7.1 Derandomization of an Algorithm

Idea: Present a randomized algorithm that works with constant probability and only uses pairwise independent random variables. Then we can iterate through all possible choices of those pairwise independent random variables to find a deterministic choice that is guaranteed to succeed.

7.2 Maximal Independent Set Algorithm

7.2.1 Sequential Algorithm

Definition 7.1 For graph $G = (V, E)$, the independent set (IS) is some set of vertices $S \subset V$ such that $\forall (v, w) \in E, V \notin S$ or $w \notin S$.

Definition 7.2 The maximal independent set (MIS) $S$ of graph $G = (V, E)$ is the one that it is not a subset of any other independent sets. In other words, it satisfies: $\forall v \in V, v \in S \text{ or } N(v) \cap S \neq \emptyset$ ($N(v)$ is the neighborhood of $v$ in $G$).

Consider the following sequential algorithm, which may take $O(n)$ rounds:

\begin{algorithm}
\caption{Sequential Algorithm for Maximal Independent Set}
\begin{algorithmic}[1]
\State \textbf{input} : A graph $G = (V, E)$
\State \textbf{output}: A maximal independent set $I$
\State Initialization: $I = \emptyset$, $V' = V$;
\While{$V' \neq \emptyset$}
\State Choose any $v \in V'$;
\State Set $I \leftarrow I \cup \{v\}$;
\State Set $V' \leftarrow V' \setminus \{v\} \cup N(v)$
\EndWhile
\State Output $I$;
\end{algorithmic}
\end{algorithm}

7.2.2 Parallel Algorithm

[Luby '85] Parallel algorithm for MIS with $O(\log n)$ rounds and $\text{poly}(n)$ processes under CREW PRAM model (concurrent read & exclusive write).

Goal: Instead of adding single vertex to $I$ in each round, we add an independent set $S$ of $G'$ to $I$. If $S \cup N(S)$ is a constant fraction of $G'$, then we only need $O(\log n)$ rounds.
How to find \(S\)? Every vertex \(v \in G'\) adds itself to \(S\) with probability \(p(v)\) independently (or pairwise independently). To make sure that \(S\) is an independent set: For all edges \((v, w) \in E\), if \(v\) and \(w\) are in \(S\), then remove the lower degree vertex from \(S\).

This idea yields the following algorithm:

**Algorithm 2:** Parallel Algorithm for Maximal Independent Set

**input:** A graph \(G = (V, E)\)

**output:** A maximal independent set \(I\)

1. Initialization: \(I = \emptyset, G' = G, V' = V\);
2. while \(V' \neq \emptyset\) do
   3. Set \(S = \emptyset\);
   4. for each \(v \in V'\) do
      5. Add \(v\) to \(S\) with probability \(\frac{1}{2d_{G'}(v)}\), where \(d_{G'}(v)\) is the degree of \(v\) in \(G'\);
   6. for each edge \((v, w) \in G'\) do
      7. if \(v \in S, w \in S\) then
         8. Drop the lower degree vertex in \(\{v, w\}\) (If tie, pick a random one);
   9. Let \(S'\) be the remaining vertices;
   10. \(I \leftarrow I \cup S', V' \leftarrow V' \setminus \{S \cup N(S')\}, G' =\) induced subgraph on \(V'\);
11. Output \(I\);

**Lemma 7.3** Let \(G_j = (V_j, E_j)\) be the graph after round \(j\) and \(G_0 = G\). Then \(\mathbb{E}[|E_{j+1}| | E_j] < |E_j|(1 - \frac{1}{24})\).

**Corollary 7.4** With \(l = O(\log n), G_l = \emptyset\).

**Proof:** [Proof of corollary 7.4]

\[
\mathbb{E}[|E_j|] \leq |E_0|(1 - \frac{1}{12})^j
\]

\[
\leq m \exp(-\frac{j}{12})
\]

\[
\leq 1 \quad \text{for } j > 12 \log m \quad (7.1)
\]

Moreover,

\[
\Pr(E_j \neq \emptyset) = \Pr(E_j \geq 1)
\]

\[
\leq \mathbb{E}[|E_j|]
\]

\[
\leq \frac{1}{4} \quad \text{for } j > 48 \log m \quad (7.2)
\]

Thus, with probability at least \(3/4\), we have \(O(\log n)\) rounds. (This is an RNC algorithm for MIS.)

