Lecture 12

Approximation algorithms for local Hamiltonian problems

Scribes: Zihan Hu and Zijing Di

In this lecture, we study approximation algorithms for local Hamiltonian problems. Specifically, we have a k-local Hamiltonian instance $H = \frac{1}{m}\sum_{i=1}^{m} H_i$ on n qudits with $H_i \geq 0$ in mind, and ideally we would like to find an approximation α of $\lambda_{\min}(H)$ within multiplicative error β, i.e.

$$(1-\beta)\cdot\lambda_{\min}(H) \leq \alpha \leq (1+\beta)\cdot\lambda_{\min}(H).$$

We will approximate $\lambda_{\min}(H)$ with a product state. The rest of the lecture is devoted to answering the following two questions:

1. What is the best approximation factor for a product state?

2. Can we find the best product state?

12.1 Product state approximation

In this section, we will answer the first question.

Theorem 12.1. *Suppose that H is k-local on n qudits (d dimensional) such that $H_i \geq 0, \forall i$. Then there exists a product state $|\phi\rangle = |\phi_1\rangle \otimes |\phi_2\rangle \otimes \cdots \otimes |\phi_n\rangle \in (\mathbb{C}^d)^{\otimes n}$ such that*

$$\langle\phi|H|\phi\rangle \geq \lambda_{\max}(H)/d^{k-1}\ .$$

Remark 12.2. Here we use $\lambda_{\max}$ instead of $\lambda_{\min}$ because on the one hand, it is too good to be true to find a product state $|\phi\rangle$ with $\langle\phi|H|\phi\rangle \leq \lambda_{\min}(H)/2^{k-1}$ (when $\lambda_{\min}(H)=0$, it can be the case that all ground states are highly entangled); on the other hand, we can always use simple transformation to swap between $\lambda_{\min}$ and $\lambda_{\max}$, so it suffices to consider $\lambda_{\max}$.

Remark 12.3. The coefficient $1/d^{k-1}$ is the best we can hope for. Consider the following example: the Hamiltonian $H = I - \frac{1}{2}(|00\rangle + |11\rangle)(\langle 00| + \langle 11|)$ is 2-local on 2 qubits. Then $\lambda_{\max}(H) = 1$ while for any product state $|\phi\rangle$, $\langle\phi|H|\phi\rangle = \frac{1}{2}$.

Before we show a formal proof of Theorem 12.1, we first see a simple case where $d = 2, n = 2$ to illustrate the proof idea.

Let us fix an instance, a Hamiltonian H on 2 qubits. Suppose $|\psi\rangle$ is the 2-qubit state that achieves the desired $\lambda_{\max}(H)$. To relate it with a product state, we do a Schmidt decomposition on it

$$|\psi\rangle = \sum_{i=1}^{2} \sqrt{\lambda_i}|u_i\rangle|v_i\rangle$$

where $\lambda_1 + \lambda_2 = 1$, and $\{|u_i\rangle\}_i, \{|v_i\rangle\}_i$ are basis for $\mathbb{C}^2$.

$$\lambda_{\max}(H) = \langle\psi|H|\psi\rangle = \sum_{i=1}^{2} \lambda_i\langle u_i, v_i|H|u_i, v_i\rangle + 2\sqrt{\lambda_1\lambda_2}\mathsf{Re}(\langle u_1, v_1|H|u_2, v_2\rangle)$$

Notice that for any vectors $|a\rangle, |b\rangle$,

$$\begin{aligned}\mathsf{Re}(\langle a|H|b\rangle) =& \mathsf{Re}(\langle a|\sqrt{H}\cdot\sqrt{H}|b\rangle)\\ \leq& \|\sqrt{H}|a\rangle\|\cdot\|\sqrt{H}|b\rangle\|\\ =& \sqrt{\langle a|H|a\rangle}\sqrt{\langle b|H|b\rangle}\\ \leq& \frac{1}{2}(\langle a|H|a\rangle + \langle b|H|b\rangle)\end{aligned} \tag{12.1}$$

