LIMITS ON QUANTUM ADIABATIC OPTIMIZATION WARNING: ROUGH MANUSCRIPT!

WIM VAN DAM AND UMESH VAZIRANI

1. Introduction

We exhibit a family of 3SAT instances for which quantum adiabatic optimization provably fails, since the spectral gap is exponentially small. This failure highlights the inability of the algorithm to distinguish between local and global optima.

2. Physics Background

2.1. Quantum states, Hamiltonians, Schrödinger's equation.

- A quantum state Ψ is a linear superposition of basis states $\{|x\rangle\}_x$ and is described by the summation $|\Psi\rangle = \sum_x \alpha_x |x\rangle$, with *amplitudes* $\alpha_x \in \mathbb{C}$ and $|\alpha_x^2|$ the probability of observing the state x.
- Formally, for an N dimensional system, the quantum state is represented by a ℓ_2 normalized vector in the N dimensional Hilbert with complex amplitudes.
- Throughout this article we assume that our quantum system is finite dimensional (the dimension is typically denoted by N).
- A *change* of the state Ψ is represented by a rotation of the vector $|\Psi\rangle$ in the Hilbert space.
- How the state changes is determined by the *Hamiltonian* H, which describes the forces that act on the state Ψ.
- The Hamiltonian, which can be time dependendent, acts as a linear operator on the state vector, and hence is described by a matrix $\in \mathbb{C}^{N\times N}$. The time evolution is expressed by Schrödinger's differential equation:

$$\mathrm{i}\hbar \frac{\mathrm{d}|\Psi(t)\rangle}{\mathrm{d}t} = H(t)|\Psi(t)\rangle, \tag{1}$$

  where $\hbar$ is Planck's constant $h \approx 6.6262 \times 10^{-34}$ Joule seconds dived by 2π.
- The requirement that the vector $|\Psi\rangle$ stays normalized is reflected by the requirement that the matrix H is *Hermitian* ($H^\dagger = H$).
- Crucial for an understanding of a Hamiltonian is the spectral decomposition into its eigenstates (=eigenvectors) and corresponding eigen energies(?) (=eigenvalues) of $H(t)$. The units of the eigenvalues is energy and the state with the smallest eigenvalue is called the *ground state.*
- For a constant Hamiltonian $H = \sum_j \lambda_j |\phi_j\rangle\langle\phi_j|$ and $|\Psi(0)\rangle = \sum_j \beta_j |\phi_j\rangle$, then $|\Psi(T)\rangle = \sum_j \mathrm{e}^{2\pi \mathrm{i}\lambda_j T}\beta_j|\phi_j\rangle$.

2.2. **Quantum Adiabatic Theorem.**

- Born and Fock'27: Let $H(s)$ for $0 \leq s \leq 1$ describe a continuous evolution from the 'initial Hamiltonian' $H(0)$ to the 'final Hamiltonian' $H(1)$. If 1) the initial state of the system is in the ground state $H(0)$ 2) for every s the ground state of $H(s)$ is unique and the evolution from $H(0)$ to $H(1)$ is done 'slow enough', then the final state will be the ground state of $H(1)$.
- quantitative version: 'Slow enough' means: not faster than $1/\text{gap}^2(H(s))$, where $\text{gap}(H(s))$ is the gap between the two smallest eigenvalues of $H(s)$.

3. Quantum Adiabatic Optimization

3.1. **The Idea by Farhi et al.**

- The ground state of $H(1)$ can be the solution to an minimization problem: For example, let $H(1) = \sum_z f(z)|z\rangle\langle z|$ with f a cost function over a large domain.
- Crucial for this to work are the requirements
  1. The evolution $H(0) \to H(1)$ must be efficiently implementable
  2. The ground state of $H(0)$ must be efficiently constructable
  3. The gap of $H(s)$ during the evolution has to be big (in comparison with the dimension = size of domain).
- Farhi et al. Suggested that $H(s) = (1-s)H_0 + sH_f$ with $H_0 := \sum_j \sigma_x^{(j)}$ and $H_f := \sum_z f(z)|z\rangle\langle z|$ with $z \in \{0,1\}^n$ in the computational basis.

