., Goldstone, J., Gutmann, S. and Sipser, M. (2000). Quantum computation by adiabatic evolution. arXiv:0001106v1.

Farhi, E., Goldstone, J., Gutmann, S. and Sipser, M. (2002). Quantum adiabatic evolution algorithms versus simulated annealing. arXiv:0201031v1.

Farhi, E., Goldstone, J. and Gutmann, S. (2014). A quantum approximate optimization algorithm. arXiv:1411.4028.

Feynman, R. P. (1982). Simulating physics with computers. International Journal of Theoretical Physics 21, 467-488.

Hajek, B. (1988). Cooling schedules for optimal annealing. Mathematics of Operations Research 13, 311-329.

Hastings, M. B. (2020). The power of adiabatic quantum computation with no sign problem. arXiv:2005.03791.

Hen, I., Job, J., Albash, T., Rønnow, T. F., Troyer, M. and Lidar, D. A. (2015). Probing for quantum speedup in spin glass problems with planted solutions. arXiv:1502.01663v2

Holevo, A. S. (1982). Probabilistic and Statistical Aspects of Quantum Theory. Springer, North-Holland, Amsterdam.

Hu, J. and Wang, Y. (2021). Quantum annealing via path-integral Monte Carlo with data augmentation. Journal of Computational and Graphical Statistics 30, 284-296.

Johnson, M. W., M. H. S. Amin, S. Gildert, T. Lanting, F. Hamze, N. Dickson, R. Harris, A. J. Berkley, J. Johansson, P. Bunyk, E. M. Chapple, C. Enderud, J. P. Hilton, K. Karimi, E. Ladizinsky, N. Ladizinsky, T. Oh, I. Perminov, C. Rich1, M. C. Thom, E. Tolkacheva, C. J. S. Truncik, S. Uchaikin, J. Wang, B. Wilson and G. Rose (2011). Quantum annealing with manufactured spins. Nature 473, 194-198.

Katzgraber, H. G., Hamze, F. and Andrist, R. S. (2014). Glassy Chimeras could be blind to quantum speedup: Designing better benchmarks for quantum annealing machines. Physical Review X 4, 021008.

Kato, T. (1978). Trotter's product formula for an arbitrary pair of self-adjoint contraction semi- groups. In Topics in Functional Analysis (Essays Dedicated to M. G. Krein on the Occasion of His 70th Birthday). Adv. in Math. Suppl. Stud. 3 185-195. Academic Press, Boston, MA.

Kirkpatrick, S., Gelatt, C. D., Vecchi, M. P. (1983). Optimization by simulated annealing. Science 220, 671-680.

Lanting, T., A. J. Przybysz, A. Yu. Smirnov, F. M. Spedalieri, M. H. Amin, A. J. Berkley, R. Harris, F. Altomare, S. Boixo, P. Bunyk, N. Dickson, C. Enderud, J. P. Hilton, E. Hoskinson, M. W. Johnson, E. Ladizinsky, N. Ladizinsky, R. Neufeld, T. Oh, I. Perminov, C. Rich, M. C.

Thom, E. Tolkacheva, S. Uchaikin, A. B. Wilson and G. Rose (2014). Entanglement in a quantum annealing processor. Physical Review X 4, 021041.

Majewski, J., Li, H. and Ott, J. (2001). The Ising model in physics and statistical genetics. Am. J. Hum. Genet. 69 853-862.

Martin-Mayor, V. and Hen, I. (2015). Unraveling quantum annealers using classical hardness. arXiv:1502.02494v1.

Martoňák, R., Santoro, G. E. and Tosatti, E. (2002). Quantum annealing by the path-integral Monte Carlo method: The two-dimensional random Ising model. Physical Review B 66, 094203.

McGeoch, C. C. (2014). Adiabatic Quantum Computation and Quantum Annealing. Synthesis Lectures on Quantum Computing. Morgan \& Claypool.

Morita, S. and Nishimori, H. (2008). Mathematical foundation of quantum annealing. Journal of Mathematical Physics 49, 125210.

Nielsen, M. and Chuang, I. (2010). Quantum Computation and Quantum Information. 10th Anniversary edition. Cambridge: Cambridge University Press.

O'Gorman, B., Babbush, R., Perdomo-Ortiz, A., Aspuru-Guzik, A., and Smelyanskiy, V. (2014). Bayesian network structure learning using quantum annealing. arXiv:1407.3897v2.

Preskill, J. (2018). Quantum computing in the NISQ era and beyond. Quantum

2 (2018): 79.
Perdomo-Ortiz, A., Dickson, N., Drew-Brook, M., Rose, G. and Aspuru-Guzik, A. (2012). Finding low-energy conformations of lattice protein models by quantum annealing. Scientific Reports 2, 571 .

Perdomo-Ortiz, A., Fluegemann, J., Narasimhan, S., Biswas and R., Smelyanskiy, V. N. (2014). A quantum annealing approach for fault detection and diagnosis of graph-based systems. arXiv:1406.7601v2

Pudenz, K. L., Albash, T. and Lidar, D. A. (2014). Error corrected quantum annealing with hundreds of qubits. Nature Communications 5, 3243, arXiv:1307.8190

Rieffel, E., Venturelli, D., O’Gorman, B., Do, M., Prystay, E., and Smelyanskiy, V. (2014). A case study in programming a quantum annealer for hard operational planning problems. arXiv:1407.2887v1.

Santoro, G. E., Martoňák, R., Tosatti, E. and Car, R. (2002). Theory of quantum annealing of an Ising spin glass. Science 295, 2427.

Sakurai, J. J. and Napolitano, J. (2010). Modern Quantum Mechanics. Reading: Addison-Wesley. Second edition.

Shankar, R. (1994). Principles of Quantum Mechanics. Springer. Second edition.
Smolin, J. A. and Smith, G. (2014). Classical signature of quantum annealing. Frontiers in Physics 2, 52. arXiv: 1305.4904v1.

Suzuki, M. (1976). Generalized Trotter's formula and systematic approximants of exponential operators and inner derivations with applications to
many-body problems. Comm. Math. Phys. 51: 183-190.
Trotter, H. F. (1959). On the product of semi-groups of operators. Proc. Amer. Math. Soc. 10, 545-551.

Venturelli, D., Mandrá, S., Knysh, S., O’Gorman, B., Biswas, R., Smelyanskiy, V. N. (2014). Quantum optimization of fully-connected spin glasses. arXiv:1406.7553v1.

Vinci, W., Albash, T., Mishra, A., Warburton, P. A., and Lidar, D. A. (2014). Distinguishing classical and quantum models for the D-Wave device. arXiv: 1403.4228 v 1 .

Wang, Y. (2012). Quantum computation and quantum information. Statistical Science 27, 373-394..

Wang, Y. (2021). When quantum computation meets data science: Making data science quantum. Harvard Data Science Review (under review).

Wang, Y. and Liu, H. (2022). Quantum computing in a statistical context. Annual Review of Statistics and Its Application (to appear).

Wang, Y. and Song, X. (2020). Quantum science and quantum technology. Statistical Science 35, 51-74.