Thus we have that

$$\lambda_{\max}(H) \leq 2\sum_{i=1}^{2} \lambda_i\langle u_i, v_i|H|u_i, v_i\rangle = 2\mathsf{Tr}(H\rho) \ ,$$

where $\rho = \sum_{i=1}^{2} \lambda_i|u_i, v_i\rangle\langle u_i, v_i|$. And therefore, $\max_i\langle u_i, v_i|H|u_i, v_i\rangle \geq \mathsf{Tr}(H\rho) \geq \frac{1}{2}\lambda_{\max}(H)$.

The general case is similar, except that we decompose the state qudit by qudit. We first define the transformation to decompose the state qudit by qudit.

Definition 12.4. For an n-qudit state $|\psi\rangle$, we first do Schmidt decomposition on $|\psi\rangle$ on the first qudit and the rest of the system:

$$|\psi\rangle = \sum_{i_1=1}^{d} \sqrt{\lambda_{i_1}}|u_{i_1}\rangle|v_{i_1}\rangle \ .$$

Then we continue to do Schmidt decomposition on each $|v_{i_1}\rangle$ on the first qudit and the rest of the system: for each i_1,

$$|v_{i_1}\rangle = \sum_{i_2=1}^{d} \sqrt{\lambda_{i_1,i_2}}|u_{i_1,i_2}\rangle|v_{i_1,i_2}\rangle \ .$$

We repeat the process until there are no more qudits. We can write the state $|\psi\rangle$ as

$$|\psi\rangle = \sum_{i_1=1}^{d}\cdots\sum_{i_{n-1}=1}^{d} \sqrt{\lambda_{i_1}\lambda_{i_1,i_2}\cdots\lambda_{i_1,i_2,\cdots,i_{n-1}}}|u_{i_1}\rangle|u_{i_1,i_2}\rangle\cdots|u_{i_1,i_2,\cdots,i_{n-1}}\rangle|v_{i_1,i_2,\cdots,i_{n-1}}\rangle \ .$$

The state $\rho(|\psi\rangle)$ is defined as

$$\rho(|\psi\rangle) = \sum_{i_1=1}^{d}\cdots\sum_{i_{n-1}=1}^{d} \lambda_{i_1}\cdots\lambda_{i_1,i_2,\cdots,i_{n-1}}|u_{i_1},\cdots u_{i_1,i_2,\cdots,i_{n-1}}, v_{i_1,i_2,\cdots,i_{n-1}}\rangle\langle u_{i_1},\cdots u_{i_1,i_2,\cdots,i_{n-1}}, v_{i_1,i_2,\cdots,i_{n-1}}|$$

Note that this is a well-defined density matrix.

We will show the following:

Lemma 12.5. *For every Hamiltonian H_i acting on k out of n qudits and every n-qudit state $|\psi\rangle$,*

$$d^{k-1}\mathsf{Tr}(H_i\rho(|\psi\rangle)) \geq \langle\psi|H_i|\psi\rangle \ .$$

Before we prove the lemma, we first show how it can be used to prove Theorem 12.1.

Proof of Theorem 12.1. Now we have an instance $H = \frac{1}{m}\sum_{i=1}^{m} H_i$ on n qudits where each H_i is acting on k out of n qudits. Suppose $|\psi\rangle$ is the n-qudit state that achieves the desired $\lambda_{\max}(H)$.

By Lemma 12.5 and the linearity,

$$d^{k-1}\mathsf{Tr}(H\rho(|\psi\rangle)) \geq \langle\psi|H|\psi\rangle = \lambda_{\max}(H) \ .$$

Notice that $\rho(|\psi\rangle)$ is a mixture over product states $|\phi\rangle = |\phi_1\rangle \otimes |\phi_2\rangle \otimes \cdots \otimes |\phi_n\rangle \in (\mathbb{C}^d)^{\otimes n}$. By convexity,

$$\max_{|\phi\rangle=|\phi_1\rangle\otimes|\phi_2\rangle\otimes\cdots\otimes|\phi_n\rangle\in(\mathbb{C}^d)^{\otimes n}} \langle\phi|H|\phi\rangle \geq \mathsf{Tr}(H\rho(|\psi\rangle)) \geq \lambda_{\max}(H)/d^{k-1} \ ,$$

which concludes the proof. □

Now we are left with the proof of Lemma 12.5.

