12 Randomized Communication Complexity (April 8–10)

In the previous lectures, we considered the minimum number of bits that must be *deterministically* transmitted between two parties to compute a function of their inputs *with certainty*. Now we'll introduce both randomness and potential error into the picture.

12.1 Internal Randomization

A *randomized communication protocol* is a probability distribution over a set of possible deterministic protocols. The definition is exactly analogous to randomized decision trees, which is not surprising, since we can model any deterministic protocol as a pair of decision trees, one for Alice and one for Bob. The *complexity* of a randomized protocol is the expected number of transmitted bits. As usual, we require that the last bit transmitted is the output $f(x, y)$ and that both players *know* when this bit is transmitted, even though Alice and Bob do *not* know each other's random bits.

A randomized protocol R for a function f has *one-sided error* if the following probability bound holds for all inputs x and y:

$$\Pr[R(x,y) = 0] = 1 \qquad \text{if } f(x,y) = 0$$
$$\Pr[R(x,y) = 0] \leq \frac{1}{2} \qquad \text{if } f(x,y) = 1$$

In other words, the protocol sometimes randomly give false negatives, but it never returns a false positive.

The power of this model is illustrated by the following remarkable result.

Theorem 1. *The randomized communication complexity of the n-bit inequality function with one-sided error is $O(\log n)$.*

Proof: In fact, the $O(\log n)$ bound holds for the following randomized *one-way* protocol. Alice and Bob are given positive integers x and y, each less than 2^n. To begin the protocol, Alice chooses p randomly from the first $2n$ prime numbers, and then transmits both p and $x \bmod p$ to Bob. If $x \bmod p \neq y \bmod p$, then Bob (correctly!) replies with 1; otherwise, Bob replies with 0.

Now suppose $x \neq y$. Let $p_1, p_2, \ldots, p_k$ be distinct primes such that $x \bmod p_i = y \bmod p_i$ for all i. Then $x \bmod P = y \bmod P$, where $P = p_1 p_2 \cdots p_k$. We easily observe that $P > 2^k$. This implies that $k < n$, since otherwise, $x \bmod P = x$ and $y \bmod P = y$. It follows that $x \bmod p \neq y \bmod p$ for at most n distinct primes, which means the protocol returns an incorrect answer with probability at most $1/2$. □

12.2 External Randomization

Alternately, we can consider the situation where the protocol is deterministic but the *input* is random. To keep things simple, let's assume that Alice's input x and Bob's input y are independently and uniformly distributed over the set $\{0,1\}^n$. Now the *distributional complexity*[1] of a deterministic protocol P is

$$\frac{1}{2^{2n}} \sum_{x,y} C(P, (x,y)),$$

[1]This is not an entirely standard term, but rather one I chose on the fly to distinguish this concept from the worst-case expected complexity of a randomized algorithm with deterministic inputs.

where $C(P,(x,y))$ is the number of bits transmitted by P given inputs x and y, as in the previous lecture. We can define the *distributional complexity* of a function is the minimum distributional complexity of any protocol that computes it.

Now, as above, let's allow our algorithms to be incorrect occasionally. The *bias* of a protocol P for a function f is defined as follows:

$$\text{bias}(P,f) = \Pr_{x,y}[P(x,y) = f(x,y)] - \Pr_{x,y}[P(x,y) \neq f(x,y)].$$

The bias of a perfectly correct protocol is 1; the bias of a protocol that is always incorrect is -1. Without loss of generality, we assume that the bias of any protocol is non-negative—the protocol is right as least as often as it is wrong—since we can always flip its last bit otherwise. The following properties are trivial:

$$0 \leq \text{bias}(P,f) \leq 1$$

$$\Pr_{x,y}[P(x,y) = f(x,y)] = \frac{\text{bias}(P,f)+1}{2}$$

$$\Pr_{x,y}[P(x,y) \neq f(x,y)] = \frac{\text{bias}(P,f)-1}{2}$$

We can now define the *communication complexity of f with bias ε* as the worst-case complexity of the best protocol P such that $\text{bias}(P,f) > \varepsilon$:

$$C_\varepsilon(f) = \min_{P:\text{bias}(P,f)>\varepsilon} C(P)$$

This definition only makes sense if $f(x,y)$ has approximately the same number of zeros and ones. For example, the protocol "Alice transmits 0" has bias $1-2^{-n}$ for the n-bit equality function. In particular, if we fix ε, then the trivial protocol has bias less than ε for all sufficiently large n.

