Balanced 2-coloring

Prove this claim:

Claim 1 Let \( \mathbf{a} \) be a \( 1 \times n \) row vector with \( a_i \in \{0, 1\} \) and \( \mathbf{x} \) be a \( n \times 1 \) column vector with \( x_i \in \{\pm 1\} \) for \( i \in [n] \). Then

\[
\Pr \left[ \mathbf{a} \mathbf{x} > 4\sqrt{n \ln n} \right] < \frac{1}{n^2}. \tag{1}
\]

1 What’s Wrong with the Slides \langle ra20040407.ppt \rangle?

Let’s rewrite a statement on Page 17 of the slides \langle ra20040407.ppt \rangle\(^{‡}\) as the following theorem:

Theorem 1 Given \( \mathbf{a} \in \{0, 1\}^n \), then for any \( \mathbf{x} \in \{-1, +1\}^n \) the following inequality

\[
\Pr \left[ \mathbf{a} \mathbf{x} > 4\sqrt{n \ln n} \right] \leq \Pr \left[ |\mathbf{x}| > 4\sqrt{n \ln n} \right] \tag{hu}
\]

always holds, where \( \mathbf{a} \mathbf{x} \overset{\Delta}{=} \sum_{i=1}^{n} a_i x_i \) and \( |\mathbf{x}| \overset{\Delta}{=} \sum_{i=1}^{n} x_i \).

However, the goal of this section is to disprove Theorem 1:

1. Let the number of \(-1\)'s in \( \mathbf{x} \) as \( k \in [n] \), it’s clear that \( |\mathbf{x}| = n - 2k \).

2. Define event \( \mathcal{S} : |\mathbf{x}| > 4\sqrt{n \ln n} \).

If \( n - 2k \) is less than \( 4\sqrt{n \ln n} \) (that is, \( k < \frac{n - 4\sqrt{n \ln n}}{2} \)), then \( \mathcal{S} \) always happens.

\(^{*}\) \( x \in \mathbb{R} \): \( x \) is randomly chosen from \( X \) under uniform distribution.

\(^{†}\) \([n] \overset{\Delta}{=} \{1, 2, \ldots, n\} \).

\(^{‡}\) See \( \langle \text{http://www.iis.sinica.edu.tw/~hil/random/ra20040407.ppt} \rangle \).
3. For analyzing when event \( s : ax > 4\sqrt{n \ln n} \) always happens, it's convenient to define some actions and symbols:

- Compute the inner product \( ax \) of vector \( a \in \{0, 1\}^n \) and vector \( x \in (-1, +1)^n \). We find if \( a_i = 0 \) for \( i \in [n] \), then \( a_i x_i = 0 \); call \( x_i \) is disabled by \( a_i \).
- \( \ell (\in [n]) \): the number of \( x_i \)'s disabled by some \( a_i = 0 \) of \( a \).
- \( y \): the vector after removing disabled \( x_i \)'s of \( x \).
  So \( y \in \{-1, +1\}^{n - \ell} \).
- Similarly, we also have \( k' \), the number of \(-1\)'s in \( y \).
  When \( k' < \frac{(n-\ell)-4\sqrt{n \ln n}}{2} \), the event \( s \) always holds.

4. Follow the above settings, we have

\[
\text{Pr}[S] = \sum_{i=0}^{k} \binom{n}{i},
\]
(2)

\[
\text{Pr}[s] = \sum_{i=0}^{k'} \binom{n-\ell}{i} = \frac{2^\ell \sum_{i=0}^{k'} \binom{n-\ell}{i}}{2^n},
\]
(3)

and

\[
\frac{\text{Pr}[s]}{\text{Pr}[S]} = \frac{3}{2} = \frac{2^\ell \sum_{i=0}^{k'} \binom{n-\ell}{i}}{\sum_{i=0}^{k} \binom{n}{i}}.
\]
(4)

5. In the case \( \ell = 1 \) and \( \frac{n-4\sqrt{n \ln n}}{2} - \lfloor \frac{n-4\sqrt{n \ln n}}{2} \rfloor > \frac{1}{2} \),
we can set \( k = \lfloor \frac{n-4\sqrt{n \ln n}}{2} \rfloor - 1 < \frac{n-4\sqrt{n \ln n}}{2} \),
and \( k' = [\frac{n-1-4\sqrt{n \ln n}}{2}] - 1 = k < \lfloor \frac{n}{2} \rfloor \).
Thus (4) becomes

\[
2 \sum_{i=0}^{k} \binom{n-1}{i} \frac{1}{\sum_{i=0}^{k} \binom{n}{i}} > 2 \frac{1}{2} = 1
\]
(5)

because

\[
\binom{n-1}{i} = \frac{n-i}{n} \binom{n}{i} > 1 \frac{1}{2} \binom{n}{i} \Rightarrow \sum_{i=0}^{k} \binom{n-1}{i} > 1 \sum_{i=0}^{k} \binom{n}{i}
\]
\[
\Leftrightarrow \frac{\sum_{i=0}^{k} \binom{n-1}{i}}{\sum_{i=0}^{k} \binom{n}{i}} > \frac{1}{2}, \text{ for } i \in \{0, 1, \ldots, k\}.
\]
(6)
The result (5) is in contradict with Theorem 1 because

\[
\frac{\Pr[S]}{\Pr[S]} = \frac{\Pr[ax > 4\sqrt{n \ln n}]}{\Pr[|x| > 4\sqrt{n \ln n}]} > 1
\]

\[\iff \Pr[ax > 4\sqrt{n \ln n}] > \Pr[|x| > 4\sqrt{n \ln n}]. \tag{7}\]

Therefore, we complete the disproof of Theorem 1.

2 A Proof of Theorem 1

Omitted. Because the Chernoff inequality given in Page 14 of the slides ⟨ra20040407.ppt⟩ is too considerable for us to do any heavy job. However, it’s dangerous to use the result of Theorem 1.