CS395T: Pseudorandomness and Combinatorial Constructions
Lecture 2, January 18, 2001
Lecturer: Prof. David Zuckerman. Scribe: Joel A. Tropp

1 Open Problems in Complexity Theory

The following problems about complexity classes remain open

- We know that $\mathsf{P} \subseteq \mathsf{RP}$. Is it true that $\mathsf{P} = \mathsf{RP}$?
- We know that $\mathsf{P} \subseteq \mathsf{BPP}$. Is it true that $\mathsf{P} = \mathsf{BPP}$?

2 Testing the Value of an Algebraic Expression

Suppose we have an algebraic expression x of the form

$$x = c_1 a_1^{e_1} + \cdots + c_m a_m^{e_m},$$

where $\{c_i\}$ and $\{a_i\}$ are finite collections of n-bit integers and $\{e_i\}$ is a collection of nonnegative n-bit integers. We would like to determine whether the value of the expression is zero. The straightforward solution would be to multiply the expression out. But the space complexity is exponential in n (and so therefore is the time complexity). Therefore we turn to a randomized algorithm to solve the problem.

Algorithm 2.1 *To check whether such an expression x equals zero, do the following.*

1. *Choose at random a $3n$-bit prime number p.*
2. *Test whether $x \equiv 0 \pmod{p}$.*

For now, we ignore the difficulty of testing whether a number p is prime. The other steps of this algorithm take polynomial time because we can raise a number to an n-bit integral exponent in $\mathrm{O}(n)$ steps. Let us demonstrate the easier bound of $\mathrm{O}(n^2)$ steps. The algorithm involves writing the exponent as $e = 2^n b_n + 2^{n-1} b_{n-1} + \cdots + 2b_1 + b_0$ where $e = (b_n b_{n-1} \ldots b_1 b_0)_2$ is the binary representation. Then,

$$a^e = a^{2^n b_n} a^{2^{n-1} b_{n-1}} \ldots a^{2b_1} a^{b_0}.$$

We can determine the individual factors by repeatedly squaring the base (no more than n times). Afterword, we simply multiply the n factors together to obtain the result. Now, we prove that the algorithm is correct.

Proof. If $x = 0$, then $x \equiv 0 \pmod{p}$ for every integer p. Therefore, the algorithm always yields the correct result when $x = 0$.

Suppose now that $x \neq 0$. We need to show that $x \not\equiv 0 \pmod{p}$ w.h.p. (with high probability). It is clear that the following inequality holds:

$$\Pr\left[x \equiv 0 \pmod{p}\right] \leq \frac{\text{the number of prime divisors of } x}{\text{the number of } 3n\text{-bit primes}}.$$

To estimate the numerator, we write the prime factorization $x = p_1 p_2 \dots p_k$, where the factors need not be unique. Note that k is an upper bound on the number of *unique* prime divisors of x. Now, every $p_i \geq 2$, so we obtain the inequality $2^k \leq x$. Therefore, we have the upper bound $k \leq \log_2 x$. And $\log_2 x$ dominates the numerator for every positive integer x. By examining the expression for x, we find that

$$|x| \leq m\, 2^n\, (2^n)^{2^n}.$$

Therefore, $\log_2 |x| \leq \log_2 m + n + n\, 2^n$. It is obvious that $n = \mathrm{O}\,(2^n)$, and practical considerations suggest that $\log_2 m$ is dominated by 2^n. We conclude that the number of prime factors of x is $\mathrm{O}\,(2^{2n})$.

Next, recall that the notation $\pi(N)$ is used to express the number of primes which are less than or equal to N. The Prime Number Theorem, which was originally conjectured by Gauss, states that

$$\pi(N) = (1 + o(1))\frac{N}{\ln N} \quad \text{as } N \to \infty.$$

The first demonstrations of this theorem are due to Hadamard and Poussin who independently established the result in 1896. Using this formula, we discover that for large values of N, $\pi(N) > \frac{1}{2}N/\log N$. Therefore, the number $2^{3n-1}/(3n)$ is a lower bound for the number of $3n$-bit primes.

Our estimate becomes

$$\Pr\left[x \equiv 0 \pmod{p}\right] = \mathrm{O}\left(\frac{n\, 2^{2n}}{2^{3n}}\right) = \mathrm{O}\left(\frac{n}{2^n}\right).$$

And we see that the probability of the algorithm yielding a false positive is substantially less than one-half. $\square$

3 Primality Testing

Primality testing provides a second application of randomness. We shall discuss some results which prove that $\mathsf{COMPOSITES} \in \mathsf{RP}$. It is conjectured that $\mathsf{FACTORING} \notin \mathsf{BPP}$. As a result, it is probably much easier to find prime numbers than it is to determine the prime factors of a number. This disparity has been widely exploited in modern cryptography.

