Quantum Computing Club at
Birzeit University

Paper Review
Mathematical foundation of quantum annealing

Khaled Hasan, Baha'a Asaad

January 2024

1 Introduction

Optimization problems play a crucial role in various scientific and practical domains, where the goal is to minimize or maximize a function of multiple variables known as the cost function. Combinatorial optimization, deals with discrete values for its variables such problems include famous examples such as Max Cut and the Traveling Salesman Problem.
Many of these problems are NP hard problems, where finding the exact solution will require exponential number steps and therefore they are considered impossible to solve exactly. However, it is still possible to find algorithms that provides approximate (yet accurate) solutions in smaller time complexity (polynomial time). One of these algorithms is the simulated annealing (SA), which is inspired by statistical physics, by gradually reducing the temperature of a system to drive it towards the state with the lowest energy, corresponding to the solution of the optimization problem. SA is popular for its practical applicability and descent performance, SA requires finite computation time to provide approximate solutions due to the infinite time needed to reach the exact solution.
This paper considers a quantum alternative to the classical (SA) known as Quantum annealing (QA), (QA) has a outperforms (SA) in terms of time complexity, by making use of some quantum phenomena such as quantum tunneling. As well as providing lower errors than in (SA) for fixed amount of time. However, a notable drawback of QA is its dependence on quantum computers for practical implementation, as solving the time-dependent Schrodinger equation on a large scale.

2 Results

2.1 Convergence condition of QA (Section 2):

The building block of quantum mechanics is the Schrodinger equation:

$$i\frac{d}{ds}\left|\widetilde{\psi}(s)\right\rangle = \tau\widetilde{H}(s)\left|\widetilde{\psi}(s)\right\rangle \tag{1}$$

Written in terms of the dimensionless time $s = t/\tau$, and $\widetilde{H}(s)$ is called the Hamiltonian operator, and it defines the system at hand. The energies of the system are given by the eigenvalues of the Hamiltonian operator.

$$\widetilde{H}(s)\left|k(s)\right\rangle = \varepsilon_k(s)\left|k(s)\right\rangle \tag{2}$$

Where $|k(s)\rangle$ and $\varepsilon_k(s)$ are the k^{th} instantaneous eigenstate and eigenvalue of the system (state with the k^{th} highest energy). we denote the ground state (desired solution) by $|0(s)\rangle$.

By starting with a system of Hamiltonian $\widetilde{H}(0)$ and assuming that we know the ground state of such Hamiltonian and start from it:

$$|\psi(0)\rangle = |0(0)\rangle \tag{3}$$

The amplitude of the non-ground state $|j \neq 0\rangle$ given by $c_{j\neq 0}$ as:

$$c_{j\neq 0}(s) \approx \frac{i}{\tau}\left(A_j(0) - e^{i\tau[\phi_j(s)-\phi_0(s)]}A_j(s)\right) + O(\tau^{-2}) \tag{4}$$

Then, by changing the Hamiltonian to a final $\widetilde{H}(s)$, to make sure we remain in the instantaneous ground state of the system, it was proven that adiabaticity condition should satisfy:

$$A_j(t) = \frac{1}{\left(\varepsilon_j(t) - \varepsilon_0(t)\right)^2}\left\langle j(t)\right|\frac{dH(t)}{dt}\left|0(t)\right\rangle \ll 1. \tag{5}$$

Most Combinatorial optimization problems can be written in terms of Ising Hamiltonian by mapping binary variables to spin variables, therefore it is helpful to look into the Ising model Hamiltonian:

$$H_{\text{Ising}} = -\sum_{i=1}^{N} J_i\sigma_z^i - \sum_{ij} J_{ij}\sigma_z^i\sigma_z^j - \sum_{ijk} J_{ijk}\sigma_z^i\sigma_z^j\sigma_z^k - ... \tag{6}$$