Proof of Lemma 12.5. We prove the lemma by induction on $n + k$.

For the base case $k = 1$, $\mathsf{Tr}(H_i\rho(|\psi\rangle)) = \langle\psi|H_i|\psi\rangle$, because the reduced density matrix of $\rho(|\psi\rangle)$ on any qudit is the same as the reduced density matrix of $|\psi\rangle\langle\psi|$ on the same qudit.

For the induction step, notice that for $|\psi\rangle = \sum_{i_1=1}^{d} \sqrt{\lambda_{i_1}}|u_{i_1}\rangle|v_{i_1}\rangle$, we have that

$$\rho(|\psi\rangle) = \sum_{i_1=1}^{d} \lambda_{i_1}|u_{i_1}\rangle\langle u_{i_1}| \otimes \rho(|v_{i_1}\rangle) \ .$$

If H_i acts on the first qubit,

$$\begin{aligned}\langle\psi|H_i|\psi\rangle =& \sum_{i_1=1}^{d} \lambda_{i_1}\langle u_{i_1}, v_{i_1}|H_i|u_{i_1}, v_{i_1}\rangle + \sum_{1\leq i_1\leq d, 1\leq i'_1\leq d, i_1\neq i'_1} \sqrt{\lambda_{i_1}\lambda_{i'_1}}\mathsf{Re}(\langle u_{i'_1}, v_{i'_1}|H_i|u_{i_1}, v_{i_1}\rangle)\\ \leq& d\sum_{i_1=1}^{d} \lambda_{i_1}\langle u_{i_1}, v_{i_1}|H_i|u_{i_1}, v_{i_1}\rangle \ ,\end{aligned}$$

where the last inequality is due to the same reason as equation 12.1.

By induction hypothesis for the Hamiltonian $\langle u_{i_1}|H_i|u_{i_1}\rangle$ acting on $k-1$ out of $n-1$ qudits, we have that

$$\langle u_{i_1}, v_{i_1}|H_i|u_{i_1}, v_{i_1}\rangle \leq d^{k-2}\mathsf{Tr}(\langle u_{i_1}|H_i|u_{i_1}\rangle\rho(|v_{i_1}\rangle)) = d^{k-2}\mathsf{Tr}(H_i(|u_{i_1}\rangle\langle u_{i_1}| \otimes \rho(|v_{i_1}\rangle))) \ .$$

By linearity,

$$\begin{aligned}\langle\psi|H_i|\psi\rangle \leq& d^{k-1}\mathsf{Tr}(H_i(\sum_{i_1=1}^{d} \lambda_{i_1}|u_{i_1}\rangle\langle u_{i_1}| \otimes \rho(|v_{i_1}\rangle)))\\ =& d^{k-1}\mathsf{Tr}(H_i\rho(|\psi\rangle)) \ .\end{aligned}$$

On the other hand, if H_i does not act on the first qubit, then H_i is a Hamiltonian acting on k out of $n-1$ qudits, by induction hypothesis, we have that

$$\begin{aligned}\langle\psi|H_i|\psi\rangle = &\sum_{i_1=1}^{d} \lambda_{i_1}\langle v_{i_1}|H_i|v_{i_1}\rangle \\ \leq &d^{k-1}\sum_{i_1=1}^{d} \lambda_{i_1}\mathsf{Tr}(H_i\rho(|v_{i_1}\rangle)) \\ = &d^{k-1}\mathsf{Tr}(H_i\rho(|\psi\rangle)) \ ,\end{aligned}$$

which establishes the induction step.