Yazhen Wang and Hongzhi Liu, University of Wisconsin-Madison
E-mail: yzwang@stat.wisc.edu, hliu438@wisc.edu
Shang Wu, Fudan University
E-mail: shangwu@fudan.edu.cn

## 6. Supplementary Appendix: Proofs

### 6.1 Mathematical notations and quantum concepts

Denote by superscripts $*, I$ and $\dagger$ the conjugate of a complex number, the transpose of a vector or matrix, and the conjugate transpose operation, respectively. Denote by $\mathbb{R}$ the set of all real numbers. We use $\mathbb{C}^{d}$ to denote the $d$-dimensional complex space. For a vector $\psi$ in $\mathbb{C}^{d}$, we comply with the convention in quantum mechanics and quantum computation that uses Dirac notations ket $|\cdot\rangle$ and bra $\langle\cdot|$ to indicate that $|\psi\rangle$ and $\langle\psi|$ are column and row vectors, respectively. We define a natural inner product in $\mathbb{C}^{d}$ to be $\langle u \mid v\rangle=\sum_{j=1}^{d} u_{j}^{*} v_{j}=\left(u_{1}^{*}, \cdots, u_{d}^{*}\right)\left(v_{1}, \cdots, v_{d}\right)^{\prime}$, where $\langle u|=\left(u_{1}, \cdots, u_{d}\right)$ and $|v\rangle=\left(v_{1}, \cdots, v_{d}\right)^{\prime}$, and the modulus (or norm) $\|u\|=\sqrt{\langle u \mid u\rangle}$. We call a matrix $\mathbf{A}$ Hermitian if $\mathbf{A}=\mathbf{A}^{\dagger}$, and a matrix $\mathbf{U}$ is said to be unitary if $\mathbf{U} \mathbf{U}^{\dagger}=\mathbf{U}^{\dagger} \mathbf{U}=\mathbf{I}$, where $\mathbf{I}$ is an identity matrix. For any matrix A, we denote its spectral norm by $\|\mathbf{A}\|$.

Consider a quantum system in a state $|\psi\rangle$, where the quantum state $|\psi\rangle$ is a unit vector in $\mathbb{C}^{d}$. Measurements on the quantum system are described through observables, where an observable is defined as a Hermitian matrix on $\mathbb{C}^{d}$. Given an observable $\mathbf{M}$, we assume that it has the following eigen-decomposition,

$$
\begin{equation*}
\mathbf{M}=\sum_{a=1}^{d} x_{a} \mathbf{Q}_{a} \tag{6.12}
\end{equation*}
$$

where $x_{a}$ are the real eigenvalues of $\mathbf{M}$, and $\mathbf{Q}_{a}$ are the corresponding projections onto the eigen-spaces of $\mathbf{M}$. When performing a measurement on $\mathbf{M}$ for the quantum system in state $|\psi\rangle$, we adopt a measure space $(\Omega, \mathcal{F})$ to accommodate its possible measurement outcomes, where the measurement result is treated as a
random variable $X$ with a probability distribution $P_{\psi}$ that is defined on $(\Omega, \mathcal{F})$. The random variable $X$ takes values in $\left\{x_{1}, x_{2}, \cdots, x_{d}\right\}$, and the probability of obtaining measurement outcome $x_{a}$ is given by

$$
P_{\psi}\left(X=x_{a}\right)=\langle\psi| \mathbf{Q}_{a}|\psi\rangle, \quad a=1,2, \cdots, d .
$$

For the quantum annealing case, the quantum system is described by the quantum Hamiltonian $\mathbf{H}_{Q A}(t)$, and its state at time $t$ is $|\psi(t)\rangle$. At the end of the annealing process, $t=t_{f}$, and $\mathbf{H}_{Q A}\left(t_{f}\right)=B\left(t_{f}\right) \mathbf{H}_{I}^{q}$. As $B\left(t_{f}\right)$ is a known scalar, we may ignore it and treat $\mathbf{H}_{Q A}\left(t_{f}\right)$ the same as $\mathbf{H}_{I}^{q}$. Measuring the quantum system to find its energy at $t_{f}$ corresponds to performing a measurement on $\mathbf{H}_{I}^{q}$ for the quantum system in the state $\left|\psi\left(t_{f}\right)\right\rangle$. According to the described measuring scheme, we may find the lowest energy of the quantum system with probability $\left\langle\psi\left(t_{t}\right)\right| \mathbf{Q}\left|\psi\left(t_{f}\right)\right\rangle$, where $\mathbf{Q}$ denotes the projection onto the eigen-space of $\mathbf{H}_{I}^{q}$ corresponding to the smallest eigenvalue.

Denote by C's generic constants whose values are free of model parameters and may change from appearance to appearance.

### 6.2 Proof of Theorem 1

Let $\mathbf{e}_{j,+1}=(1,0)^{\dagger}$ and $\mathbf{e}_{j,-1}=(0,1)^{\dagger}, j=1, \cdots, d$. Then $\mathbf{e}_{j, \pm 1}$ are eigenvectors of Pauli matrix $\boldsymbol{\sigma}_{j}^{z}$ corresponding to eigenvalues $\pm 1$. For the classical Ising model, given a configuration $\mathbf{s}=\left(s_{1}, \cdots, s_{b}\right)$ with energy $\mathbf{H}_{I}^{c}(\mathbf{s})$ in 2.1), define a unit vector in $\mathbb{C}^{2^{d}}, \mathbf{e}_{\mathbf{s}}=\mathbf{e}_{1, s_{1}} \otimes \mathbf{e}_{2, s_{2}} \cdots \otimes \mathbf{e}_{b, s_{d}}$. We show that $\mathbf{e}_{\mathbf{s}}$ is an eigenvector of
$\mathbf{H}_{I}^{q}$ with corresponding eigenvalue $\mathbf{H}_{I}^{c}(\mathbf{s})$. Indeed,

$$
\begin{aligned}
& \boldsymbol{\sigma}_{j}^{z} \mathbf{e}_{j, s_{j}}=s_{j} \mathbf{e}_{j, s_{j}}, \\
& \left(\mathbf{I}_{2} \otimes \cdots \mathbf{I}_{2} \otimes \boldsymbol{\sigma}_{j}^{z} \otimes \mathbf{I}_{2} \cdots \otimes \mathbf{I}_{2}\right)\left(\mathbf{e}_{1, s_{1}} \otimes \mathbf{e}_{2, s_{2}} \cdots \otimes \mathbf{e}_{d, s_{d}}\right) \\
& =s_{j} \mathbf{e}_{1, s_{1}} \otimes \mathbf{e}_{2, s_{2}} \cdots \otimes \mathbf{e}_{d, s_{d}}=s_{j} \mathbf{e}_{\mathbf{s}}, \\
& \left(\mathbf{I}_{2} \otimes \cdots \mathbf{I}_{2} \otimes \boldsymbol{\sigma}_{i}^{z} \otimes \mathbf{I}_{2} \cdots \mathbf{I}_{2} \otimes \boldsymbol{\sigma}_{j}^{z} \otimes \mathbf{I}_{2} \cdots \otimes \mathbf{I}_{2}\right)\left(\mathbf{e}_{1, s_{1}} \otimes \mathbf{e}_{2, s_{2}} \cdots \otimes \mathbf{e}_{d, s_{d}}\right) \\
& =s_{i} s_{j} \mathbf{e}_{1, s_{1}} \otimes \mathbf{e}_{2, s_{2}} \cdots \otimes \mathbf{e}_{d, s_{d}}=s_{i} s_{j} \mathbf{e}_{\mathbf{s}},
\end{aligned}
$$

and

$$
\begin{aligned}
& \mathbf{H}_{I}^{q} \mathbf{e}_{\mathbf{s}}=-\sum_{(i, j) \in \mathcal{E}(\mathcal{G})} J_{i j}\left(\mathbf{I}_{2} \otimes \cdots \mathbf{I}_{2} \otimes \boldsymbol{\sigma}_{i}^{z} \otimes \mathbf{I}_{2} \cdots \mathbf{I}_{2} \otimes \boldsymbol{\sigma}_{j}^{z} \otimes \mathbf{I}_{2} \cdots \otimes \mathbf{I}_{2}\right) \\
& \quad\left(\mathbf{e}_{1, s_{1}} \otimes \mathbf{e}_{2, s_{2}} \cdots \otimes \mathbf{e}_{d, s_{d}}\right) \\
& -\sum_{j \in \mathcal{V}(\mathcal{G})} h_{j}\left(\mathbf{I}_{2} \otimes \cdots \mathbf{I}_{2} \otimes \boldsymbol{\sigma}_{j}^{z} \otimes \mathbf{I}_{2} \cdots \otimes \mathbf{I}_{2}\right)\left(\mathbf{e}_{1, s_{1}} \otimes \mathbf{e}_{2, s_{2}} \cdots \otimes \mathbf{e}_{d, s_{d}}\right) \\
& =-\sum_{(i, j) \in \mathcal{E}(\mathcal{G})} J_{i j} s_{i} s_{j} \mathbf{e}_{\mathbf{s}}-\sum_{j \in \mathcal{V}(\mathcal{G})} h_{j} s_{j} \mathbf{e}_{\mathbf{s}} \\
& =\left[-\sum_{(i, j) \in \mathcal{E}(\mathcal{G})} J_{i j} s_{i} s_{j}-\sum_{j \in \mathcal{V}(\mathcal{G})} h_{j} s_{j}\right] \mathbf{e}_{\mathbf{s}}=\mathbf{H}_{I}^{c}(\mathbf{s}) \mathbf{e}_{\mathbf{s}} .
\end{aligned}
$$

