LECTURE 20

Last time

- Testing linearity
- Tolerant testing and distance approximation

Today

- Approximating the distance to sortedness (length of LIS) of 0/1 sequences

Thank you for signing up to grade HW 4

Approximating Distance to Monotonicity for 0/1 Sequences

Input: Parameter $\varepsilon \in (0,1/2]$ and

a list of n zeros and ones (equivalently, $f: [n] \to \{0,1\}$)

Question: How far is this list to being sorted?

(Equivalently, how far is f from monotone?)

$\text{dist}(f, MONO)$ =distance from f to monotone

$\text{Dist}(f, MONO) = n \cdot \text{dist}(f, MONO)$

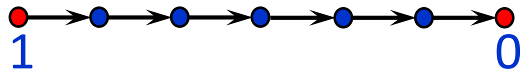
Note: $\text{Dist}(f, MONO) = n - |LIS|$,
where LIS is the longest increasing subsequence

Output: An estimate $\hat{\varepsilon}$ such that w.p. $\geq \frac{2}{3}$

$$|\hat{\varepsilon} - \text{dist}(f, MONO)| \leq \varepsilon$$

Today: can answer in $O\left(\frac{1}{\varepsilon^2}\right)$ time [Berman Raskhodnikova Yaroslavtsev]

Distance to Monotonicity over POset Domains

- Let f be a function over a partially ordered domain D.
- Violated pair: 

- The violation graph G_f is a directed graph with vertex set D whose edge set is the set of pairs (x, y) violated by f.
- VC_f is a minimum vertex cover of G_f
- MM_f is a maximum matching in G_f

Characterization of $Dist(f, \mathrm{Mono})$ for $f: D \to \{0,1\}$ [FLNRRS 02]

$$Dist(f, \mathrm{Mono}) = \left|\mathrm{MM}_f\right| = |VC_f|$$

Distance to Monotonicity for 0/1 Sequences

- Let $f : [n] \to \{0,1\}$
- Great notation switch: $g_i = (-1)^{f(i)}$ for $i \in [n]$
- Cumulative sums: $s_0 = 0$ and $s_i = s_{i-1} + g_i$ for $i \in [n]$
- Final sum: $s_f = s_n$
- Maximum sum: $m_f = \max_{i=0}^{n} s_i$

dist$(f, Mono)$ for $f : [n] \to \{0,1\}$ [Berman Raskhodnikova Yaroslavtsev]

$$Dist(f, \mathrm{Mono}) = \frac{n - 2m_f + s_f}{2}$$

Proof:

1. Construct a matching of that size
2. Construct a vertex cover of that size.

Distance to Monotonicity for 0/1 Sequences

Characterization dist$(f, Mono)$ for $f: [n] \to \{0,1\}$

$$Dist(f, \mathrm{Mono}) = \frac{n - 2m_f + s_f}{2}$$

Proof: (1) Construct a matching that leaves $2m_f - s_f$ nodes unmatched

Distance to Monotonicity for 0/1 Sequences

Characterization dist$(f, Mono)$ for $f: [n] \to \{0,1\}$

$$Dist(f, \text{Mono}) = \frac{n - 2m_f + s_f}{2}$$

Proof: (2) Construct a vertex cover.

Distance to Monotonicity: Algorithm

Algorithm (**Input:** ε, n; query $acess\ to\ f: [n] \to \{0,1\}$)

1. Sample a random subset $\boldsymbol{S} \subset [n]$
   where each element is included w.p. s/n independently
2. Let $\tilde{f} = f_{|S}$
3. Compute $\tilde{\varepsilon} = Dist(\tilde{f}, Mono)/s$
4. **Return** $\tilde{\varepsilon}$

- Let $\varepsilon_f = dist(f, Mono) = Dist(f, Mono)/n$

Theorem

$$\varepsilon_f - \sqrt{2\varepsilon_f/s} \leq \mathbb{E}[\tilde{\varepsilon}] \leq \varepsilon_f$$

$$\mathrm{Var}[\tilde{\varepsilon}] = O(\varepsilon_f/s)$$

Proof idea: Let $Z(S) = Dist(\tilde{f}, Mono)$

We'll define random variables $X(S)$ and $\mathrm{Y}(S)$, such that $X(S) \leq Z(S) \leq Y(S)$

$X(S)$ will be in terms of matching MM_f; $Y(S)$ in terms of vertex cover VC_f

Upper Bound on $Z(S)$

- Define $Y(S) = |VC_f \cap S|$

Upper Bound Lemma

(a) $Z(S) \le Y(S)$, (b) $\mathbb{E}[Y(S)] = \varepsilon_f \cdot s$ and $\mathrm{Var}[Y(S)] \le \varepsilon_f \cdot s$