At the beginning we consider another Hamiltonian whose eigenstates are well known, and which acts like an additional kinetic energy term:

$$H_{\text{TF}}(t) \equiv -\Gamma(t)\sum_{i=1}^{N}\sigma_x^i \tag{7}$$

Where $\Gamma(t)$ plays the same role as temperature in the (SA) Algorithm.
It can be shown that satisfying the adiabaticity condition for this problem yields the following time dependence:

$$\Gamma(t) = a(\delta t + c)^{-\frac{1}{2N-1}} \tag{8}$$

And by considering a more general transverse Hamiltonian (Many body transverse interaction Hamiltonian):

$$H_{\mathrm{MTI}}(t) = -\Gamma_{\mathrm{MTI}}(t)\Pi_{i=1}^{N}(1+\sigma_x^i) \tag{9}$$

give rise to the following dependence:

$$\Gamma_{\mathrm{MTI}}(t) = a\frac{2^{N-2}}{\delta t} \tag{10}$$

where $\delta \ll 1$

2.2 QA outperforms SA (Section 2.3)

The convergence of the QA scheme is demonstrated by reducing the factor $\Gamma(t)$ to small values ϵ, at which point, the Hamiltonian will approach the desired cost function, we use this time to investigate the complexity of the algorithm.
For the QA with H_{MTI} the time is given by equation 10 to be:

$$t \propto \frac{2^N}{\delta\epsilon t} \tag{11}$$

while the time dependence of temperature in the SA scheme is give as:

$$T(t) = \frac{pN}{\alpha t + 1} \tag{12}$$

which is clearly slower than Γ as given in equations 8 and 10, despite both needing exponential time to converge.

2.3 Errors of finite time evolution (Section 4)

It was established that the exact solution for the optimization problems defined by equation 6 needs infinite time, and for finite times QA provides an approximation with certain error, this error can be represented as the probability of the excited states, which can be obtained by the norm squared of equation 4 which gives:

$$|c_{j\neq 0}(1)|^2 \leq \frac{1}{\tau^2}\left(|A_j(0)| + |A_j(1)|\right)^2 + O(\tau^{-3}) \tag{13}$$

This error can be further reduced by applying the different annealing schedules, that imposes the following condition:

$$\frac{d^{k-1}}{ds^{k-1}}\widetilde{H}(s)\bigg|_{s=\{0,1\}} = 0 \tag{14}$$

Under which the error is given by:

$$|c_{j\neq 0}(1)|^2 \leq \frac{1}{\tau^{2k}}\left(|A_j^{(k)}(0)| + |A_j^{(k)}(1)|\right)^2 + O(\tau^{-2k-1}) \tag{15}$$

Where $A_j^{(k)}(s)$ is given by:

$$A_j^{(k)}(s) \equiv \frac{1}{\Delta_j(s)^{k+1}}\langle j(s)|\frac{d^k\widetilde{H}(s)}{ds^k}|0(s)\rangle \tag{16}$$

A generalization for time-dependent Hamiltonian with finite annealing time is composed of the potential term and the kinetic energy term,

$$\widetilde{H}(s) = f(s)H_{\mathrm{pot}} + [1-f(s)]H_{\mathrm{kin}} \tag{17}$$

where H_{pot} and H_{kin} generalize H_{Ising} and H_{TF}, respectively. And $f(s)$ is such that $f(0) = 0$ and $f(1) = 0$.

2.3.1 An example: Database search

Assuming we search for an element defined by the state $|m\rangle$.

consider a Hamiltonian:

$$\widetilde{H}_{\text{pot}}(s) = \mathbb{1} - |m\rangle\langle m| \tag{18}$$

while:

$$H_{\text{kin}} = \mathbb{1} - \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}|i\rangle\langle j| \tag{19}$$

the initial instantaneous eigenstate:

$$|\psi(0)\rangle = |0(0)\rangle = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}|i\rangle \tag{20}$$

Applying Adiabaticity condition defined in equation 5 provide the following differential equation:

$$\frac{\sqrt{N-1}}{\tau N\left(\sqrt{\left(1-\frac{4}{N-1}\frac{Nf(s)(1-f(s))}{N}\right)}\right)^3}\frac{df}{ds} = \delta \tag{21}$$

The solution is:

$$f_{\text{opt}}(s) = \frac{1}{2} + \frac{2s-1}{2\sqrt{N-(N-1)(2s-1)^2}} \tag{22}$$

and the τ is given by:

$$\tau = \frac{\sqrt{N-1}}{\delta} \tag{23}$$

Which implies that the complexity of the above algorithm is of order $\sqrt{N}$.