**Proof:** [Proof of Lemma 7.3]

For \(v \in V_j\), define \(H(v) = \{w \in N_{G_j}(v) : d_{G_j}(w) > d_{G_j}(v)\}\) and \(L(v) = \{w \in N_{G_j}(v) : d_{G_j}(w) \leq d_{G_j}(v)\}\), where \(N_{G_j}(v)\) denotes the neighborhood of \(v\) in the induced subgraph \(G'\).
We say that for any \( v \in V(G') \), \( v \) in BAD if \( |H(v)| \geq \frac{2}{3}d_{G'}(v) \) and GOOD if \( |L(v)| > \frac{1}{3}d_{G'}(v) \).

Further, we say edge \((v, w) \in E_j\) is BAD if \( |H(v)| \geq \frac{2}{3}d_{G'}(v) \) and GOOD if \( |L(v)| > \frac{1}{3}d_{G'}(v) \).

To prove that \( \Pr(w \text{ is deleted } | w \text{ is GOOD}) \geq \frac{1}{12} \), it is sufficient to show the following two claims (Note that \( w \) is deleted iff \( w \in S \cup N(S) \)):

**Claim 1** Let \( E_G \) be the GOOD edges. Then \( |E_G| \geq \frac{|E|}{2} \).

**Claim 2** \( \Pr(\text{edge } e \text{ is deleted } | e \text{ is GOOD}) \geq \frac{1}{12} \).

With these two claims,

\[
\mathbb{E}[|E_{j+1}| - |E_j|] = \sum_{e \in E_j} (1 - \Pr(e \text{ gets deleted})) \\
\leq |E_j| - \frac{1}{12}|E_G| \\
\leq |E_j|(1 - \frac{1}{12}) \tag{7.3}
\]

**Proof:** [Proof of Claim 1]

Let \( E_B \) be BAD edges of \( G_j \). We will define a mapping \( f : E_B \to \left( \frac{E_j}{2} \right) \) so that for all \( e_1 \neq e_2 \in E_B, f(e_1) \cap f(e_2) = \emptyset \). Thus, each \( e \in E_B \) has a distinct pair of edges in \( E_j \) and hence \( |E_B| \leq \frac{|E_j|}{2} \), which proves the claim.

The mapping \( f \) is defined with the following procedure:

For each \((v, w) \in E_j\), direct it from the lower degree endpoint to the higher degree one (choose arbitrary if tie).

Suppose \((v, w) \in E_B\) is directed as \( v \to w \). So \( d_{G'}(v) \leq d_{G'}(w) \). Since \((v, w) \in E_B\), both \( v \) and \( w \) are bad.

Since \( v \) is BAD, at least \( \frac{2}{3} \) of its neighbors are of degree \( \geq d_{G'}(v) \) and at most \( \frac{1}{3} \) of the edges incident to \( v \). Therefore, \( \geq 2 \) times as many out-edges from \( v \) as in-edges into \( v \).

Hence, for each BAD edge \( e \) directed into \( v \), there are a pair of out edges out of \( v \) that we can uniquely assign to \( e \).

**Proof:** [Proof of Claim 2]

We will show:

(1): \( \Pr(w \in S' | w \in S) \geq \frac{1}{2} \).

(2): \( \Pr(N(v) \cap S \neq \emptyset | v \text{ is GOOD}) \geq \frac{1}{6} \).
Proof of (1):
\[
Pr(w \in S' | w \in S) = Pr(H(w) \cap S \neq \emptyset | w \in S)
\leq \sum_{z \in H(w)} Pr(z \in S | w \in S) \quad \text{(Union Bound)}
\leq \sum_{z \in H(w)} \frac{Pr(z \in S, w \in S)}{Pr(w \in S)}
= \sum_{z \in H(w)} Pr(z \in S) \quad \text{(Pairwise Independence)}
= \sum_{z \in H(w)} \frac{1}{2d_G(z)}
\leq \sum_{z \in H(w)} \frac{1}{2d_G(w)}
\leq \frac{1}{2} \quad (7.4)
\]