Thus, the $2^{d}$ eigenvalues of $\mathbf{H}_{I}^{q}$ are $\mathbf{H}_{I}^{c}(\mathbf{s}), \mathbf{s} \in\{+1,-1\}^{d}$, which are actually the diagonal entries of $\mathbf{H}_{I}^{q}$. If $\mathbf{s}_{0}$ achieves the global minimum of $\mathbf{H}_{I}^{c}(\mathbf{s})$, then $\mathbf{H}_{I}^{c}\left(\mathbf{s}_{0}\right)$ is the smallest eigenvalue of $\mathbf{H}_{I}^{q}$.

Remark 1. The quantum system governed by quantum Hamiltonian $\mathbf{H}_{I}^{q}$ in (3.5) can recover the Boltzmann distribution for observing configurations in the classical Ising model governed by $\mathbf{H}_{I}^{c}$ in 2.1). Indeed, the partition function and

Gibbs state of the quantum Ising model are, respectively, given by

$$
Z=\operatorname{tr}\left[e^{-\beta \mathbf{H}_{I}^{q}}\right] \text { and } \boldsymbol{\rho}=\frac{e^{-\beta \mathbf{H}_{I}^{q}}}{Z}
$$

Since $\mathbf{H}_{I}^{q}$ has eigenvalues $\mathbf{H}_{I}^{c}(\mathbf{s})$ with eigenvectors $\mathbf{e}_{\mathbf{s}}$, it is easy to compute the partition function as follows,

$$
Z=\sum_{\mathbf{s}}\left\langle\mathbf{e}_{\mathbf{s}}\right| e^{-\beta \mathbf{H}_{I}^{q}}\left|\mathbf{e}_{\mathbf{s}}\right\rangle=\sum_{\mathbf{s}} e^{-\beta \mathbf{H}_{I}^{c}(\mathbf{s})}
$$

The probability of observing configuration $\mathbf{s}$ is given by

$$
\left\langle\mathbf{e}_{\mathbf{s}}\right| \boldsymbol{\rho}\left|\mathbf{e}_{\mathbf{s}}\right\rangle=\frac{1}{Z}\left\langle\mathbf{e}_{\mathbf{s}}\right| e^{-\beta \mathbf{H}_{I}^{q}}\left|\mathbf{e}_{\mathbf{s}}\right\rangle=\frac{1}{Z} e^{-\beta \mathbf{H}_{I}^{c}(\mathbf{s})},
$$

which is equal to the classical Boltzmann distribution.

### 6.3 Proof of Theorem 2

Let $u=t / t_{f}$ be the dimensionless time, and set

$$
\begin{equation*}
\mathbf{H}(u) \equiv \mathbf{H}_{Q A}\left(u t_{f}\right)=\mathbf{H}_{Q A}(t) \text { and }|\varphi(u)\rangle \equiv\left|\psi\left(u t_{f}\right)\right\rangle=|\psi(t)\rangle . \tag{6.13}
\end{equation*}
$$

Since state $|\psi(t)\rangle$ of the quantum system in nature time $t$ follows the Schrödingier equation with Hamiltonian $\mathbf{H}_{Q A}(t),|\varphi(u)\rangle$ satisfies

$$
\begin{equation*}
i \frac{d|\varphi(u)\rangle}{d u}=t_{f} \mathbf{H}(u)|\varphi(u)\rangle, \tag{6.14}
\end{equation*}
$$

where we reserve $i$ for the unit imaginary number $\sqrt{-1}$ in the proof of Theorem 2.

Because our goal is to study how close the quantum state is to the ground state, we naturally express the quantum state by the instantaneous eigenstates
of $\mathbf{H}(u)$. We denote by $\lambda_{0}(u)<\lambda_{1}(u)<\cdots<\lambda_{k}(u)<\cdots$ the eigenvalues of $\mathbf{H}(u)$ listed in a strictly increasing order, and $\left|k^{(\varpi)}(u)\right\rangle$ denotes the normalized instantaneous eigenstates of $\mathbf{H}(u)$ corresponding to the $k$-th eigenvalue $\lambda_{k}(u)$, where index $\varpi$ is for eigenvalue multiplicity. Denote by $\left|0^{(\varpi)}(u)\right\rangle, \varpi=1, \cdots, \zeta$, those normalized instantaneous eigenstates so that $\left\{\left|0^{(\varpi)}(1)\right\rangle\right\}_{1 \leq \varpi \leq \zeta}$ are the $\zeta$ ground states of $\mathbf{H}_{I}^{q}$. That is, $u=1$ corresponds to the end of the quantum annealing process, and there are $\zeta$ ground states $\left|0^{(\varpi)}(1)\right\rangle, \varpi=1, \cdots, \zeta$, corresponding to the smallest eigenvalue $\lambda_{0}(1)$. Since our main analysis is targeted for the ground states, we focus on $\left|0^{(\varpi)}(u)\right\rangle$, their amplitudes and relationships with other eigenstates.

Note that we adopt two equivalent notations for the eigen-analysis of $\mathbf{H}(u)$ : $\left\{\xi_{j}(u),\left|v_{j}(u)\right\rangle\right\}$ stated in Theorem 2 and $\left\{\lambda_{k}(u),\left|k^{(\varpi)}(u)\right\rangle\right\}$ defined above, where the former allows repetitions of eigenvalues $\xi_{j}(u)$ with multiplicity greater than 1 , and the latter has strictly increasing eigenvalues $\lambda_{k}(u)$ and the multiplicity index $\varpi$.

The main idea of the long proof arguments is described as follows: (1) characterize the eigenstates $\left|k^{(\varpi)}(u)\right\rangle ;(2)$ obtain an expression for the quantum state $|\psi(u)\rangle$ in terms of these eigenstates; (3) derive the amplitudes in the expression corresponding to $\left|0^{(\varpi)}(u)\right\rangle$; and (4) find the norm of these amplitudes.