3.2. **Efficient Implementation of QAO.**

3.3. **Quantum Walks on the Hypercube and Other Graphs.**

underlying graph + quantum walks changing walk - simulated annealing gap ... trouble computing ground state.

delete rest of subsection

Before we analyze the power of the quantum adiabatic method we will first try to get some understanding of this approach to optimization. A standard approximation technique tells us that for fixed Hamiltonian $H = (1-s)H_0 + sH_f$ we have $\mathrm{e}^{\mathrm{i}H} \approx (\mathrm{e}^{\mathrm{i}(1-s)H_0/m} \cdot \mathrm{e}^{\mathrm{i}sH_f/m})^m$, which gets more precise as m gets bigger. We can decompose the effect of H into the effect of H_0 and H_f.

The Effect of H_f: The operation of H_f is clear: a phase change $|z\rangle \mapsto \mathrm{e}^{\mathrm{i}f(z)}|z\rangle$ proportional to the function value $f(z)$.

Effect of H_0: Equally important is the effect of the initial Hamiltonian, which is a combination the single-qubit Hamiltonians

$$\sigma_x := \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \tag{2}$$

that is applied to each of the n qubits: $H_0 := \sum_{j=1}^n \sigma_x^{(j)}$ (where we used the convention of omitting the trivial Hamiltonians such that $\sigma_x^{(j)}$ stands for $I^{(1,\dots,j-1)} \otimes \sigma_x^{(j)} \otimes I^{(j+1,\dots,n)}$). The eigenstates of σ_x are the two uniform superpositions $|\hat{0}\rangle := \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ (ground state with eigenvalue 0) and $|\hat{1}\rangle := \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$ (excited state with eigenvalue 1). With this knowledge it is straightforward to see that

- H_0 has as ground state the uniform superposition $\frac{1}{\sqrt{2^n}}\sum_z |z\rangle$ (with energy 0)

- The coupling $\langle z'|H_0|z\rangle$ between two different strings is non-zero only for those (z, z') that have Hamming distance 1. This is reflected by the fact that the non-zero off-diaginal terms of H_0 define the adjacency matrix of the hypercube $\{0,1\}^n$.

Joint Effect of $(1-s)H_0 + sH_f$: By viewing the effect of the Hamiltonian $H(s)$ as the joint effect of $(1-s)H_0$ and sH_f we see a combination of a 'mixing operation' (defined by H_0) along the edges of the hypercube $\{0,1\}^n$ and a phase change according to the values of the cost function f.

This setting and its numerous has proven to be extremely useful in the theory of quantum computation. To position the quantum adiabatic method, we list the following distinctions that researchers have made.

- **Continuous time versus discrete time evolution:** The adiabatic algorithm uses a Hamiltonian that changes smoothly in a continuous time fashion. Algorithms expressed in the circuit model will have use the unitary transformations $\mathrm{e}^{\mathrm{i}H_0}$ and $\mathrm{e}^{\mathrm{i}H_f}$ in a discrete time fashion such that the Trotter approximation $\mathrm{e}^{\mathrm{i}(1-s)H_0+\mathrm{i}sH_f} \approx \mathrm{e}^{\mathrm{i}(1-s)H_0} \cdot \mathrm{e}^{\mathrm{i}sH_f}$ does not necessarly hold.
- **Constant Hamiltonians versus Changing Hamiltonians:** The adiabatic method uses a Hamiltonian $H(s)$ that changes as the algorithm proceeds from $s=0$ to $s=1$. If one keeps the Hamiltonian fixed, then it is often possible to analyze the behaviour of the algorithm in greater detail. Typically in this setting the state of the system will find the solution in a periodic fashion (cf. Grover's search).
- **Searching databases versus complicated cost functions:** In a database search the cost function (which needs to be minimized) is a simple delta function with $f(z)=0$ for the target value, and $f(z')=1$ otherwise. In heuristics we deal with a more complicated function f that defines an (energy) landscape over the graph of search items. This allows an algorithm to use the shape of the landscape as a guidance towards the solution, but also carries the problem of local versus global minima.
- **Different search graphs:** Besides the hypercube, we can use other mixing operations that correspond to other grpahs like the d dimensional grid $(\mathbb{Z}/N\mathbb{Z})^d$ and the complete graph K_N. With the complete graph the algorithm does not have any notion of distance between the strings, hence this correspond where we treat each search items as a set without any further structure (typical in the database search algorithm).
- **Searching versus mixing** An important application of walks on a graph is the possibility of rapidly mixing to count the number elements of a structure. (Note that one can use the adiabatic algorithm in reverse to establish this effect.)