On the other hand, if we want to find a function f such that $C_\varepsilon(f)$ is large, it's not enough for f to be balanced—consider the function that computes the parity of Alice's input x. We need a stronger notion of "balance" that applies to sub-matrices as well. The *bias* of a 0/1 matrix is the absolute difference between the number of 0s and the number of 1s. The *discrepancy* of a matrix (or function) f, denoted $\text{disc}(f)$, is the maximum bias of any minor of f.

For example, the following 4×4 matrix has bias 0 and discrepancy 6:

$$\begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

Theorem 2. $C_\varepsilon(f) \geq 2n + \lceil \lg \varepsilon - \lg \text{disc}(f) \rceil$

Proof: Let $k = C_\varepsilon(f)$, and let P be a protocol with complexity k and bias ε with respect to f. By definition of bias, we have

$$\text{bias}(P \oplus f) \geq 2^{2n}\varepsilon,$$

where $P \oplus f$ is the bitwise exclusive-or of the matrix f and the matrix (computed by) P. All I've done is replace the probabilities with an explicit count.

Let $M[X \times Y]$ denote the minor of a matrix M where $X \subseteq [2^n]$ and $Y \subseteq [2^n]$ specify the subset of rows and columns, respectively. The protocol P naturally divides the matrix f into at

most 2^k disjoint, *not* necessarily monochromatic, minors; the corresponding minors of matrix P are monochromatic. For some minor $X \times Y$ in this partition, we have

$$\text{bias}(P[X \times Y] \oplus f[X \times Y]) \geq \frac{2^{2n}\varepsilon}{2^k}.$$

Suppose $P[X \times Y]$ is all 1s; the argument for all 0s is similar. Then

$$\text{bias}(f[X \times Y]) = \text{bias}(P[X \times Y] \oplus f[X \times Y]) \geq \frac{2^{2n}\varepsilon}{2^k},$$

which implies that the discrepancy of f is at least $2^{2n-k}\varepsilon$. Solving for k gives the desired lower bound. □

12.3 Lower Bounds for Inner Product

Now consider the n-bit inner product function

$$F(x, y) = \sum_{i=1}^{n} x_i y_i \bmod 2.$$

In the homework, you were asked to prove upper and lower bounds for the deterministic communication complexity of this function. Here we derive the following lower bound for the communication complexity of this function with bias ε. (Recall that $\varepsilon < 1$!)

Theorem 3. *Let F be the n-bit inner product function, and let $N = 2^n$. We have $\text{disc}(F) \leq N^{3/2}$, which implies that $C_\varepsilon(F) \geq n/2 + \lceil \lg \varepsilon \rceil$.*

The lower bound on $C_\varepsilon(F)$ follows from the discrepancy bound by Theorem 2. To get a feel for this bound, observe that $C_\varepsilon(F) \to n/2$ as $\varepsilon \to 1$, and $C_\varepsilon(F) \to 0$ as $\varepsilon \to 1/2^{n/2}$. The discrepancy bound itself follows from the following remarkable fact about the minors of F.

Lemma 4. $\text{bias}(F[X \times Y]) \leq \sqrt{N|X||Y|}$ *for any minor $X \times Y$.*

Proof: Recall that an *eigenvector* of a square matrix M is a non-zero vector $\vec{u}$ such that $M\vec{u} = \lambda\vec{u}$ for some scalar λ, called an *eigenvalue*.