Since both the base case and the induction step have been proved to be true, by mathematical induction, Lemma 12.5 holds. □

For some specific interaction graph G, we might even have better approximation of $\lambda_{\min}(H)$.

1. If the graph G is planar, there exists a PTAS for $\lambda_{\min}(H)$ that finds $\lambda_{\min}(H) \pm \epsilon$ in time $n^{O(1/\epsilon^2)}$. This is because a planar graph can be cut up in pieces and then we can approximate $\lambda_{\min}(H)$ as we did for the 2D grid.

2. If the graph G is D-regular, then $\forall|\psi\rangle$, there exists a product state $|\phi\rangle = |\phi_1\rangle \otimes |\phi_2\rangle \otimes \cdots \otimes |\phi_n\rangle$ such that
$$|\langle\psi|H|\psi\rangle - \langle\phi|H|\phi\rangle| \leq O((\frac{d^2 \ln d}{D})^{1/3}) \ .$$
This suggests that for $d = 2$, if $D \gg 1/\epsilon^3$, then there exists a good product state approximation. We must emphasize that it might be NP-hard to find such state in general. But for dense graph with $D \gg \delta n$, there exists a PTAS to find such a product state. We will see how to find such a good product state for the quantum MAXCUT problem on D-regular graph where $D \gg \delta n$, which partially answers the second question.

12.2 Classical approximation algorithm for MAXCUT on dense graphs

Before we show how to find a good approximation product state for the quantum MAXCUT problem on dense graphs, we first consider its classical version.

Given an undirected graph $G = (V, E)$, where $|V| = n$, the goal of MAXCUT problem is to find $S \subseteq V$ such that $|E(S, \bar{S})| := |\{\{v, w\} \text{ s.t } v \in S, w \in \bar{S}\}|$ is maximized, where $\bar{S} := V \setminus S$. The problem can also be formulated as
$$\max_{x\in\{0,1\}^n} \sum_{\{i,j\}\in E} x_i(1-x_j).$$

Intuitively, this is because x is like a partition of V into set S or $\bar{S}$ depending on whether $x_i = 0$ or 1. If a cut fails to cut an edge $\{i, j\}$, then $x_i = x_j$ and $x_i(1-x_j) = 0$. A maximum cut should maximize the number of non-zero $x_i(1-x_j)$, and hence the formulation.

For every $i \in V$, we define $N(i)$ to be the set of neighbors of i: $N(i) := \{j : \{i,j\} \in E\}$. Then for every x, we denote

$$\begin{aligned} L(x) &:= \sum_{\{i,j\}\in E} x_i(1-x_j) \\ &= \sum_{i\in V} x_i \sum_{j\in N(i)} (1-x_j); \\ R_i(x) &:= \sum_{j\in N(i)} (1-x_j). \end{aligned}$$

Theorem 12.6. *For every $\epsilon > 0$, given an undirected graph $G = (V,E)$ where $|V| = n$, there exists a polynomial time $(\epsilon n^2 + 2n\sqrt{2n\log n})$-additive approximation algorithm for* MAXCUT.

The algorithm has three main procedures:

1. Choose a set S of $O(\log n)$ vertices at random and guess the values $\{x_i\}_{i\in S}$.
2. Use the fixed values $\{x_i\}_{i\in S}$ to estimate $R_i(x)$: Use linear programming the solve the following in polynomial time.
$$\max_{y:\forall i, 0\le y_i\le 1} \sum_i y_i R_i$$
3. "Round" each $y_i \in [0,1]$ to $x_i \in \{0,1\}$.

Roughly speaking, the strategy is reducing the quadratic problem to linear problem by guessing a small part of the vertices and estimating $R_i(x)$, and then reducing the integer problem to linear programming by relaxing the constraints.

Step 1.

