Analysis of Algorithms and Data Structures: Problem set 8

Giosue Castellano

Problem 8

Given the 2-CNF-SAT problem, we consider a randomized algorithm that works as follows: start with an arbitrary assignment of all variables set to false, and as long as there exists an unsatisfied clause, uniformly at random pick one of the two variables in that clause and flip its value. We assert that if a satisfying assignment exists, the algorithm will find it in $O(n^2)$ expected time, where n is the number of variables.

We want to show that if there is a satisfying assignment τ^*, then this algorithm reaches a satisfying assignment in $O(n^2)$ expected time.

Let τ^* be the starting assignment and τ_i the assignment after i iterations. We define the "Hamming distance" d_i between τ_i and τ^* as the number of coordinates in which the two differ, where $d_i \in \{0, 1, \ldots, n\}$.

Observation: The Hamming distance at the next step d_{i+1} can only be $d_i - 1$ or $d_i + 1$, since we flip exactly one variable at each iteration for any given i and this flipped variable can either be now flipped from incorrect to correct or the reverse way for assignment τ^*.

Claim: $\mathcal{P}[d_{i+1} = d_i - 1] \geq \frac{1}{2}$.

Proof of Claim: Consider iteration i which examines an unsatisfied clause $C = (x \vee y)$. The satisfying assignment τ^* must satisfy at least one of x or y, while τ_i satisfies neither. Hence, when we pick a literal randomly, with probability at least $\frac{1}{2}$, we choose the one where τ_i and τ^* differ, reducing d_i by 1.

Let us prove that we can find a correct assignment using this Claim. For $k > \ell$, let $T_{k,\ell}$ be the expected number of iterations it takes to reach Hamming distance ℓ for the first time when starting from distance k. We have the recurrence relation for the expected number of steps to reduce the Hamming distance from $i + 1$ to i as follows:

$$T_{i+1,i} \leq \frac{1}{2} \cdot (1) + \frac{1}{2} \cdot (1 + T_{i+2,i})$$

Simplifying this using the linearity of expectation, we obtain:

$$T_{i+1,i} \leq \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot (1 + T_{i+2,i+1} + T_{i+1,i})$$

By isolating $T_{i+1,i}$ on one side, we find:

$$T_{i+1,i} \leq 2 + T_{i+2,i+1}$$

Applying this simplification recursively and considering $T_{n,n-1} = 1$ (since flipping the one differing variable will achieve the satisfying assignment), we establish the following sequence of inequalities:

$$T_{i+1,i} \leq 2 + T_{i+2,i+1} \leq 4 + T_{i+3,i+2} \leq \ldots \leq O(n) + T_{n,n-1} \leq O(n)$$

Summing up the expected times to reach each consecutive Hamming distance from $n-1$ to 0, we conclude that:

$$T_{n,0} \leq T_{n,n-1} + T_{n-1,n-2} + \ldots + T_{1,0} = O(n^2)$$

This completes the proof that the algorithm reaches a satisfying assignment in $O(n^2)$ expected time if such an assignment τ^* exists.