First we need to adjust the eigenstates to meet certain conditions in order to facilitate our analysis. Taking derivatives on both sides of $\mathbf{H}(u)\left|j^{(l)}(u)\right\rangle=$ $\lambda_{j}(u)\left|j^{(l)}(u)\right\rangle$, we obtain

$$
\frac{d \mathbf{H}(u)}{d u}\left|j^{(l)}(u)\right\rangle+\mathbf{H}(u) \frac{d\left|j^{(l)}(u)\right\rangle}{d u}=\frac{d \lambda_{j}(u)}{d u}\left|j^{(l)}(u)\right\rangle+\lambda_{j}(u) \frac{d\left|j^{(l)}(u)\right\rangle}{d u},
$$

and thus for $k \neq j$,

$$
\begin{align*}
& \left\langle k^{(\varpi)}(u)\right| \frac{d \mathbf{H}(u)}{d u}\left|j^{(l)}(u)\right\rangle+\left\langle k^{(\varpi)}(u)\right| \mathbf{H}(u) \frac{d\left|j^{(l)}(u)\right\rangle}{d u} \\
& =\left\langle k^{(\varpi)}(u)\right| \frac{d \lambda_{j}(u)}{d u}\left|j^{(l)}(u)\right\rangle+\left\langle k^{(\varpi)}(u)\right| \lambda_{j}(u) \frac{d\left|j^{(l)}(u)\right\rangle}{d u} . \tag{6.15}
\end{align*}
$$

For orthonormal $\left|j^{(l)}(u)\right\rangle$ and $\left|k^{(\varpi)}(u)\right\rangle$, we have $\left\langle j^{(l)}(u)\right\rangle\left|k^{(\varpi)}(u)\right\rangle=0$ and $\left\langle k^{(\varpi)}(u)\right| \mathbf{H}(u)=\lambda_{k}(u)\left\langle k^{(\varpi)}(u)\right|$. Substituting these into 6.15 yields

$$
\begin{aligned}
& \left\langle k^{(\varpi)}(u)\right| \frac{d \mathbf{H}(u)}{d u}\left|j^{(l)}(u)\right\rangle+\lambda_{k}(u)\left\langle k^{(\varpi)}(u)\right| \frac{d\left|j^{(l)}(u)\right\rangle}{d u} \\
& =\frac{d \lambda_{j}(u)}{d u}\left\langle k^{(\varpi)}(u) \mid j^{(l)}(u)\right\rangle+\lambda_{j}(u)\left\langle k^{(\varpi)}(u)\right| \frac{d\left|j^{(l)}(u)\right\rangle}{d u} \\
& =\lambda_{j}(u)\left\langle k^{(\varpi)}(u)\right| \frac{d\left|j^{(l)}(u)\right\rangle}{d u},
\end{aligned}
$$

which immediately leads to

$$
\begin{align*}
& \left\langle k^{(\varpi)}(u)\right| \frac{d}{d u}\left|j^{(l)}(u)\right\rangle \equiv\left\langle k^{(\varpi)}(u)\right| \frac{d\left|j^{(l)}(u)\right\rangle}{d u} \\
& =\frac{1}{\lambda_{j}(u)-\lambda_{k}(u)}\left\langle k^{(\varpi)}(u)\right| \frac{d \mathbf{H}(u)}{d u}\left|j^{(l)}(u)\right\rangle, \quad j \neq k . \tag{6.16}
\end{align*}
$$

Let $\left|\check{j}^{(l)}(u)\right\rangle=\exp \left\{i \eta_{j}^{(l)}(u)\right\}\left|j^{(l)}(u)\right\rangle$ be a time-dependent phase shift of $\left|j^{(l)}(u)\right\rangle$, that is, we add an accent mark ${ }^{\text {n }}$ to mean a time-dependent shift for the eigenstates, where $\eta_{j}^{(l)}(u)$ satisfies

$$
\left\langle j^{(l)}(u)\right| \frac{d}{d u}\left|j^{(l)}(u)\right\rangle+i \frac{d \eta_{j}^{(l)}(u)}{d u}=0,
$$

which is possible since

$$
\left\langle j^{(l)}(u)\right| \frac{d}{d u}\left|j^{(l)}(u)\right\rangle+\left(\left\langle j^{(l)}(u)\right| \frac{d}{d u}\left|j^{(l)}(u)\right\rangle\right)^{\dagger}=\frac{d}{d u}\left\langle j^{(l)}(u) \mid j^{(l)}(u)\right\rangle=0
$$

and hence $\left\langle j^{(l)}(u)\right| \frac{d}{d u}\left|j^{(l)}(u)\right\rangle$ is a pure imaginary number. Thus, we have

$$
\begin{align*}
& \left\langle\check{j}^{(l)}(u)\right| \frac{d}{d u}\left|\check{j}^{(l)}(u)\right\rangle=e^{i \eta_{j}^{(l)}(u)}\left\langle\check{j}^{(l)}(u)\right| \frac{d}{d u}\left|j^{(l)}(u)\right\rangle+i \frac{d \eta_{j}^{(l)}(u)}{d u}\left\langle\check{j}^{(l)}(u) \mid \check{j}^{(l)}(u)\right\rangle \\
& =e^{i \eta_{j}^{(l)}(u)} e^{-i \eta_{j}^{(l)}(u)}\left\langle j^{(l)}(u)\right| \frac{d}{d u}\left|j^{(l)}(u)\right\rangle+i \frac{d \eta_{j}^{(l)}(u)}{d u} \\
& =\left\langle j^{(l)}(u)\right| \frac{d}{d u}\left|j^{(l)}(u)\right\rangle+i \frac{d \eta_{j}^{(l)}(u)}{d u}=0 . \tag{6.17}
\end{align*}
$$

Of course, $\left\{\left|\breve{j}^{(l)}(u)\right\rangle, j=0,1, \cdots,\right\}$ remains to be orthonormal, the pair $\left(\lambda_{j}(u),\left|\check{j}^{(l)}(u)\right\rangle\right)$ still satisfies

$$
\mathbf{H}(u)\left|\check{j}^{(l)}(u)\right\rangle=e^{i \eta_{j}^{(l)}(u)} \mathbf{H}(u)\left|j^{(l)}(u)\right\rangle=e^{i \eta_{j}^{(l)}(u)} \lambda_{j}(u)\left|j^{(l)}(u)\right\rangle=\lambda_{j}(u)\left|\check{j}^{(l)}(u)\right\rangle
$$

and for $k \neq j$,

$$
\begin{align*}
& \left\langle\check{k}^{(\varpi)}(u) \left\lvert\, \frac{d}{d u} \check{j}^{(l)}(u)\right.\right\rangle=e^{i \eta_{j}^{(l)}(u)}\left\langle\check{k}^{(\varpi)}(u)\right| \frac{d}{d u}\left|j^{(l)}(u)\right\rangle+i \frac{d \eta_{j}^{(l)}(u)}{d u}\left\langle\check{k}^{(\varpi)}(u) \mid j^{(l)}(u)\right\rangle \\
& =e^{i\left[\eta_{j}^{(l)}(u)-\eta_{k}^{(\varpi)}(u)\right]}\left\langle k^{(\varpi)}(u)\right| \frac{d}{d u}\left|j^{(l)}(u)\right\rangle \\
& =\frac{e^{i\left[\eta_{j}^{(l)}(u)-\eta_{k}^{(\varpi)}(u)\right]}}{\lambda_{j}(u)-\lambda_{k}(u)}\left\langle k^{(\varpi)}(u)\right| \frac{d \mathbf{H}(u)}{d u}\left|j^{(l)}(u)\right\rangle \\
& =\frac{1}{\lambda_{j}(u)-\lambda_{k}(u)}\left\langle\check{k}^{(\varpi)}(u)\right| \frac{d \mathbf{H}(u)}{d u}\left|\check{j}^{(l)}(u)\right\rangle, \tag{6.18}
\end{align*}
$$

where the third equality is due to (6.16).
Now since (6.17) and 6.18) are satisfied by the instantaneous eigenstates $\check{j}^{(l)}(u)$ of $\mathbf{H}(u)$, we use them to express the quantum state $|\varphi(u)\rangle$ as follows:

$$
\begin{equation*}
|\varphi(u)\rangle=\sum_{j, l \geq 0} \alpha_{j}^{(l)}(u)\left|\check{j}^{(l)}(u)\right\rangle . \tag{6.19}
\end{equation*}
$$