Lemma 12.7. *Let $a_1,\dots,a_n \in [-M,M]$. Choose $S \subseteq \{1,\dots,n\}$ at random and with replacement such that $|S| = g\log n$. Then*

$$\sum_{i=1}^n a_i - n\cdot M\cdot\sqrt{\frac{2}{g}} \le \sum_{i\in S} a_i \cdot \frac{n}{|S|} \le \sum_{i=1}^n a_i + n\cdot M\cdot\sqrt{\frac{2}{g}}$$

with probability at least $1 - 1/n^2$ over choice of S.

Proof. We rely on Hoeffding's inequality.

Lemma 12.8 (Hoeffding's). *For independent random variables $z_1,\dots,z_n$, if for every $i \in [n]$, $-M \le z_i \le M$, then*

$$\Pr\left[\left|\sum_i z_i - \mathbb{E}[\sum_i z_i]\right| \ge t\right] \le 2e^{-2t^2/(nM^2)}$$

For $j = 1,\dots,g\log n$ we define the random variable X_j to be the j-th element whose index is chosen to be in the set S, so that these variables are independent (because we chose elements of S with replacement) and for every j, $\mathbb{E}[X_j] = \frac{1}{n}\sum_{i=1}^n a_i$. Applying Hoeffding's inequality with $t = M\cdot|S|\cdot\sqrt{\frac{2}{g}}$ gives

$$\Pr\left[\left|\sum_j X_j - \mathbb{E}[\sum_j X_j]\right| \ge t\right] = \Pr\left[\left|\sum_{i\in S} a_i - \frac{g\log n}{n}\sum_{i=1}^n a_i\right| \ge t\right] \le 2e^{-2t^2/(g\log nM^2)} \le \frac{1}{n^2}.$$

Multiplying $\sum_{i\in S} a_i - \frac{g\log n}{n}\sum_{i=1}^n a_i$ by $\frac{n}{|S|} = \frac{n}{g\log n}$ we get the desired bound. □

We now show how to apply Lemma 12.7 for Step 1. Fix $i \in V$, define

$$a_j := \begin{cases} 1 - x_j & \text{if } j \in N(i) \\ 0 & \text{otherwise} \end{cases}$$

Then we have $\sum_j a_j = R_i(x)$. Randomly choose S such that $|S| = g \log n$. For every $i \in S$, fix $x_i \in \{0, 1\}$. Apply Lemma 12.7 with $g = 1/\epsilon^2$, $M = 1$, then with probability at least $1 - 1/n^2$,

$$\rho_i := \frac{n}{|S|} \sum_{j \in S} a_j \in [R_i(x) - n\epsilon, R_i(x) + n\epsilon]. \tag{12.2}$$

For the same S, union bounding over all $i \in V$, we have that with probability at least $1 - 1/n$, for every $i \in V$, $\rho_i \in [R_i(x) - n\epsilon, R_i(x) + n\epsilon]$.

If our guess of $\{x_i\}_{i \in S}$ is correct, we can compute such ρ_i for each $i \in V$.

Step 2. Suppose that our guess of $\{x_i\}_{i \in S}$ is correct and we have computed ρ_i such that $\rho_i \in [R_i(x) - n\epsilon, R_i(x) + n\epsilon]$ for each $i \in V$ where x is the optimal solution. In this step, we solve the following linear program in polynomial time:

$$OPT^* := \sup_{\substack{y : \forall i \in [n], 0 \le y_i \in 1 \\ \rho_i - n\epsilon \le R_i(y) \le \rho_i + n\epsilon}} \sum_{i \in [n]} y_i \rho_i$$

By the linear program, we obtain y such that

$$\begin{aligned}
\sum_{i \in [n]} y_i R_i(y) &\ge \sum_{i \in [n]} y_i (\rho_i - n\epsilon) \\
&\ge \sum_{i \in [n]} y_i \rho_i - n^2 \epsilon && \text{(since } \textstyle\sum_i y_i \le n\text{)} \\
&\ge \sum_{i \in [n]} x_i \rho_i - n^2 \epsilon && \text{By optimality of linear program} \\
&\ge \sum_{i \in [n]} x_i (R_i(x) - n\epsilon) - n^2 \epsilon && \text{(By Equation 12.2)} \\
&\ge \sum_{i \in [n]} x_i R_i(x) - 2n^2 \epsilon = \mathsf{MAXCUT}(G) - 2n^2 \epsilon && (12.3)
\end{aligned}$$

