7 Appendix

7.1 Proof of Proposition 1

Proof. Suppose Algorithm 1 outputs "$|S| \geq s\lfloor 2^{j-3} \rfloor$" but $|S| < s\lfloor 2^{j-3} \rfloor$. We will show that this happens with probability at most Δ. Let the iteration with $m = i$ be the final iteration where a break would have resulted in a correct output, i.e., $i = \arg\max_{i'} s\lfloor 2^{i'-3} \rfloor \leq |S|$. This means that $s\lfloor 2^{i-3} \rfloor \leq |S| < s\lfloor 2^{i-2} \rfloor$. Also note that $i \geq 2$ because of the floor operator. The algorithm outputs an incorrect bound if and only if the while-loop on m breaks with $m = j$ such that $j \geq i + 1$. For this to happen, the while loop would *not* have been broken in all iterations with $m \leq i$. In particular, we would have observed $\sum_{t=1}^{T} w^t \geq sT/2$ on iteration i. However, this is an unlikely event, as we now show. Observe that $S(h^i) = S \cap (h^i)^{-1}(b)$ by definition, and

$$\mathbf{E}[|S(h^i)|] = \frac{|S|}{2^i} < \frac{s\lfloor 2^{i-2} \rfloor}{2^i} \leq \frac{s}{4}.$$

This results in the inequalities

$$\begin{aligned} \mathbf{E}[w^t] &= \mathbf{E}\left[\min\left\{s, |S(h^i)|\right\}\right] \\ &\leq \min\left\{\mathbf{E}\left[s\right], \mathbf{E}\left[|S(h^i)|\right]\right\} \leq s/4. \end{aligned} \tag{2}$$

Since $w^t \in [0, s]$, we can apply Hoeffding's inequality and use Equation 2 to obtain

$$\begin{aligned} \Pr\left[\frac{1}{T}\sum_{t=1}^{T} w^t \geq \frac{s}{2}\right] &\leq \exp\left(-\frac{2T}{s^2}\left(\frac{s}{2} - \frac{s}{4}\right)^2\right) \\ &= \exp\left(-\frac{T}{8}\right). \end{aligned}$$

Setting $T = \lceil 8 \ln \frac{1}{\Delta} \rceil$, we have $\exp\left(-\frac{T}{8}\right) \leq \Delta$. Therefore, the probability of observing $\sum_{t=1}^{T} w^t \geq sT/2$ in iteration i (making the output of Algorithm 1 incorrect) is bounded above by Δ. $\square$

7.2 Proof of Proposition 2

Proof. Suppose Algorithm 2 outputs "$|S| \geq s\lfloor 2^{j-3} \rfloor$" but $|S| < s\lfloor 2^{j-3} \rfloor$. We will show that this happens with probability at most Δ. Let the iteration with $m = i$ be the final iteration where a break would have resulted in a correct output, i.e., $i = \arg\max_{i'} s\lfloor 2^{i'-3} \rfloor \leq |S|$. This means that $s\lfloor 2^{i-3} \rfloor \leq |S| < s\lfloor 2^{i-2} \rfloor$. Also note that $i \geq 2$ because of the floor operator.

The algorithm outputs an incorrect bound if and only if the while-loop on m breaks with $m = j$ such that $j \geq i + 1$. For this to happen, the while loop would *not* have been broken in all iterations with $m \leq i$. In particular, we would have observed $\sum_{t=1}^{T} w^t \geq sT/2$ on iteration i. However, this is an unlikely event, as we now show. Observe that $S(h^i) = S \cap (h^i)^{-1}(b)$ by definition, and

$$\mathbf{E}[|S(h^i)|] = \frac{|S|}{2^i} < \frac{s\lfloor 2^{i-2} \rfloor}{2^i} \leq \frac{s}{4}.$$

This property holds because b is chosen uniformly at random on line 8 of Algorithm 2. Crucially, this property holds regardless of how the matrices A_m are constructed on line 7.