4. Skeleton

4.1. **Notation.** Time dependent Hamiltonian $H(t)$; ground state ψ_0 of the time dependent Hamiltonian; smallest eigenvalue λ_0, next smallest eigenvalue: λ_1 such that the gap is $\lambda_1 - \lambda_0$.

4.2. **For general Connected Graphs.** We will assume that the initial Hamiltonian H_0 is of the form $H_0 = -cA + dI$ where A is the adjacency matrix of an undirected connected graph, and c is positive. Since the final Hamiltonian H_f is diagonal, the convex combination $H_f(\tau)$ also has a similar form to H_0: namely $-cA + D$ where c is a positive constant and D is diagonal.

The following lemma proves a useful property of the ground state of the instantenuous Hamiltonian $H_f(\tau)$ at every point during the adiabatic process.

Lemma 1. *Let H be a Hamiltonian of the form $-cA + D$ with A the adjacency matrix of a undirected connected graph, c is positive and D is diagonal. Then H has a unique ground state $|\psi_0\rangle = \sum_x \alpha_x |x\rangle$. Moreover we can assume that all α_x are positive reals.*

Proof. Since we are interested in the ground state of H, we can equivalently consider the principal eigenvector (the eigenvector corresponding to the largest eigenvalue) of $M := I - \delta H$, with $\delta > 0$ small enough such that M is a nonnegative matrix and the diagonal entries of M are all positive. The pseudo-graph of M equals the graph of A where we include all self-loops (due to the positive diagonal of M). Because M is nonnegative we have $M^w > 0$ with w the finite width or diameter of the graph, hence M is primitive (see [1, Chapter 8]). This shows that H has a unique smallest eigenvalue and all entries of the corresponding eigenvector can be assumed to be positive. $\square$

The above lemma applies to all but the final ground state of the adiabatic process (where $c = 0$). For the final Hamiltonian $H(1) = H_f$, the states $|s\rangle$ that achieve the minimum $f(s)$ are the ground states of $H(1)$. Hence if f has a unique minimum, then $H_f(\tau)$ will have a unique ground state throughout the whole evolution $\tau = 0 \to \tau = 1$. If f has several minima, then the ground state of H_f is a space spanned by the vectors $|x_1\rangle$, $|x_2\rangle$ for the x_j that achieve this minimum.

We now derive a lowerbound on the spectral gap in the very general setting of connected graph Hamiltonians, thus providing an upperbound on the running time of the adiabatic optimization algorithm. We note that we must add one additional constraint here — that the cost function being optimized is polynomially bounded. In terms of the notation above, this means that the diagonal matrix D in $H_f(\tau) = -cA + D$ has entries bounded by a polynomial in $\log N$, where N is the dimension of the Hamiltonian.