Proof: (a) $Z(S) = Dist(\tilde{f}, Mono) = \left|\mathrm{VC}_{\tilde{f}}\right|$

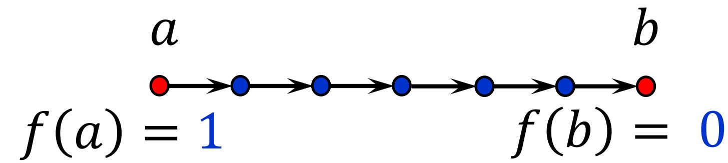
- Each pair violated by $\tilde{f}$ is also violated by f
- $VC_f \cap S$ is a vertex cover (not necessarily minimum) of $G_{\tilde{f}}$

$$Z(S) = Dist(\tilde{f}, Mono) = \left|\mathrm{VC}_{\tilde{f}}\right| \le \left|VC_f \cap S\right| = Y(S)$$

(b) Recall that $\left|VC_f\right| = \varepsilon_f \cdot n$

- Each element of VC_f appears in S independently w.p. s/n
- $Y(S)$ is binomial with mean $\left|VC_f\right| \cdot \frac{s}{n} = \varepsilon_f \cdot s$ and variance $\left|VC_f\right| \cdot \frac{s}{n}\left(1 - \frac{s}{n}\right) \le \varepsilon_f \cdot s$

Lower Bound on $Z(S)$



- Let $\ell = |MM_f| = \varepsilon_f \cdot n$
- MM_f consist of ℓ pairs of the form (a, b)
- Let $a_1 < a_2 < \cdots < a_\ell$ be the lower endpoints of pairs in MM_f
- Let $b_1 < b_2 < \cdots < b_\ell$ be the upper endpoints of pairs in MM_f
- Then $a_i < b_i$ for all $i \in [\ell]$
- Guaranteed edges are pairs of the form (a_i, b_j) where $i \le j$

$f(a_i) = 1$

$f(b_i) = 0$

- Let $\widetilde{MM}(S)$ denote a maximum matching that consists of guaranteed edges
- Define $X(S) = |\widetilde{MM}(S)|$

Lower Bound Lemma

(a) $X(S) \le Z(S)$, (b) $\mathbb{E}[X(S)] \ge \varepsilon_f \cdot s - \sqrt{4.5\varepsilon_f \cdot s}$ and $\mathrm{Var}[X(S)] = O(\varepsilon_f \cdot s)$

Proof of Lower Bound Lemma: Random Walk

- Recall: $X(S) = |\widetilde{MM}(S)|$
- Let $X'(S) = |V\left(MM_f\right) \cap S|$
- $U(S) =$ number of element of $V\left(MM_f\right) \cap S$ left unmatched by $\widetilde{MM}(S)$
- Then $X(S) = \frac{X'(S)-U(S)}{2}$
- $X'(S)$ is binomial with mean $2\varepsilon_f \cdot s$ and variance $\leq 2\varepsilon_f \cdot s$
- To understand $U(S)$ define a random walk that at step $i \in [\ell]$ moves by

$$g_i = \begin{cases} 1 & \text{if } \{a_i, b_i\} \cap S = \{b_i\} \\ -1 & \text{if } \{a_i, b_i\} \cap S = \{a_i\} \\ 0 & \text{otherwise} \end{cases}$$

- Define $m(S) =$ the maximum value reached by the walk
- Define $p(S) =$ the final position reached by the walk

Claim

$U(S) \leq 2m(S) - p(S)$

Analyzing $2m(S) - p(S)$

Claim 2

$\Pr[m(S) \geq z] \leq \Pr[|p(S)| \geq z]$ for all $z \in [\ell]$

Analyzing the Expectation and Variance of $U(S)$

Claim

$U(S) \leq 2m(S) - p(S)$

Claim 2

$\Pr[m(S) \geq z] \leq \Pr[|p(S)| \geq z]$ for all $z \in [\ell]$

$$\mathbb{E}[U] \leq \mathbb{E}[2m(S) + |p(S)|] \leq 3\mathbb{E}[|p(S)|]$$

Completing the Analysis

Lower Bound Lemma

(a) $X(S) \leq Z(S)$, (b) $\mathbb{E}[X(S)] \geq \varepsilon_f \cdot s - \sqrt{4.5\varepsilon_f \cdot s}$ and $\mathrm{Var}[X(S)] = O(\varepsilon_f \cdot s)$

Upper Bound Lemma

(a) $Z(S) \leq Y(S)$, (b) $\mathbb{E}[Y(S)] = \varepsilon_f \cdot s$ and $\mathrm{Var}[Y(S)] \leq \varepsilon_f \cdot s$

Theorem

$$\varepsilon_f - \sqrt{2\varepsilon_f / s} \leq \mathbb{E}[\tilde{\varepsilon}] \leq \varepsilon_f$$

$$\mathrm{Var}[\tilde{\varepsilon}] = O(\varepsilon_f / s)$$