This results in the inequalities

$$\begin{aligned} &\mathbf{E}[w_k] = \mathbf{E}\left[\min\left\{sK, |S(h^{j-1})|\right\}\right] \leq s/4 \\ &\mathbf{E}\left[\frac{1}{K}\sum_{k=1}^{K} w_k\right] \leq s/4 \\ &\mathbf{E}[w^t] = \mathbf{E}\left[\min\left\{s, \frac{1}{K}\sum_{k=1}^{K} w_k\right\}\right] \leq s/4 \end{aligned} \tag{3}$$

Since $w^t \in [0, s]$, we can apply Hoeffding's inequality and use Equation 3 to obtain

$$\begin{aligned} \Pr\left[\frac{1}{T}\sum_{t=1}^{T} w^t \geq \frac{s}{2}\right] &\leq \exp\left(-\frac{2T}{s^2}\left(\frac{s}{2} - \frac{s}{4}\right)^2\right) \\ &= \exp\left(-\frac{T}{8}\right). \end{aligned}$$

Setting $T = \lceil 8 \ln \frac{1}{\Delta} \rceil$, we have $\exp\left(-\frac{T}{8}\right) \leq \Delta$. Therefore, the probability of observing $\sum_{t=1}^{T} w^t \geq sT/2$ in iteration i (making the output of Algorithm 1 incorrect) is bounded above by Δ. $\square$

7.3 Upper Bound

Proof. Suppose Algorithm 4 outputs "$|S| \leq s2^{j+1}$", but this is incorrect, and $s2^{j+2} \geq |S| > s2^{j+1}$. That is, the output is the largest invalid upper bound. We will show that Algorithm 1 outputs this, or any other smaller invalid bound, with probability at most Δ. For the algorithm to output the smallest valid upper bound, 2^{j+2}, the iteration with $m = j + 2$ would have resulted in breaking the while-loop on m. Thus, in *every* prior iteration $i \leq j+1$, we would have observed $\sum_{t=1}^{T} w^t \geq sT/2$. We will use the union bound to upper bound the probability of observing $\sum_{t=1}^{T} w^t < sT/2$ for *some* $i \leq j + 1$.

Fix any $i \leq j + 1$. Then, $\mathbf{E}[|S(h^i)|] = \mu_i = |S|/2^i = 2^{j-i}\,|S|/2^j > s2^{j-i+1}$ by our assumption. Let the variance be $\mathbf{Var}[|S(h^i)|] = \sigma_i^2$. We first observe that the $\min$ operation with sK on line 10 of Algorithm 1 serves only an optimization purpose, and does not alter the outcome of the algorithm (because of the subsequent $\min$ operation when computing w^t). Thus, for the sake of analysis, we can let $w_k = |S(h^i)|$ without loss of generality.

For brevity of notation, let $\overline{w}_K = \frac{1}{K}\sum_{k=1}^{K} w_k$. Then, $\mathbf{E}[\overline{w}_K] = \mathbf{E}[|S(h^i)|] > s2^{j-i+1}$ and $\mathbf{Var}[\overline{w}_K] \leq \sigma_i^2/K$. Applying Cantelli's inequality:

$$\begin{aligned}\Pr[\overline{w}_K \leq s] &= \Pr\left[\overline{w}_K \leq \mathbf{E}\left[\overline{w}_K\right] - \left(\mathbf{E}\left[\overline{w}_K\right] - s\right)\right] \\ &\leq \frac{\sigma_i^2/K}{\sigma_i^2/K + s^2(2^{j-i+1}-1)^2} \\ &\leq \frac{\sigma_i^2/K}{\sigma_i^2/K + s^2 4^{j-i}}\end{aligned}$$

Hence, $\Pr[\overline{w}_K \geq s] \geq \frac{s^2 4^{j-i}}{\sigma_i^2/K + s^2 4^{j-i}}$. Since $w^t = \min\{s, \overline{w}_K\}$, we also have $\Pr[w^t \geq s] \geq \frac{s^2 4^{j-i}}{\sigma_i^2/K + s^2 4^{j-i}}$.