Plugging expression (6.19) into the Schrödinger equation (6.14) we obtain

$$
\sum_{j, l \geq 0} i \frac{d}{d u}\left[\alpha_{j}^{(l)}(u)\left|\check{j}^{(l)}(u)\right\rangle\right]=\sum_{j, l \geq 0} t_{f} \mathbf{H}(u)\left[\alpha_{j}^{(l)}(u)\left|\check{j}^{(l)}(u)\right\rangle\right],
$$

and simple manipulations lead to

$$
\begin{aligned}
& \sum_{j, l \geq 0} i\left[\frac{d \alpha_{j}^{(l)}(u)}{d u}\left|\check{j}^{(l)}(u)\right\rangle+\alpha_{j}^{(l)}(u) \frac{d}{d u}\left|\check{j}^{(l)}(u)\right\rangle\right]=\sum_{j, l \geq 0} t_{f} \alpha_{j}^{(l)}(u) \mathbf{H}(u)\left|\check{j}^{(l)}(u)\right\rangle \\
& =\sum_{j, l \geq 0} t_{f} \alpha_{j}^{(l)}(u) \lambda_{j}(u)\left|\check{j}^{(l)}(u)\right\rangle .
\end{aligned}
$$

Taking products with state $\left\langle\breve{k}^{(\varpi)}(u)\right|$ on both sides of above equation and noting the scalar nature of $t_{f}, \alpha_{j}^{(l)}(u)$ and $\lambda_{j}(u)$, we arrive at

$$
\begin{aligned}
& \sum_{j, l \geq 0} i\left[\frac{d \alpha_{j}^{(l)}(u)}{d u}\left\langle\check{k}^{(\varpi)}(u) \mid \check{j}^{(l)}(u)\right\rangle+\alpha_{j}^{(l)}(u)\left\langle\check{k}^{(\varpi)}(u)\right| \frac{d}{d u}\left|\check{j}^{(l)}(u)\right\rangle\right] \\
& =\sum_{j, l \geq 0} t_{f} \lambda_{j}(u) \alpha_{j}^{(l)}(u)\left\langle\check{k}^{(\varpi)}(u) \mid \check{j}^{(l)}(u)\right\rangle,
\end{aligned}
$$

which can be simplified by using 6.17 and the orthonormality of $\left|\check{j}^{(l)}(u)\right\rangle$ as

$$
\begin{align*}
& \frac{d \alpha_{k}^{(\varpi)}(u)}{d u}+\sum_{l \neq \varpi} \alpha_{k}^{(l)}\left\langle\check{k}^{(\varpi)}(u)\right| \frac{d}{d u}\left|\check{k}^{(l)}(u)\right\rangle+\sum_{j \neq k} \sum_{l} \alpha_{j}^{(l)}(u)\left\langle\check{k}^{(\varpi)}(u) \left\lvert\, \frac{d}{d u} \check{j}^{(l)}(u)\right.\right\rangle \\
& =-i t_{f} \lambda_{k}(u) \alpha_{k}^{(\varpi)}(u) . \tag{6.20}
\end{align*}
$$

### 6.3.1 Establishing the probability lower bound (3.6)

Recall that we list $2^{d}$ eigenvalues $\xi_{1}(u) \leq \xi_{2}(u) \leq \cdots \leq \xi_{2^{d}}(u)$ of $\mathbf{H}(u)$ in an increasing order along with the corresponding normalized eigenvectors $v_{1}(u), v_{2}(u), \cdots, v_{2^{d}}(u)$, where for any eigenvalue with a multiplicity greater than 1 , the eigenvalue is repeated in the list with the number of repetitions equal to its multiplicity, and the multiple eigenvectors corresponding to the same eigenvalue are ordered as a group.

We use the ordered $2^{d}$ normalized eigenvectors $v_{1}(u), v_{2}(u), \cdots, v_{2^{d}}(u)$ to represent the quantum state $|\psi(u)\rangle$ defined in 6.13). Vector-matrix forms are
adopted to facilitate our analysis below. Denote by $\boldsymbol{\Lambda}(u)$ the diagonal matrix whose diagonal entries are the ordered $2^{d}$ eigenvalues $\xi_{1}(u), \xi_{2}(u), \ldots, \xi_{2^{d}}(u)$. Denote by $\mathbf{V}(u)$ the matrix whose columns are formed by the ordered $2^{d}$ normalized eigenvectors $v_{1}(u), v_{2}(u), \ldots, v_{2^{d}}(u)$, and denote by $\boldsymbol{\alpha}(u)$ the coefficient vector in the representation of $|\psi(u)\rangle$ by the ordered $2^{d}$ normalized eigenvectors. Then from the equations (6.19) and 6.20), we have

$$
\begin{equation*}
|\varphi(u)\rangle=\mathbf{V}(u) \boldsymbol{\alpha}(u), \tag{6.21}
\end{equation*}
$$

and

$$
\begin{equation*}
\frac{d}{d u} \boldsymbol{\alpha}(u)=-i t_{f} \boldsymbol{\Lambda}(u) \boldsymbol{\alpha}(u)-\left(\mathbf{V}^{\dagger}(u) \frac{d}{d u} \mathbf{V}(u)\right) \boldsymbol{\alpha}(u) . \tag{6.22}
\end{equation*}
$$

Since the probability lower bound (3.6) depends on the first section of $\boldsymbol{\alpha}(u)$, we need to break matrix $\mathbf{V}(u)$ and vector $\boldsymbol{\alpha}(u)$ into two parts. Take an integer $r$ between 1 and $2^{d}$, denote by $\boldsymbol{\alpha}_{r}(u)$ and $\boldsymbol{\alpha}_{r+}(u)$ the vectors formed by the first $r$ components of $\boldsymbol{\alpha}(u)$ and the remaining $2^{d}-r$ components, respectively. Similarly, denote by $\mathbf{V}_{r}(u)$ and $\mathbf{V}_{r+}(u)$ the matrices formed by the first $r$ columns of $\mathbf{V}(u)$ and the $2^{d}-r$ columns, respectively. Denote by $\boldsymbol{\Lambda}_{r}(u)$ the diagonal matrix formed by retaining the first $r$ rows and columns of $\boldsymbol{\Lambda}(u)$. Then equation 6.22) immediately leads to

$$
\begin{align*}
& \frac{d}{d u} \boldsymbol{\alpha}_{r}(u)=-i t_{f} \boldsymbol{\Lambda}_{r}(u) \boldsymbol{\alpha}_{r}(u)-\left(\mathbf{V}_{r}^{\dagger}(u) \frac{d}{d u} \mathbf{V}(u)\right) \boldsymbol{\alpha}(u) \\
& =-i t_{f} \boldsymbol{\Lambda}_{r}(u) \boldsymbol{\alpha}_{r}(u)-\left(\mathbf{V}_{r}^{\dagger}(u) \frac{d}{d u} \mathbf{V}_{r}(u)\right) \boldsymbol{\alpha}_{r}(u) \\
& \quad-\left(\mathbf{V}_{r}^{\dagger}(u) \frac{d}{d u} \mathbf{V}_{r+}(u)\right) \boldsymbol{\alpha}_{r+}(u) \\
& =\mathbf{D}_{r}(u) \boldsymbol{\alpha}_{r}(u)-\left(\mathbf{V}_{r}^{\dagger}(u) \frac{d}{d u} \mathbf{V}_{r+}(u)\right) \boldsymbol{\alpha}_{r+}(u), \tag{6.23}
\end{align*}
$$

where $\mathbf{D}_{r}(u)=-i t_{f} \boldsymbol{\Lambda}_{r}(u)-\mathbf{V}_{r}^{\dagger}(u) \frac{d}{d u} \mathbf{V}_{r}(u)$. The solution of the linear differential equation system 6.23 has an expression

$$
\begin{equation*}
\boldsymbol{\alpha}_{r}(u)=\mathbf{U}_{r}(u) \boldsymbol{\alpha}_{r}(0)-\mathbf{U}_{r}(u) \int_{0}^{u}\left[\mathbf{U}_{r}(x)\right]^{-1}\left(\mathbf{V}_{r}^{\dagger}(x) \frac{d}{d x} \mathbf{V}_{r+}(x)\right) \boldsymbol{\alpha}_{r+}(x) d x, \tag{6.24}
\end{equation*}
$$

where $\mathbf{U}_{r}(u)$ is a fundamental matrix for the homogeneous linear differential equation system (6.23) with initial condition $\mathbf{U}_{r}(0)=\mathbf{I}-$ namely, the columns of $\mathbf{U}_{r}(u)$ form a complete linearly independent set of solutions for the homogeneous linear differential equation system,

$$
\frac{d}{d u} \boldsymbol{\alpha}_{r}(u)=\mathbf{D}_{r}(u) \boldsymbol{\alpha}_{r}(u) .
$$