Let $H = J - 2F$, where F is the matrix of the inner product function and J is an $N \times N$ matrix of 1s. Equivalently, let H be the matrix obtained by replacing every 1 in F with -1, and every 0 in F with 1. It turns out that H is a so-called *Walsh-Hadamard matrix*, a standard worst-case example in many linear algebra problems. If we let H_n denote the matrix H for a particular value of n, we have $H_0 = [1]$ and

$$H_n = H_{n-1} \otimes H_1 = \begin{bmatrix} H_{n-1} & H_{n-1} \\ H_{n-1} & -H_{n-1} \end{bmatrix}$$

For example:

$$H_3 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{bmatrix}$$

The following facts are easy to prove from this definition:

- H is symmetric about the main diagonal.
- Every pair of rows (and therefore any pair of columns) in H is orthogonal.
- $H^2 = N \cdot I$, where I is the $N \times N$ identity matrix. (This follows from the previous two facts.)
- The eigenvectors $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_N$ of H form an orthonormal basis of $\mathbb{R}^N$. (This is true of any real symmetric $N \times N$ matrix.)
- The only possible eigenvalues of H are $\pm\sqrt{N}$. To see this, observe that for any eigenvector $\vec{u}$, we have both $H^2\vec{u} = N\vec{u}$ and $H^2\vec{u} = H(H\vec{u}) = \lambda^2\vec{u}$, so $\lambda^2 = N$.

Now we can express the bias of any minor of F in terms of the corresponding minor of H, as follows.

$$\text{bias}(F[X \times Y]) = \left|\sum_{i\in X}\sum_{j\in Y} H_{ij}\right| = \left|\sum_{i=1}^{n}\sum_{j=1}^{n}[i \in X]H_{ij}[j \in Y]\right| = \left|i_X^\top H\, i_Y\right|$$

Here, i_X and i_Y denote the 0/1 indicator vectors for the subsets X and Y:

$$i_X = ([0 \in X], [1 \in X], [2 \in X], \ldots, [N-1 \in X])$$

Note that the Euclidean length of i_X is exactly $\sqrt{|X|}$. To continue the derivation, we switch from the standard orthonormal basis of $\mathbb{R}^N$ to the orthonormal basis of eigenvectors of H. Specifically, we write

$$i_X = \sum_{i=1}^{N}\alpha_i\vec{u}_i \qquad i_Y = \sum_{i=1}^{N}\beta_i\vec{u}_i$$

Changing from one orthonormal basis to another does not change the length of a vector. It follows that $\sum_i \alpha_i^2 = \|i_X\| = \sqrt{|X|}$ and similarly, $\sum_i \beta_i^2 = \|i_Y\| = \sqrt{|Y|}$.

Now we can continue our derivation.

$$\begin{aligned}
\text{bias}(F[X \times Y]) &= \left|i_X^\top H\, i_Y\right| \\
&= \left|\left(\sum_{i=1}^{N}\alpha_i\vec{u}_i^\top\right) H \left(\sum_{i=1}^{N}\beta_i\vec{u}_i\right)\right| && \text{[change of basis]} \\
&= \left|\left(\sum_{i=1}^{N}\alpha_i\vec{u}_i^\top\right) \left(\sum_{i=1}^{N}\beta_i H\vec{u}_i\right)\right| && \text{[scalar mult commutative]} \\
&= \left|\left(\sum_{i=1}^{N}\alpha_i\vec{u}_i^\top\right) \left(\sum_{i=1}^{N}\beta_i\lambda_i\vec{u}_i\right)\right| && \text{[eigenvector def'n]} \\
&= \left|\sum_{i=1}^{N}\alpha_i\beta_i\lambda_i\right| && \text{[}\{\vec{u}_i\}\text{ is orthonormal basis]} \\
&\le \sqrt{N}\left|\sum_{i=1}^{N}\alpha_i\beta_i\right| && \text{[}|\lambda_i| = \sqrt{N}\text{]} \\
&= \sqrt{N}\,|\langle i_X, i_Y\rangle| && \text{[inner product def'n]} \\
&\le \sqrt{N}\sqrt{\|i_X\|\,\|i_Y\|} && \text{[Cauchy-Schwartz]} \\
&\le \sqrt{N|X||Y|}
\end{aligned}$$

The second to last step uses the *Cauchy-Schwartz inequality* $\langle u, v\rangle^2 \le \langle u, u\rangle\langle v, v\rangle = \|u\|^2\|v\|^2$; in English, the dot product of two vectors is less than the product of their lengths. □