Lemma 2. *For connected graph Hamiltonians, the spectral gap will be no smaller than $2^{-w\log w}$, where w is the width/diameter of the graph.*

Proof. [**Here we will prove that on the cube $\{0,1\}^n$ the quantum adiabatic method will always succeed in time $2^{O(n\log n)}$.**] Given the Hamiltonians H_0 and H_f, define

$$M := -H_0 - \lambda H_f, \tag{3}$$

such that the off-diagonal terms of M are 0 or +1 and we the gap $g(M)$ between the biggest two eigenvalues of M equals the $(1+\lambda)$ times the gap of the two smallest eigenvalues of $\frac{1}{\lambda+1}H_0 + \frac{\lambda}{1+\lambda}$. For $\lambda \geq n+1$ be the adjusted Hamiltonian that has off diagonal entries 0 or 1, is non-negative (entry wise) and positive definite (eigenvalue wise). For this, p, q don't have to be bigger than $\mathrm{poly}(n)$. By the mixing properties of the hypercube, we see that M^n is a positive matrix with entries ≥ 1, which shows that the gap of M^n between the two biggest eigenvalues is bigger than 1. Hence, the gap of H is not smaller than $2^{-\mathrm{poly}(n)}$. $\square$

4.3. **Adiabatic Optimization on the Hypercube.** We now consider adiabatic algorithms for optimization problems over binary strings. In this context the natural graph to consider is the n-dimensional hypercube. Luckily the corresponding initial Hamiltonian H_0 is natural and easy to implement, since $H_0 = \sigma_x^{(1)} + \ldots \sigma_x^{(n)}$. This is the initial Hamiltonian suggested by Farhi, et al.

The underlying graph plays a crucial role in the evolution of the adiabatic process, since the Hamiltonian only allows interaction between adjacent vertices. The key to our analysis of the adiabatic process is to view it as a quantum analog of local search. We show that if the cost function is monotone on the hypercube, then so is the ground state at every point during the adiabatic process.

Definition 1 (Monotonicity on the hypercube)**.** *Consider the hypercube* $\{0,1\}^n$ *with the partial ordering on its strings* $x \preccurlyeq y$ *if and only if* $x_i \leq y_i$ *for all* $i = 1, \ldots, n$*. A function* $f : \{0,1\}^n \to \mathbb{R}$ *is* monotonically decreasing *if* $x \preccurlyeq y$ *implies* $f(x) \geq f(y)$*. Similarly for* monotonically increasing *functions.*

Recall that the ground state ψ_0 of $H_f(s)$ has positive real amplitudes α_x. We say that ψ_0 is monotone decreasing if the function $x \mapsto \alpha_x$ is monotone decreasing.

Lemma 3. *For a monotone increasing cost function* f*, the ground state* ψ_0 *of* $H_f(s)$ *is monotone decreasing as well.*

Proof. Let $M = I - \delta H_f(\tau)$. First observe that since ϕ is the unique principal eigenvector of M, we can use the limit $|\phi\rangle \propto \lim_{n\to\infty} M^n|0^n\rangle$. Also M has interesting structure — it can be written as a sum of two operators G_0 and G_1, where G_0 is the adjacency matrix of the hypercube, and G_1 is diagonal and monotonically decreasing. $|0^n\rangle$ is monotone, and moreover G_0 and G_1 each have the property that when they operate on a monotone vector the result is a monotone vector. It follows that $M^n|0^n\rangle$ is monotone, and therefore so is $|\phi\rangle$. $\square$

The intuitive picture behind the above lemma is that the adiabatic process starts off with the uniform superposition on the hypercube, and the ground state progressively shifts amplitude towards the optimum vertex (the base of the hypercube) until it has the entire weight there at $\tau = 1$.

We now use the monotonicity of the ground state to show that the total weight of the ground state at the top levels of the hypercube is neglibile throughout the adiabatic process.