Step 3. In the last step, we round the real value solution y into an integer solution. For each $i \in [n]$, we set independently

$$z_i = \begin{cases} 1 & \text{w.p. (with probability) } y_i \\ 0 & \text{w.p. } 1 - y_i \end{cases} \tag{12.4}$$

Lemma 12.9. *Suppose $ay = b$ and for every $i \in [n]$, $|a_i| \le M$. Define z as in Equation 12.4. Then with probability at least $1 - 1/n^2$, by Hoeffding's inequality,*

$$a \cdot z \in [b - M\sqrt{2n \log n}, b + M\sqrt{2n \log n}]$$

In our setting, we set a such that $a \cdot y = R_i(y)$. Then, by the above lemma (setting $M = 1$), for every $i \in [n]$,

$$R_i(z) \in [R_i(y) - \sqrt{2n \log n}, R_i(y) + \sqrt{2n \log n}]$$

with probability at least $1 - 1/n$ by union bound.

Finally, we have

$$\begin{aligned}
\sum_{i\in[n]} z_i \cdot R_i(z) &\geq \sum_{i\in[n]} z_i \cdot (R_i(y) - \sqrt{2n\log n}) \\
&\geq \sum_{i\in[n]} z_i \cdot R_i(y) - n\sqrt{2n\log n} \\
&\geq \sum_{i\in[n]} y_i \cdot R_i(y) - n\sqrt{2n\log n} - n\sqrt{2n\log n} \\
&\qquad\qquad\qquad\qquad \text{(By applying Lemma 12.9 for } a_i = R_i(y)\text{)} \\
&\geq \mathsf{MAXCUT}(G) - \epsilon n^2 - 2n\sqrt{2n\log n} \qquad \text{(By Equation 12.3)}
\end{aligned}$$

12.3 Quantum product state approximation algorithm on dense graphs

Let H be a 2-local Hamiltonian on n qubits. The strategy to find a good product state approximation is roughly the same as the strategy to find an approximation for the classical MAXCUT problem.

Instead of doing a linear program, in the quantum case, we will solve some semi-definite programming problem in the form of

$$\begin{aligned}
\mathsf{sup}_{\rho=\sum_k p_k\rho_k^{(1)}\otimes\cdots\otimes\rho_k^{(n)}} \mathsf{Tr}(H\rho) =& \mathsf{sup}_{\rho=\sum_k p_k\rho_k^{(1)}\otimes\cdots\otimes\rho_k^{(n)}} \sum_{i,j}\sum_{P,Q} \alpha_{P,Q}^{(i,j)} \mathsf{Tr}(P\otimes Q\rho) \\
=& \mathsf{sup}_{\rho=\sum_k p_k\rho_k^{(1)}\otimes\cdots\otimes\rho_k^{(n)}} \sum_k p_k \sum_{i,P} \mathsf{Tr}(\rho_k^{(i)} P) \left(\sum_{j\in N(i)} \sum_Q \alpha_{P,Q}^{(i,j)} \mathsf{Tr}(Q\rho_k^{(j)}) \right) ,
\end{aligned}$$

where $H = \sum_{i,j} H_{i,j}$ and $H_{i,j} = \sum_{P,Q\in\{I,X,Y,Z\}} \alpha_{P,Q}^{(i,j)} P\otimes Q$.

Here roughly speaking, $\sum_{j\in N(i)} \sum_Q \alpha_{P,Q}^{(i,j)} \mathsf{Tr}(Q\rho_k^{(j)})$ can be viewed as a quantity similar to $R_i(x)$ in the previous section.