1 Another Hard Problem

- Let’s talk about a variant of the SAT problem.
- In this variant, we are given a CNF formula \( \phi \), and we want to find a truth assignment \( A \) that is “close” to satisfying.
- “Close” here means that \( A \) satisfies as many clauses of \( \phi \) as possible.
- Finding \( A \) that satisfies max number of clauses of \( \phi \) is called MAX-SAT.
- MAX-SAT models constraint satisfaction problems.
- Constraints are specified as logical formulas that must be true, e.g. “at most one job can run at time \( t \)”
- We may not be able to satisfy all constraints at once, but we can try to satisfy as many as possible.

Hardness?

- \textit{Decision problem}: does \( \phi \) have a truth assignment that satisfies at least \( m \) clauses?
- If we set \( m \) equal to \# clauses in \( \phi \), this decision problem is equivalent to SAT on CNF formulas.
- Hence, MAX-SAT is an NP-optimization problem.
- Problem remains hard even if \( \phi \) is a 3-CNF formula, since 3-SAT is still NP-complete.
- This variant is called MAX-3SAT.
- \textit{Interesting fact}: MAX-2SAT, in which formula is 2-CNF, is still an NP-optimization problem.
- (True even though 2-SAT is in \( P \))
2 Approximability on Average

Sometimes, it’s trivial to get a good approximation on average.

- Consider following algorithm RAND-ASSIGN for MAX-3SAT.
- Let $\phi$ be given, and let $X = \{x_1 \ldots x_n\}$ be list of variables in $\phi$.
- For each $1 \leq i \leq n$, flip a fair coin.
- If toss comes up heads, set $x_i = 1$; otherwise, set $x_i = 0$.
- (Hence, $\Pr(x_i = 1) = \Pr(x_i = 0) = \frac{1}{2}$).
- Return resulting random truth assignment.

How good is assignment obtained by RAND-ASSIGN?

- **Claim**: Suppose every clause of $\phi$ has literals over 3 distinct variables. (That is, a clause never contains the same literal twice, or two contradictory literals.) Then RAND-ASSIGN has an expected approx ratio of $8/7$.

What does “expected approx ratio” mean?

- Let $v^*$ be value of an opt solution.
- Suppose randomized algorithm $Q$ generates solution $y$ with probability $p(y)$, and $y$ has value $v_y$.
- Expected value of $Q$’s solution is then
  $$\bar{v}_Q = \sum_y p(y) \cdot v_y$$

- Expected approximation ratio for $Q$ is then $\frac{v^*}{\bar{v}_Q}$.

Cool. Let’s prove it.

- **Important note**: we are assuming that we never have the same variable appear twice in a clause! Easy to show that problem is still hard.

- **Pf**: What is average solution value for RAND-ASSIGN?
- Let $A$ be a truth assignment chosen by RAND-ASSIGN.
- For each clause $C_j$ of $\phi$, define indicator variable
  $$s_j = \begin{cases} 1 & \text{if } A \text{ satisfies } C_j \\ 0 & \text{otherwise} \end{cases}$$
- Consider $E[s_j]$, expected value of $s_j$ over choice of $A$. 


• Observe that
\[
E[s_j] = 1 \cdot \Pr(s_j = 1) + 0 \cdot \Pr(s_j = 0) = \Pr(s_j = 1) = 1 - \Pr(s_j = 0)
\]

• Now \(C_j\) contains three literals on distinct variables \(x_1 \neq x_2 \neq x_3\).

• Chance that a random assignment \(A\) sets each of \(x_1, x_2, x_3\) so that none of these 3 literals is true is \(1/2 \times 1/2 \times 1/2 = 1/8\).

• Hence,
\[
E[s_j] = 1 - \frac{1}{8} = \frac{7}{8}.
\]

• Now suppose \(\phi\) has \(m\) total clauses, and let \(S\) be number of clauses satisfied by a random truth assignment \(A\).

• Observe that \(S = \sum_{j=1}^{m} s_j\).

• By linearity of expectation, we have
\[
E[S] = \sum_{j=1}^{m} E[s_j] = \frac{7m}{8}
\]

• Conclude that \(v\) for RAND-ASSIGN is \(7m/8\).

• But \(v^* \leq m\).