Let y^t denote a 0-1 indicator variable that is 1 when $w^t \geq s$. Then $y^t \leq w^t$ and $\mathbf{E}[y^t] \geq \frac{s^2 4^{j-i}}{\sigma_i^2/K + s^2 4^{j-i}}$. By a precondition of the theorem, $s^2 4^{j-i} \geq \mu_i^2/16 > \sigma_i^2/K$, which implies $\mathbf{E}[y^t] > 1/2$, making it unlikely to observe the sum of T_i such y^t variables to be smaller than $T_i/2$. We thus have:

$$\begin{aligned}&\Pr\left[\sum_{t=1}^{T_i} w^t < \frac{sT_i}{2}\right] \leq \Pr\left[\sum_{t=1}^{T_i} y^t < \frac{T_i}{2}\right] \\ &\leq \exp\left(-\frac{2}{T_i}\left(\mathbf{E}\left[\sum_{t=1}^{T_i} y^t\right] - \frac{T_i}{2}\right)^2\right) \\ &\leq \exp\left(-\frac{2}{T_i}\left(\frac{s^2 4^{j-i} T_i}{\sigma_i^2/K + s^2 4^{j-i}} - \frac{T_i}{2}\right)^2\right) \\ &= \exp\left(-\frac{T_i}{2}\left(\frac{s^2 4^{j-i} - \sigma_i^2/K}{s^2 4^{j-i} + \sigma_i^2/K}\right)^2\right) \\ &\leq \exp\left(-\frac{T_i}{2}\left(\frac{\mu_i^2/16 - \sigma_i^2/K}{\mu_i^2/16 + \sigma_i^2/K}\right)^2\right) \\ &= \exp\left(-\frac{T_i}{2}\left(\frac{1 - 16\gamma_i^2/K}{1 + 16\gamma_i^2/K}\right)^2\right),\end{aligned}$$

where the second inequality follows from Hoeffding's inequality and the last inequality follows because $s^2 4^{j-i} \geq \mu_i^2/16$. This expression is at most Δ/n because T_i is set to $\left\lceil 2\left(\frac{1+16\gamma_i^2/K}{1-16\gamma_i^2/K}\right)^2 \ln\frac{n}{\Delta}\right\rceil$ in line 4 of Algorithm 4. Applying the union bound over all $i \leq j+1$, the probability of observing $\sum_{t=1}^{T_i} w^t < sT_i/2$ in *any* iteration $i \leq j+1$, and thus possibly outputting an incorrect upper bound, is bounded above by Δ. $\square$

When the linear search in Algorithm 4 is replaced with more efficient search procedures, the definition of T_i can be modified to achieve the desired probability of correctness.

Algorithm 4 Upper Bound with Variance Reduction

Inputs: K: Number of repetitions per trial
s: Solution cutoff
Δ: Failure probability
$\mathcal{O}_S$: A SAT oracle
$\{\mathcal{A}^m\}_{m=1}^n$: For each $m \in [1, n]$, a distribution over parity matrices with known variance bounds that satisfy $16\sigma_m^2 < K\mu_m^2$, where $\mathbf{Var}[|S(h^m)|] \leq \sigma_m^2$ and $\mu_m = \mathbf{E}[|S(h^m)|]$

Output: A probabilistic upper bound on $|S|$

1: $m = 1$
2: **while** $m \leq n$ **do**
3: $\quad\gamma_m^2 = \sigma_m^2/\mu_m^2$
4: $\quad T_m = \left\lceil 2\left(\frac{1+16\gamma_m^2/K}{1-16\gamma_m^2/K}\right)^2 \ln\frac{n}{\Delta}\right\rceil$
5: $\quad$**for** $t = 1, \cdots, T$ **do**
6: $\quad\quad$**for** $k = 1, \cdots, K$ **do**
7: $\quad\quad\quad$Sample $A^m \sim \mathcal{A}^m$, denote $h^m(x) = A^m x$
8: $\quad\quad\quad$Sample $b \sim \mathrm{Uniform}(\mathbb{F}_2^m)$
9: $\quad\quad\quad w_k \leftarrow \min\left\{sK, |S \cap (h^m)^{-1}(b)|\right\}$ {Invoke oracle $\mathcal{O}_S$ up to sK times to check whether the input formula with additional constraints $A^m x = b$ has at least sK distinct solutions}
10: $\quad\quad w^t \leftarrow \min\left\{s, \frac{1}{K}\sum_{k=1}^{K} w_k\right\}$
11: $\quad$**if** $\sum_{t=1}^{T} w^t < sT/2$ **then**
12: $\quad\quad$break
13: $\quad m = m + 1$
14: Output "$|S| \leq s2^{m+1}$"