The fundamental matrix $\mathbf{U}_{r}(u)$ has an expression through the so-called Magnus expansion (Blanes et al. (2009)),

$$
\mathbf{U}_{r}(u)=\exp \left\{\boldsymbol{\Xi}^{(r)}(u)\right\}, \quad \boldsymbol{\Xi}^{(r)}(u)=\sum_{k=1}^{\infty} \boldsymbol{\Xi}_{k}^{(r)}(u)
$$

where $\boldsymbol{\Xi}_{k}^{(r)}(u)$ in the Magnus expansion can be computed by a recursive procedure through the matrices $\boldsymbol{\Upsilon}_{k}^{(j)}(u)$ as follows,

$$
\begin{equation*}
\mathbf{\Xi}_{1}^{(r)}(u)=\int_{0}^{u} \mathbf{D}_{r}(v) d v, \quad \mathbf{\Xi}_{k}^{(r)}(u)=\sum_{j=1}^{k-1} \frac{\nu_{j}}{j!} \int_{0}^{u} \boldsymbol{\Upsilon}_{k}^{(j)}(v) d v, \quad k \geq 2, \tag{6.25}
\end{equation*}
$$

$\nu_{j}$ are Bernoulli numbers,

$$
\begin{equation*}
\mathbf{\Upsilon}_{k}^{(1)}(u)=\left[\boldsymbol{\Xi}_{k-1}^{(r)}(u), \mathbf{D}_{r}(u)\right], \quad \mathbf{\Upsilon}_{k}^{(j)}(u)=\sum_{l=1}^{k-j}\left[\mathbf{\Xi}_{l}^{(r)}(u), \mathbf{\Upsilon}_{k-l}^{(j-1)}(u)\right], \quad j=2, \cdots, k-1, \tag{6.26}
\end{equation*}
$$

and $\left[\mathbf{A}_{1}, \mathbf{A}_{2}\right]=\mathbf{A}_{1} \mathbf{A}_{2}-\mathbf{A}_{2} \mathbf{A}_{1}$ is the matrix commutator of $\mathbf{A}_{1}$ and $\mathbf{A}_{2}$.
We therefore have the following lemma for $\mathbf{U}_{r}(u)$.

Lemma 1. The fundamental matrix $\mathbf{U}_{r}(u)$ is unitary for integer $r$ and $u \in[0,1]$.

Proof. A square matrix $\mathbf{A}$ is said to be skew-Hermitian if and only if $\mathbf{A}^{\dagger}=-\mathbf{A}$. Denote by $\mathbf{S H}(r)$ the set formed by all skew-Hermitian matrice of size $r$. Then $\mathbf{S H}(r)$ is closed under any linear transformation with real coefficients, integration, and matrix commutator. Indeed, it is evident that

1. $\mathbf{B}_{1}, \mathbf{B}_{2} \in \mathbf{S H}(r), \alpha, \beta \in \mathbb{R} \Rightarrow \alpha \mathbf{B}_{1}+\beta \mathbf{B}_{2} \in \mathbf{S H}(r)$,
2. $\mathbf{B}(u) \in \mathbf{S H}(r), u \in[a, b] \Rightarrow \int_{a}^{b} \mathbf{B}(u) d u \in \mathbf{S H}(r)$,
3. $\mathbf{B}_{1}, \mathbf{B}_{2} \in \mathbf{S H}(r) \Rightarrow\left[\mathbf{B}_{1}, \mathbf{B}_{2}\right] \in \mathbf{S H}(r)$.

Now we will show $\mathbf{D}_{r}(u) \in \mathbf{S H}(r)$, since

$$
\begin{align*}
\mathbf{D}_{r}(u)^{\dagger} & =\left(-i t_{f} \boldsymbol{\Lambda}_{r}(u)\right)^{\dagger}-\left(\mathbf{V}_{r}^{\dagger}(u) \frac{d}{d u} \mathbf{V}_{r}(u)\right)^{\dagger} \\
& =i t_{f} \boldsymbol{\Lambda}_{r}(u)-\left(\left(\frac{d}{d u} \mathbf{V}_{r}^{\dagger}(u)\right) \mathbf{V}_{r}(u)\right) \\
& =i t_{f} \boldsymbol{\Lambda}_{r}(u)-\left(\frac{d}{d u}\left(\mathbf{V}_{r}^{\dagger}(u) \mathbf{V}_{r}(u)\right)-\mathbf{V}_{r}^{\dagger}(u) \frac{d}{d u} \mathbf{V}_{r}(u)\right) \\
& =i t_{f} \boldsymbol{\Lambda}_{r}(u)+\mathbf{V}_{r}^{\dagger}(u) \frac{d}{d u} \mathbf{V}_{r}(u)=-\mathbf{D}_{r}(u) \tag{6.27}
\end{align*}
$$

Using 6.25 and 6.26, we conclude that $\mathbf{\Upsilon}_{k}^{(j)}(u) \in \mathbf{S H}(r)$ for $k>1$ and $j=$ $1,2, \ldots, k-1, \boldsymbol{\Xi}_{k}^{(r)}(u) \in \mathbf{S H}(r)$ for $k \geq 1$, and thus $\boldsymbol{\Xi}^{(r)}(u) \in \mathbf{S H}(r)$. Then we have

$$
\begin{aligned}
& \mathbf{U}_{r}^{\dagger}(u) \mathbf{U}_{r}(u)=\exp \left\{\left[\boldsymbol{\Xi}^{(r)}(u)\right]^{\dagger}\right\} \exp \left\{\boldsymbol{\Xi}^{(r)}(u)\right\} \\
& =\exp \left\{-\boldsymbol{\Xi}^{(r)}(u)\right\} \exp \left\{\boldsymbol{\Xi}^{(r)}(u)\right\}=\mathbf{I}_{r},
\end{aligned}
$$

which proves the lemma.