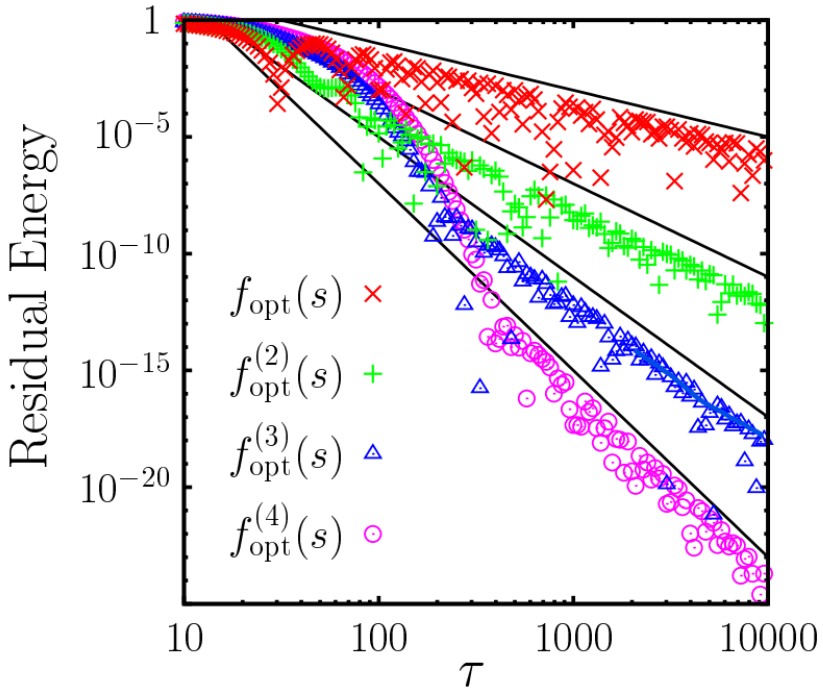

Figure 1: (Figure 6 in the paper) The annealing-time dependence of the residual energy for the database search problem, the scales are logarithmic, the fitting lines varifies equation 15. where $f^{(k)}_{\text{opt}}(s)$ are annealing schedules such that $f^{(k)}_{\text{opt}}(s) = f_{\text{opt}}(f_k(s))$, such that $\frac{d^{k-1}f_k(s)}{ds}\Big|_{\{0,1\}} = 0$ and $f^{(k)}_{\text{opt}}(0,1) = \{0,1\}$.

2.4 Quantum Monte-Carlo (Section 5)

Since Quantum computers are not yet mass produced, and the time to solve the Schrodinger equation increases exponentially, it is necessary to consider other methods to approximate the solutions and use it to apply and test quantum annealing by classical computers.
The method that is used to solve the Schrodinger equation is the path integral Monte Carlo method (PIMC), this section discussed and proved the convergence condition (Strong and weak ergodicity).

2.4.1 Monte Carlo method:

Denote the space of possible discrete states by $\mathcal{S}$. With each "time" step, we denote the transition probability from state x to state y at time t is given by the transition matrix $G(x, y; t)$ defined as:

$$G(y,x;t) = \begin{cases} P(y,x)A(y,x;t) & \text{if } x \neq y, \\ 1 - \sum_{z \in S} P(z,x)A(z,x;t) & \text{if } x = y. \end{cases} \tag{24}$$

Where $P(y, x)$ is the generation probability, and $A(x, y)$ is the acceptance probability, and $P(y, x)$ should satisfy the following:

$$\forall x, y \in S : P(y,x) = P(x,y) \geq 0, \tag{25}$$

$$\forall x \in S : P(x,x) = 0, \tag{26}$$

$$\forall x \in S : \sum_{y \in S} P(y,x) = 1. \tag{27}$$

$$\forall x, y \in S, \exists n > 0, \exists z_1, \ldots, z_{n-1} \in S : \prod_{k=0}^{n-1} P(z_{k+1}, z_k) > 0, \quad z_0 = x, \quad z_n = y. \tag{28}$$