• Conclude that \(v^*/v \leq 8/7\). QED

3 What Does Average Buy Us?

What good is an average approximation ratio?

• No guarantee that a single, randomly chosen assignment will achieve our ratio!

• How can we take advantage of this result?

• Option 1: apply Markov’s inequality:
\[
\Pr(x \geq a) \leq \frac{\bar{v}}{a}.
\]

• Let \(\phi\) be a 3-CNF formula with \(m\) clauses, each with 3 distinct vars.

• Let \(\bar{v}\) be expected number of clauses left unsatisfied by a random assignment.

• We have \(\bar{v} = m - \overline{v} = m/8\).
• Hence,
\[ \Pr(u \geq m/8 + 1) \leq \frac{m/8}{m/8 + 1} = \frac{m}{m + 8} \]

• Conclude that a randomly chosen assignment satisfies at least 7m/8 clauses with probability at least
\[ p = 1 - \frac{m}{m + 8}. \]

• Run z times and take the best result, and it will satisfy the desired number of clauses with prob \( 1 - (1 - p)^z \).

• (Works great for small m, maybe up to 100; for very large m, z gets huge!)

This is not a very satisfying result. (But there may be stronger tail inequalities that guarantee a good solution w/high prob after a small number of trials.)

• **Option 2:** derandomization!

• Because a random assignment satisfies at least 7m/8 clauses on average, *some* assignment \( A^* \) satisfies at least this many clauses.

• If we could find \( A^* \), we’d achieve stated approx ratio.

• Not always easy to find \( A^* \), but it *is* easy for this problem.

• Will use *method of conditional expectations*.

*Idea:* gradually fix truth values in \( A^* \), always ensuring that avg assignment to remaining vars satisfies enough clauses.

• As before, let \( S \) be number of satisfied clauses in \( \phi \).

• Observe that over random choice of truth assignment,
\[
E[S] = \Pr(x_1 = \text{true}) \cdot E[S|x_1 = \text{true}] + \Pr(x_1 = \text{false}) \cdot E[S|x_1 = \text{false}]
\]
\[
= 1/2 \cdot E[S|x_1 = \text{true}] + 1/2 \cdot E[S|x_1 = \text{false}].
\]

• Now first expectation is over all assignments \( A \) that set \( x_1 \) true, while second is over all assignments that set \( x_1 \) false.

• Avg of two expectations is 7m/8, so *one* of them is at least 7m/8.

• Compute the two conditional expectations above. If the first is bigger, set \( x_1 \) true; otherwise, set \( x_1 \) false.

• (Computations are easy by computing \( E[s_j|x_1 = v] \) for each clause \( C_j \) by itself and adding up results.)

• If we set \( x_1 = v \) for truth value \( v \), we know that there exists a truth assignment with \( x_1 = v \) that satisfies at least 7m/8 clauses!
• Repeat this process by observing that
\[ E[S|x_1 = v] = \frac{1}{2} \cdot E[S|(x_1 = v) \land (x_2 = \text{true})] + \frac{1}{2} \cdot E[S|(x_1 = v) \land (x_2 = \text{false})], \]
computing two conditional expectations, and selecting value for \( x_2 \) to match larger expectation.

• Continue in this way until all \( x_i \) are set.

• Will end up with one assignment \( A^* \) that does at least as well as average!

4 Best Known Results for MAX-SAT

• MAX-\( k \)-SAT for any fixed \( k \) (i.e. each clause has exactly \( k \) distinct literals) is approximable to within \( 1/(1 - 2^{-k}) \), by easy generalization of our MAX-3SAT result to arbitrary \( k \). This is known as Johnson’s algorithm (1974).

• If each clause may contain at most 3 literals, can still achieve guaranteed 8/7 ratio for satisfiable formulas (Karloff and Zwick 1997) and a ratio of 1.249 for any formula, satisfiable or not (Trevisan, Sorkin, Sudan, and Williamson 1996).

• MAX-\( k \)-SAT is not approximable to within better than \( 1/(1 - 2^{-k}) \) unless P = NP (Hastad 1997).

• For MAX-SAT with no restriction on number of literals per clause, best known ratio is 1.2987 (Asano, Hori, Ono, and Hirata 1997). Known not to have a PTAS (Papadimitriou and Yannakakis 1991).