Now we show the probability lower bound (3.6). The quantum system initializes at the ground state of $\mathbf{H}(0)$-namely, $|\psi(0)\rangle$ is equal to the ground state. Thus, we have 1 for the first component of $\boldsymbol{\alpha}(0)$ and 0 for the rest of its components, which implies $\left\|\boldsymbol{\alpha}_{r}(0)\right\|=1$ for any $r \geq 1$. Using (6.24) we obtain

$$
\begin{align*}
& \left\|\boldsymbol{\alpha}_{r}(1)\right\| \geq\left\|\mathbf{U}_{r}(1) \boldsymbol{\alpha}_{r}(0)\right\| \\
& \quad \quad-\left\|\mathbf{U}_{r}(1) \int_{0}^{1}\left[\mathbf{U}_{r}(u)\right]^{-1}\left(\mathbf{V}_{r}^{\dagger}(u) \frac{d}{d u} \mathbf{V}_{r+}(u)\right) \boldsymbol{\alpha}_{r+}(u) d u\right\| \\
& =1-\left\|\int_{0}^{1}\left[\mathbf{U}_{r}(u)\right]^{-1}\left(\mathbf{V}_{r}^{\dagger}(u) \frac{d}{d u} \mathbf{V}_{r+}(u)\right) \boldsymbol{\alpha}_{r+}(u) d u\right\| \\
& \geq 1-\int_{0}^{1}\left\|\left(\mathbf{V}_{r}^{\dagger}(u) \frac{d}{d u} \mathbf{V}_{r+}(u)\right) \boldsymbol{\alpha}_{r+}(u)\right\| d u . \tag{6.28}
\end{align*}
$$

Note that $\left\|\boldsymbol{\alpha}_{r+}(u)\right\| \leq\|\boldsymbol{\alpha}(u)\|=1,\left\|\mathbf{V}_{r+}^{\dagger}(u)\right\|=1, \mathbf{V}_{r}^{\dagger}(u) \mathbf{V}_{r+}(u)=0$, and

$$
\begin{aligned}
& \mathbf{V}_{r}^{\dagger}(u) \frac{d}{d u} \mathbf{V}_{r+}(u)=\frac{d}{d u}\left(\mathbf{V}_{r}^{\dagger}(u) \mathbf{V}_{r+}(u)\right)-\left(\frac{d}{d u} \mathbf{V}_{r}^{\dagger}(u)\right) \mathbf{V}_{r+}(u) \\
& =\quad-\left(\mathbf{V}_{r+}^{\dagger}(u) \frac{d}{d u} \mathbf{V}_{r}(u)\right)^{\dagger}
\end{aligned}
$$

Then

$$
\begin{aligned}
& \left\|\left(\mathbf{V}_{r}^{\dagger}(u) \frac{d}{d u} \mathbf{V}_{r+}(u)\right) \boldsymbol{\alpha}_{r+}(u)\right\| \leq\left\|\mathbf{V}_{r}^{\dagger}(u) \frac{d}{d u} \mathbf{V}_{r+}(u)\right\|\left\|\boldsymbol{\alpha}_{r+}(u)\right\| \\
& =\left\|\mathbf{V}_{r+}^{\dagger}(u) \frac{d}{d u} \mathbf{V}_{r}(u)\right\| \\
& \leq\left\|\mathbf{V}_{r+}^{\dagger}(u)\right\|\left\|\frac{d}{d u} \mathbf{V}_{r}(u)\right\| \\
& =\left\|\frac{d}{d u} \mathbf{V}_{r}(u)\right\| .
\end{aligned}
$$

Plugging the above inequality into $(6.28)$ we arrive at

$$
\left\|\boldsymbol{\alpha}_{r}(1)\right\| \geq 1-\mathcal{K}_{r},
$$

where

$$
\mathcal{K}_{r}=\int_{0}^{1}\left\|\frac{d}{d u} \mathbf{V}_{r}(u)\right\| d u
$$

Finally, the probability for the quantum annealing process to stay in a ground state at the final annealing time is equal to $\left\|\boldsymbol{\alpha}_{\zeta}(1)\right\|^{2}$, and for any $1 \leq r \leq \zeta$,

$$
\left\|\boldsymbol{\alpha}_{\zeta}(1)\right\|^{2} \geq\left\|\boldsymbol{\alpha}_{r}(1)\right\|^{2} \geq\left[\left(1-\mathcal{K}_{r}\right)_{+}\right]^{2} .
$$

Thus, (3.6) is established.

### 6.3.2 Establishing the probability lower bound (3.7)

We utilize 6.18) to re-write 6.20 with $k=0$ as the following linear differential equation system for the amplitudes $\alpha_{0}^{(\varpi)}(u)$ of $\zeta$ ground states:

$$
\begin{align*}
& \frac{d \alpha_{0}^{(\varpi)}(u)}{d u}=-i t_{f} \lambda_{0}(u) \alpha_{0}^{(\varpi)}(u)-\sum_{l \neq \varpi} \alpha_{0}^{(l)}(u)\left\langle\check{0}^{(\varpi)}(u) \left\lvert\, \frac{d}{d u} \check{0}^{(l)}(u)\right.\right\rangle \\
& -\sum_{j>0} \sum_{l} \alpha_{j}^{(l)}(u) \frac{1}{\lambda_{j}(u)-\lambda_{0}(u)}\left\langle\check{0}^{(\varpi)}(u)\right| \frac{d \mathbf{H}(u)}{d u}\left|\check{j}^{(l)}(u)\right\rangle, \tag{6.29}
\end{align*}
$$

where $\varpi=1, \cdots, \zeta$, the sum in the second term is taken over $l=1, \cdots, \varpi-$ $1, \varpi+1, \cdots, \zeta$ for ground states, and the sums in the third term are over for all excited states.

The linear differential equation system (6.29) has the solution

$$
\begin{array}{r}
\left(\alpha_{0}^{(1)}(u), \cdots, \alpha_{0}^{(\zeta)}(u)\right)^{\prime}=\mathbf{U}(u)\left(\alpha_{0}^{(1)}(0), \cdots, \alpha_{0}^{(\zeta)}(0)\right)^{\prime}+ \\
\mathbf{U}(u) \int_{0}^{u}[\mathbf{U}(x)]^{-1} \sum_{j>0} \alpha_{j}^{(l)}(x) \frac{1}{\lambda_{j}(x)-\lambda_{0}(x)}\langle\check{\mathbf{0}}(x)| \frac{d \mathbf{H}(x)}{d x}\left|\check{j}^{(l)}(x)\right\rangle d x, \tag{6.30}
\end{array}
$$

where $\langle\check{\mathbf{0}}(x)|=\left(\left\langle\check{0}^{(1)}(x)\right|, \cdots,\left\langle\check{0}^{(\zeta)}(x)\right|\right)^{\prime}$, and $\mathbf{U}$ is a fundamental matrix for the homogeneous linear differential equation system corresponding to (6.29) with
initial condition $\mathbf{U}(0)=\mathbf{I}$. More specifically, the columns of $\mathbf{U}$ form a complete, linearly independent set of solutions for the homogeneous equation system,

$$
\begin{equation*}
\frac{d \alpha_{0}^{(\varpi)}(u)}{d u}=-i t_{f} \lambda_{0}(u) \alpha_{0}^{(\varpi)}(u)-\sum_{l \neq \varpi}\left\langle\check{0}^{(\varpi)}(u)\right| \frac{d}{d u}\left|\check{0}^{(l)}(u)\right\rangle \alpha_{0}^{(l)}(u) . \tag{6.31}
\end{equation*}
$$

Or, in a vector-matrix form,

$$
\begin{equation*}
\frac{d\left(\alpha_{0}^{(1)}(u), \cdots, \alpha_{0}^{(\zeta)}(u)\right)^{\prime}}{d u}=\mathbf{D}(u)\left(\alpha_{0}^{(1)}(0), \cdots, \alpha_{0}^{(\zeta)}(0)\right)^{\prime} \tag{6.32}
\end{equation*}
$$

where $\mathbf{D}(u)=-i t_{f} \lambda_{0}(u) \mathbf{I}-i \boldsymbol{\Psi}(u)$ is a matrix of size $\zeta, \boldsymbol{\Psi}(u)=\left(\Psi_{\varpi l}(u)\right)$ and $\Psi_{\varpi l}(u)=-i\left\langle\check{0}^{(\varpi)}(u)\right| \frac{d}{d u}\left|\check{0}^{(l)}(u)\right\rangle$ for $l \neq \varpi$ and 0 for $l=\varpi$. Since

$$
\begin{aligned}
& i \Psi_{\varpi l}(u)+\left[i \Psi_{l \varpi}(u)\right]^{*}=\left\langle 0^{(\varpi)}(u)\right| \frac{d}{d u}\left|0^{(l)}(u)\right\rangle+\left(\left\langle 0^{(l)}(u)\right| \frac{d}{d u}\left|0^{(\varpi)}(u)\right\rangle\right)^{*} \\
& =\frac{d}{d u}\left\langle 0^{(\varpi)}(u) \mid 0^{(l)}(u)\right\rangle=0
\end{aligned}
$$