Proof of (2):
\[
Pr(N(v) \cap S \neq \emptyset | v \text{ is GOOD}) = 1 - Pr(N(v) \cap S = \emptyset | v \text{ is GOOD})
= 1 - \prod_{z \in N(v)} Pr(z \notin S | v \text{ is GOOD}) \quad \text{(Mutual Independence)}
\geq 1 - \prod_{z \in L(v)} (1 - \frac{1}{2d_G'(z)})
\geq 1 - \prod_{z \in L(v)} (1 - \frac{1}{2d_G'(v)})
\geq 1 - \exp\left(-\frac{|L(v)|}{2d_G'(v)}\right)
\geq 1 - \exp(-\frac{1}{6})
\geq \frac{1}{6} \quad (7.5)
\]

With both (1) and (2), we have:
\[
Pr(v \in N_{G'}(S') | v \text{ is GOOD})
= Pr(N_{G'}(v) \cap S' \neq \emptyset | N(v) \cap S \neq \emptyset, v \text{ is GOOD}) Pr(N(v) \cap S \neq \emptyset | v \text{ is GOOD})
\geq \frac{1}{2} \cdot \frac{1}{6}
= \frac{1}{12} \quad (7.6)
\]

Since \(e = (v, w)\) is GOOD if there is at least one endpoint being GOOD, we have proved Claim 2.
7.2.3 Proof with Pairwise Independence

Instead of using mutual independence in (7.5) to obtain the lower bound on \( \Pr(N(v) \cap S \neq \emptyset | v \text{ is GOOD}) \), we can relax the condition with pairwise independence via the following lemma:

**Lemma 7.5** For pairwise independent random variables \( X_1, \ldots, X_l \in \{0, 1\} \) with \( \Pr(X_i = 1) = p_i \), we have

\[
\Pr\left( \sum_{i=1}^{l} X_i > 0 \right) \geq \frac{1}{2} \min\left\{ \frac{1}{2}, \sum_{i=1}^{l} p_i \right\}
\]

**Corollary 7.6** (Lower bound on \( \Pr(N(v) \cap S \neq \emptyset | v \text{ is GOOD}) \) with pairwise independence)

Let \( X_i = \begin{cases} 1, & \text{if } w_i \in S \\ 0, & \text{otherwise} \end{cases} \). We have

\[
\Pr(N_{G'}(v) \cap S \neq \emptyset | v \text{ is GOOD}) \geq \frac{1}{2} \min\left\{ \frac{1}{2}, \sum_{i=1}^{d_{G'}(v)} \frac{1}{2d_{G'}(v)} \right\} \\
\geq \frac{1}{2} \min\left\{ \frac{1}{2}, \sum_{i=1}^{d_{G'}(v)} \frac{|L(v)|}{2d_{G'}(v)} \right\} \\
\geq \frac{1}{12}
\]

**Proof:** [Proof of lemma 7.5]

Case 1: \( \sum_{i} p_i \leq 1 \)

\[
\Pr\left( \sum_{i=1}^{l} X_i > 0 \right) \geq \Pr\left( \sum_{i=1}^{l} X_i > 1 \right) \\
\geq \sum_{i=1}^{l} \Pr(X_i > 1) - \frac{1}{2} \sum_{i,j,i \neq j} \Pr(X_i = X_j = 1) \\
\geq \sum_{i=1}^{l} p_i - \frac{1}{2} \sum_{i \neq j} p_ip_j \\
= \sum_{i=1}^{l} p_i(1 - \frac{1}{2} \sum_{j \neq i} p_j) \\
\geq \sum_{i=1}^{l} p_i(1 - \frac{1}{2} \sum_{j=1}^{l} p_j) \\
\geq \frac{1}{2} \sum_{i=1}^{l} p_i
\]

(7.8)

Case 2: \( \sum_{i} p_i > 1 \)

We find always find some \( S \subset \{1, 2, \ldots, l\} \) such that \( \frac{1}{4} \leq \sum_{i \in S} p_i \leq 1 \), because either 1) \( p_i < \frac{1}{2} \) or 2) \( \exists j \) such that \( p_j \geq \frac{1}{2} \) (then pick \( S = \{j\} \)).
Similar to Case 1,

\[
\Pr(\sum_{i=1}^{l} X_i > 0) \geq \Pr(\sum_{i \in S} X_i > 1) \\
\geq \frac{1}{2} \sum_{i \in S} p_i \\
\geq \frac{1}{4}
\]

(7.9)

Hence, \(\Pr(\sum_{i=1}^{l} X_i > 0) \geq \frac{1}{2} \min\{\frac{1}{2}, \sum_{i=1}^{l} p_i\}\).