Assuming an initial probability distribution $p(t_0) = p_0$ the probability distribution column vector (whose elements represent the probability of a certain state $[p]_x = p(x)$) at t is given by:

$$p(t, t_0) = G(t-1)G(t-2)\ldots G(t_0)p_0. \tag{29}$$

The norm of a probability distribution is given by:

$$||p|| = \sum_{x \in S} |p(x)|. \tag{30}$$

There are two types of convergence associated with the probability distributions, one is the Weak ergodicity, which means that the final distribution $p(t, t_0)$ is independent of the initial distribution p_0, and is formally defined as:

$$\forall t_0 \geq 0 : \lim_{t \to \infty} \sup\{||p(t,t_0) - p'(t,t_0)|| \mid p_0, p_0' \in P\} = 0. \tag{31}$$

While Strong ergodicity is the condition that the probability distribution converges to a unique distribution r, which does not depend on the initial distribution as well (clearly Strong ergodicity implies Weak ergodicity).

$$\exists r \in P, \forall t_0 \geq 0 : \lim_{t \to \infty} \sup\{||p(t,t_0) - r|| \mid p_0 \in P\} = 0. \tag{32}$$

In this review, I will not state the proof of the criteria for weak and strong ergodicity, I'll only state them.
Condition for **Weak ergodicity**:

$$\sum_{t=0}^{\infty} \min \left\{ \sum_{z \in S} \min\{G(z,x), G(z,y)\} \mid x, y \in \mathcal{S} \right\} \longrightarrow \infty \tag{33}$$

Or intuitively, the sum of the transitions from the "furthest" element goes to ∞, implying that no element gets "too far".

While the conditions for **strong erdgocity** are the following:

1. Weak ergodicity,
2. For all t, there exists a stationary state $p_t \in P$ such that $p_t = G(t)p_t$,
3. p_t satisfies $\sum_{t=0}^{\infty} \|p_t - p_{t+1}\| < \infty$.

2.4.2 Path-integral Monte Carlo (PIMC)

The Path-integral Monte Carlo is a method to solve the Schrodinger equation using the path integral formulation of quantum mechanics, in the path integral montecarlo formulation, we redefine the d-dimensional transverse field ising model by a $(d+1)$-dimensional classical Hamiltonian that can be worked out on a classical computer. The partition function associated with the problem can be in general written as:

$$Z(t) = \sum_{x \in S} \exp\left(-\frac{F_0(x)}{T_0} - \frac{F_1(x)}{T_1(t)}\right) \tag{34}$$

were $F_1(x)$ is the generalization of kinetic energy Hamiltonian (transverse field) in equation 7.
Using this equation, the acceptance probability used in equation 24 is given by:

$$A(y, x; t) = g\left(q(y; t)/q(x; t)\right) \tag{35}$$

Where $g(u) = u/(1+u)$ or $g(u) = \min\{u, 1\}$ and $q(x, t)$ is given by:

$$q(x; t) = \frac{1}{Z(t)} \exp\left(-\frac{F_0(x)}{T_0} - \frac{F_1(x)}{T_1(t)}\right) \tag{36}$$

Theorem 5.3 states that the **PIMC** is strongly erdgocic, under the following condition:

$$T_1(t) \geq \frac{RL_1}{\log(t+2)} \tag{37}$$

Applying this condition to the transverse field ising model depicted in equation 7 will result into the following time dependence :

$$\Gamma(t) \geq \frac{M}{\beta} \tanh^{-1}\left(\frac{1}{(t+2)^2/RL_1}\right) \tag{38}$$

were R is the minimum number of maximum steps needed to reach any state, starting from any state, excluding the states where $F_1(x)$ has local maximum (This is essential for the proof, that I have not included here). And L_1 is the maximum change in F_1 in a single monte carlo step
For large values of t:

$$\Gamma(t) \geq \frac{M}{\beta(t+2)^{-2/RL_1}} \tag{39}$$

Which is of the same order as equation 8 since $L_1 = \mathcal{O}(N)$ (Notice that it is a maximum).