$\boldsymbol{\Psi}(u)$ is a real matrix. Equation (6.32) and the fundamental matrix $\mathbf{U}(u)$ correspond to equation (6.23) and fundamental matrix $\mathbf{U}_{r}(u)$ with $r=\zeta$ in Section 6.3.1 respectively. Hence, by Lemma 1 we have that $\mathbf{U}(u)$ is unitary, and then

$$
\begin{equation*}
\|\mathbf{U}(u)\|=1 \tag{6.33}
\end{equation*}
$$

As the system initializes at the ground state of $\mathbf{H}(0),\left\|\left(\alpha_{0}^{(1)}(0), \cdots, \alpha_{0}^{(\zeta)}(0)\right)\right\|^{2}$ $=1$. Finally, from (6.30) and 6.33) we find

$$
\begin{align*}
& \left\|\left(\alpha_{0}^{(1)}(1), \cdots, \alpha_{0}^{(\zeta)}(1)\right)\right\|^{2} \geq\left\|\mathbf{U}(1)\left(\alpha_{0}^{(1)}(0), \cdots, \alpha_{0}^{(\zeta)}(0)\right)\right\|^{2}- \\
& \left.\left|\mathbf{U}(1) \int_{0}^{1}[\mathbf{U}(x)]^{-1} \sum_{j>0} \sum_{l} \frac{\alpha_{j}^{(l)}(x)}{\lambda_{j}(x)-\lambda_{0}(x)}\langle\check{\mathbf{0}}(x)| \frac{d \mathbf{H}(x)}{d x}\right| \check{j}^{(l)}(x)\right\rangle\left. d x\right|^{2}  \tag{6.34}\\
& \left\|\mathbf{U}(1)\left(\alpha_{0}^{(1)}(0), \cdots, \alpha_{0}^{(\zeta)}(0)\right)\right\|^{2}=\left\|\left(\alpha_{0}^{(1)}(0), \cdots, \alpha_{0}^{(\zeta)}(0)\right)\right\|^{2}=1 \tag{6.35}
\end{align*}
$$

and

$$
\begin{align*}
& \left\|\mathbf{U}(u) \int_{0}^{u}[\mathbf{U}(x)]^{-1} \sum_{j>0} \sum_{l} \frac{\alpha_{j}^{(l)}(x)}{\lambda_{j}(x)-\lambda_{0}(x)}\langle\check{\mathbf{0}}(x)| \frac{d \mathbf{H}(x)}{d x}\left|\check{j}^{(l)}(x)\right\rangle d x\right\|^{2} \\
& =\left\|\mathbf{U}(u)\left[\mathbf{U}\left(x_{*}\right)\right]^{-1} \sum_{j>0} \sum_{l} \frac{\alpha_{j}^{(l)}\left(x_{*}\right)}{\lambda_{j}\left(x_{*}\right)-\lambda_{0}\left(x_{*}\right)}\left\langle\check{\mathbf{0}}\left(x_{*}\right)\right| \frac{d \mathbf{H}\left(x_{*}\right)}{d x}\left|\check{j}^{(l)}\left(x_{*}\right)\right\rangle\right\|^{2} \\
& \left.\leq \sum_{\varpi=1}^{\zeta}\left|\sum_{j>0} \sum_{l} \alpha_{j}^{(l)}\left(x_{*}\right) \frac{1}{\lambda_{j}\left(x_{*}\right)-\lambda_{0}\left(x_{*}\right)}\left\langle\check{0}^{(\varpi)}\left(x_{*}\right)\right| \frac{d \mathbf{H}\left(x_{*}\right)}{d x}\right| \check{j}^{(l)}\left(x_{*}\right)\right\rangle\left.\right|^{2} \\
& \left.\leq \sum_{\varpi=1}^{\zeta}\left|\sum_{j>0} \sum_{l} \alpha_{j}^{(l)}\left(x_{*}\right) \frac{1}{\lambda_{j}\left(x_{*}\right)-\lambda_{0}\left(x_{*}\right)}\left\langle\check{0}^{(\varpi)}\left(x_{*}\right)\right| \frac{d \mathbf{H}\left(x_{*}\right)}{d x}\right| \check{j}^{(l)}\left(x_{*}\right)\right\rangle\left.\right|^{2} \\
& \left.\leq \sum_{\varpi=1}^{\zeta}\left|\sum_{j>0} \sum_{l} \alpha_{j}^{(l)}\left(x_{*}\right) \frac{1}{\lambda_{j}\left(x_{*}\right)-\lambda_{0}\left(x_{*}\right)}\left\langle\check{0}^{(\varpi)}\left(x_{*}\right)\right| \frac{d \mathbf{H}\left(x_{*}\right)}{d x}\right| \check{j}^{(l)}\left(x_{*}\right)\right\rangle\left.\right|^{2} \\
& \leq \aleph_{0} \zeta\left[\sum_{j>0} \sum_{l}\left|\alpha_{j}^{(l)}\left(x_{*}\right)\right|\right]^{2} \\
& \leq \aleph_{0} \zeta 2^{d} \sum_{j>0} \sum_{l}\left|\alpha_{j}^{(l)}\left(x_{*}\right)\right|^{2} \\
& \leq \aleph_{0} \zeta 2^{d}\left(1-p_{0}\right) \leq \aleph_{0} \zeta 2^{d}, \tag{6.36}
\end{align*}
$$

where in the above eight arrays, the second line is from the mean value theorem and $x_{*} \in(0,1)$, the third and fifth lines are due to the spectral norm and (6.33), respectively, $\aleph_{0}$ is defined to be equal to

$$
\left.\left.\max \left\{\left|\frac{1}{\lambda_{j}(u)-\lambda_{0}(u)}\left\langle\check{0}^{(\varpi)}(u)\right| \frac{d \mathbf{H}(u)}{d u}\right| \check{j}^{(l)}(u)\right\rangle\right|^{2}, j, l \geq 1, \varpi \leq \zeta, u \in[0,1]\right\},
$$

and the last two lines are, respectively, from the Cauchy-Schwartz inequality and the facts that

$$
p_{0}=\min _{0 \leq x \leq 1} \sum_{\varpi=1}^{\zeta}\left|\alpha_{0}^{(\varpi)}(x)\right|^{2} \text { and } \sum_{j, l}\left|\alpha_{j}^{(l)}(x)\right|^{2}=1 .
$$

Applying the Cauchy-Schwartz inequality and using unit norms of $\left|0^{(\varpi)}(u)\right\rangle$ and $\left|\check{j}^{(l)}(u)\right\rangle$ and the spectral norm definition, we obtain

$$
\begin{equation*}
\aleph_{0} \leq \max _{u \in[0,1]}\left\{\left[\lambda_{1}(u)-\lambda_{0}(u)\right]^{-2}\left\|\frac{d \mathbf{H}(u))}{d u}\right\|^{2}\right\}=\aleph \tag{6.37}
\end{equation*}
$$

where $\aleph$ is defined in (3.9).
Combining (6.34)-(6.37), we conclude

$$
\left\|\left(\alpha_{0}^{(1)}(1), \cdots, \alpha_{0}^{(\zeta)}(1)\right)\right\|^{2} \geq 1-2^{d} \zeta \aleph .
$$

The probability of the quantum annealing process staying in the ground states at the final annealing time is equal to $\left\|\left(\alpha_{0}^{(1)}(1), \cdots, \alpha_{0}^{(\zeta)}(1)\right)\right\|^{2}$, and thus we establish the probability lower bound (3.7).