Lemma 4. *Let* p *be a monotonically decreasing probability distribution over the hypercube* $\{0,1\}^n$*. Then, for every* $\delta > 0$*:*

$$\frac{\sum_{w(x)\geq(\frac{1}{2}+\delta)n} p(x)}{\sum_x p(x)} \leq \frac{|\{x : w(x) \geq (\frac{1}{2}+\delta)n\}|}{2^n} = 2^{-\Omega(n)}, \tag{4}$$

where $w(x)$ *denotes the Hamming weight of the string* x*.*

So far we have established that for monotone cost functions, the ground state throughout the adiabatic process has negligible weight at the top levels of the hypercube. It follows that if we modify the cost function at the top levels of the hypercube, then the original ground state still feels almost the same force as before. Intuitively, the dynamics are essentially unchanged and if run for polynomial time the process converges as before to the base of the hypercube, even though the optimum vertex for the modified cost function may lie in the top of the hypercube. Formally, we use some of these considerations to prove an exponentially small upperbound on the spectral gap for the adiabatic process on the modified cost function.

Definition 2. *A cost function $f : \{0,1\}^n \to \mathbb{R}$ is "ϵ-deceptive monotone" if it is monotone for all but a fraction ϵ vertices comprising the top layers of the hypercube. Moreover, we require that the global minimum is not at $f(0^n)$ (and hence in the ϵ-top of the hypercube.*

Theorem 1. *For ϵ-deceptive monotone cost functions the adiabatic process has spectral gap at most ϵ.*

Proof. Let g be an ϵ-deceptive monotone function and let f be a monotone function that only differs from g at the top. How does the adiabatic algorithm perform while trying to minimize g? Since the ground state of $H_f(\tau)$ is always monotone, by the claim above, its ground state ψ_0 'feels' only an exponentially small difference between H_f and H_g. Thus if we were to switch from H_f to H_g, the state $\psi(\tau)$ of the adiabatic algorithm would not notice the difference very much. This idea can be formalized to show that there must be a critical value of τ_c such that the eigenvalue gap of the adiabatic algorithm for minimizing g is exponentially small in n.

To see this, view $H_g(\tau)$ in the rotating eigenbasis $\phi_0(\tau), \phi_1(\tau), \dots$ of $H_f(\tau)$—call this matrix $A(\tau)$. By definition, the ground state of $H_f(\tau)$ is always $\phi_0(\tau)$. At $\tau = 0$ and $\tau = 1$, H_f and H_g have the same eigenbasis and hence both $A(0)$ and $A(1)$ are diagonal matrices. Moreover, the initial $A(0)$ has the same ground state $\phi_0(0)$, while the minimum eigenvalue of the final $A(1)$ corresponds to the state $|\phi_s(1)\rangle = |s\rangle$ that minimizes g, where this s is a string with Hamming weight $\geq (\frac{1}{2} + \epsilon)n$.

The main claim is that for every $0 \leq \tau \leq 1$, the interaction between $\phi_0(\tau)$ and the other eigenstates $\phi_j(\tau)$ is expontially small. Which gets expressed by the fact that the norm of the off-diagonal entries of $A(\tau)$ corresponding to the state $|0^n\rangle$ are exponentially small. This is because when written in the standard basis, $|0^n\rangle$ is monotone, and so 'feels' only an exponentially small difference between f and g.

Now let $B(\tau)$ be the matrix that results from zeroing out the off-diagonal entries in the $|0^n\rangle$ row and column of $A(\tau)$. Now, $B(\tau)$ has two blocks, and its initial minimum eigenvalue at $\tau = 0$ starts in the first block and ends up (at $\tau = 1$) in the second block. It follows that there is a critical τ_c such that there is a level-crossing between the two smallest eigenvalues of $B(\tau)$. But since $\|A(\tau) - B(\tau)\|$ is exponentially small, it follows that at this same moment the eigenvalue gap of $A(\tau_c)$ is exponentially small as well. □

Remark: The properties in this sections hold more general for "sedimented graphs" and their corresponding monotone functions.