2.4.3 Green's Function Monte Carlo

Another way to solve the Schrodinger equation is as follows:

$$|\psi(t)\rangle = T \exp\left(-\int_0^t dt' H(t')\right) |\psi_0\rangle \tag{40}$$

The solution can be approximated as:

$$|\psi(t)\rangle = \hat{G}_0(t_{n-1})\hat{G}_0(t_{n-2}) \dots \hat{G}_0(t_1)\hat{G}_0(t_0)|\psi_0\rangle \tag{41}$$

Were $\hat{G}_0(t) \equiv 1 - \Delta t \cdot H(t)$.

In recursive form:

$$\psi_{k+1}(y) = \sum_x \hat{G}_1(y, x; t_k)\psi_k(x) \tag{42}$$

Where $\hat{G}_1(t) \equiv 1 - \Delta t(H(t) - E_T)$ and E_T is called reference energy and it is an energy close to the ground state. E_T is added in order to accelerate the convergence of **(GFMC)**.
In order to transform equation 42 to a Markov process we rewrite $\hat{G}$ as:

$$\hat{G}_1(y, x; t) = G_1(y, x; t)w(x; t) = \hat{G}_1(y, x; t)\frac{w(x; t)}{\sum_y \hat{G}_1(y, x; t)} \tag{43}$$

$$\psi_n(y) = \sum_{\{x_k\}} \delta_{y,x_n} \prod_{k=0}^{n-1} w(x_k; t_k) \cdot G_1(x_{k+1}, x_k; t_k) \cdot \psi_0(x_0) \tag{44}$$

For the case of the **TFIM** the elements of Green's function is given by:

$$\hat{G}_1(y,x;t) = \begin{cases} 1 - \Delta t[\langle x| \left(-\sum_{ij} J_{ij}\sigma_z^i\sigma_z^j\right)|x\rangle - E_T] & (\text{if } x = y) \\ \Delta t\Gamma(t) & (\text{if } x \text{ and } y \text{ differ by a single-spin flip}) \\ 0 & (\text{otherwise}) \end{cases} \tag{45}$$

The convergence theorem for **GFMC**:
The Marcov chain defined in equations 45, 43 and 44 is strongly ergodic if:

$$\Gamma(t) \geq \frac{b}{(t+1)^c}, \quad 0 < c \leq \frac{1}{N} \tag{46}$$

3 Strengths

The paper went into mathematical details and proofs of all theorems it provided, it has also stated the results in an organized way, with a clear introduction and conclusion. It was also clear in the objectives of each section and introduced the results in a clear way.
The sections were clearly divided, and further divided into subsections with a clear transition between them. Which helped the reader to navigate the paper as well as summerizing it.

4 Weaknesses

The paper introduced the concepts in a rigorous way that did not provide any intuition, it also did not have a clear definition of the annealing in general before delving into the related theorems. The intermediate steps were not clear, and some transitions in the proofs and concepts were vague, unless read by experts in the field of quantum computing.
For example, the paper had no mention of quantum tunneling, and dealt with it via the Schrodinger equation and did not emphasize its significance.

5 Discussion AND Conclusion

In conclusion, this paper explores a generic algorithm to solve optimization problems (QA) comparing it with the classical (SA), and discussing its convergence conditions and the errors associated with it.
It should be noted that the (QA) algorithm should be implemented using quantum computers, which are currently not available, therefore it is essential to find classical algorithms for the sake of implementing (QA) using currently available classical computers, the paper explored two such algorithms (PIMC and GFMC) and discussed their convergence and complexity.
Many of the convergence theorems in this paper are taken by Physicists for granted. The proofs are not written in full details in the paper, and they require prior knowledge in the basics of quantum annealing (in terms of what it is), as well as experience in dealing with the Pauli matrices and Dirac notation. In terms of mathematical knowledge, one should be familiar with probability theory (for section 5).