Corollary 1. *The following 3SAT function is ϵ-deceptive monotone.*

4.4. **A Hard Family of 3SAT Instances.** We have just shown how functions with a hidden global minimum can mislead adiabatic algorithms. Here we show how such a fooling function occurs when trying to solve the NP-complete problem 3SAT via adiabatic computing, as suggested by Farhi et al. The 3SAT problem asks us to decide whether a Boolean formula $\Psi(x_1, \dots, x_n)$ in '3-conjunctive normal form' is satisfiable by an assignment $x \in \{0,1\}^n$, or not (where we use the values 0 (1) to denote that a variable is set to be False (True)). The normal form that is used requires that the formula is a conjunction OR-clauses of three variables.

The adiabatic approach tries to solve the 3SAT problem by minimizing the number of violated clauses over the set of 2^n possible assignments. More formally, given a 3CNF formula Ψ the cost function $V_\Psi : \{0,1\}^n \to \mathbb{N}$ is defined, where $V_\Psi(x) := \#(\text{violated clauses in } \Psi(x))$. Our goal is now to define a Boolean fomula Φ that has a cost function that is 'deceptive' in the sense of the previous section. For a simple example, consider first the formula

$$\Phi(x_1,\dots,x_n) := \bigwedge_{1\le i,j,k\le 3} (x_i \vee \bar{x}_j \vee \bar{x}_k), \tag{5}$$

which has n^3 clauses. By symmetry, one sees that the cost $V_\Psi(x)$ only depends on the Hamming weight $w(x)$ of the assignment $x \in \{0,1\}^n$. Specifically, the string x with $w(x)$ assignments $x_i = \text{True}$ (combined with $(n-w(x))$False values) will have $(n-w(x))w(x)^2$ different ways of violating the $(x_i \vee \bar{x}_j \vee \bar{x}_k)$ clauses. Hence we have for the cost function $V_\Phi(x) = (n-w(x))w(x)^2$, which is a third degree polynomial in $0 \le w(x) \le n$ that has its two minima at the extreme points $x = 0^n$ and $x = 1^n$. What is important for our purposes here is that the *maximum* of V_Φ is reached for the strings with Hamming weight $\frac{2}{3}n$, and that on the part on the hypercube between 0^n and $w(x) = \frac{2}{3}n$, V_Φ is monotone increasing. To get a truelly fooling function, we need to make entry $x = 1^n$ the unique minimum of the function V, but this is easily done by the following 3CNF formula:

$$\Phi'(x_1,\dots,x_n) := (x_1 \vee x_1 \vee x_1) \quad \wedge \bigwedge_{1\le i,j,k\le n} (x_i \vee \bar{x}_j \vee \bar{x}_k). \tag{6}$$

Note that indeed the only satisfying assignment of Φ' is obtained by setting all variables $x_1,\dots,x_n$ to be true. The satisfiablity of Φ' thus gets expressed by the global minimum $V_{\Phi'}(1^n) = 0$. In this case, the value of $V_{\Phi'}(x_1,\dots,x_n)$ only gets determined by x_1 and the Hamming weight $0 \le |w(x)| \le n$ of the string $x = x_1 \dots x_n$, which indicates how many variables x_j are True:

$$V_{\Phi'}(x_1,\dots,x_n) = (1-x_1) \quad + \quad (n-w(x)) \cdot w(x)^2. \tag{7}$$

On the domain $w(x) < \frac{2}{3}n$, this functoin is monotone, while its global minimum is obtained at $V_{\Phi'}(1^n) = 0$. This thus fits the general lower bound scenario of the previous section, hence the minimization of $V_{\Phi'}$ requires an exponential slowdown for the quantum adiabatic algorithm.

References

[1] Roger A. Horn and Charles R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.

Wim van Dam
Department of Computer Science
University of California, Santa Barbara
Santa Barbara, CA 93106, USA
E-mail address: vandam@cs.ucsb.edu

Umesh Vazirani
Department of Computer Science
University of California, Berkeley
Berkeley, CA 94720, USA
E-mail address: vazirani@cs.berkeley.edu