Due to the conjectured hardness of factoring, we need to explain how to find a certificate of compositeness which is not a factor. We require the following result.

Theorem 3.1 (Fermat's Little Theorem) *If p is a prime and $p \nmid a$, then $a^{p-1} \equiv 1 \pmod{p}$.*

Proof. It is known that the map $x \mapsto ax$ is an automorphism of the field $\mathbb{Z}_p^* = \{n < p : \gcd(n, p) = 1\}$ in which arithmetic is performed modulo p. Since p is prime, $\mathbb{Z}_p^* = \{1, 2, \dots, p-1\}$. Therefore, the following relation holds.

$$1 \cdot 2 \cdots (p-1) \equiv a \cdot 2a \cdots (p-1)a \pmod{p}.$$

In other words,

$$(p-1)! \equiv a^{p-1}(p-1)! \pmod{p}.$$

Since p is prime, we know that $p \nmid (p-1)!$. Dividing through, we obtain

$$a^{p-1} \equiv 1 \pmod{p}.$$

□

We would like a certain converse to hold, namely that if p is composite and $p \nmid a$ then $a^{p-1} \not\equiv 1 \pmod{p}$ at least half the time. This statement is not true. There are composite integers n, called Carmichael numbers, which always satisfy $\gcd(a, n) = 1$ implies $a^{n-1} \equiv 1 \pmod{n}$. The smallest Carmichael number is 561.

There is a sort of converse which does hold, however. The *Jacobi symbol* $\left(\frac{a}{n}\right) \equiv \pm 1 \pmod{n}$ is a number-theoretic quantity which we can compute in polynomial time. We have the following result.

Theorem 3.2 (Solovay-Strassen) *If p is prime, then for every $a \nmid p$, we have*

$$a^{(p-1)/2} \equiv \left(\frac{a}{p}\right) \pmod{p}.$$

If, on the other hand, p is composite, then for at least half the numbers $a \in \mathbb{Z}_p^*$, we have

$$a^{(p-1)/2} \not\equiv \left(\frac{a}{p}\right) \pmod{p}.$$

Therefore, $\mathsf{COMPOSITES} \in \mathsf{RP}$. It is also true that $\mathsf{PRIMES} \in \mathsf{RP}$, although this is quite difficult to prove. An immediate corollary is that $\mathsf{PRIMES} \in \mathsf{ZPP} = \mathsf{RP} \cap \mathsf{coRP}$. This property is desirable because we can always solve problems in ZPP in expected polynomial time without the risk of failure.

Proposition 3.1 *Every language $L \in \mathsf{ZPP}$ has a randomized algorithm which executes in expected polynomial time (as a function of the input) and which never produces a false result.*

Proof. For a given input x, we alterate between the RP and the coRP algorithms until one yields a certificate that $x \in L$ or $x \notin L$. Each round has a probability of success exceeding one-half. Therefore, the expected number of rounds does not exceed two. □

It can even be shown that $\mathsf{PRIMES} \in \mathsf{P}$ if we assume the Extended Riemann Hypothesis (ERH). One implication of ERH is that the multiplicative group $\mathbb{Z}_n^*$ is generated by the primes which satisfy $p \leq c \log^2 n$ for a small constant c. That is, for every number $a \in \mathbb{Z}_n^*$, there exist (nondistinct) primes $p_1, p_2, \ldots, p_k \leq c \log^2 n$ which satisfy $a \equiv p_1 p_2 \ldots p_k \pmod{n}$. Thus, we have the corollary

Corollary 3.1 *Under ERH,* $\mathsf{PRIMES} \in \mathsf{P}$.

4 The Riemann Hypothesis

Although this material is not strictly relevant to our main discussion, it is culturally significant.

Euler defined the zeta function as

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s}.$$

This converges for complex s with real part greater than 1. Riemann extended the zeta function to all complex numbers by analytic continuation. The traditional statement of the Riemann Hypothesis (RH) is that $\zeta(s) = 0$ implies that $\operatorname{Im} s = 0$ or $\operatorname{Re} s = \frac{1}{2}$.

Bernhard Riemann made this conjecture in a paper of 1859. It has been demonstrated for the first 10^9 zeroes of the zeta function, but it has not yet been proven. It is arguably the most important open problem in all of mathematics because of its deep connections with many diverse fields. An equivalent form of the RH is the following.

Conjecture 4.1 (Riemann Hypothesis) *For every $\varepsilon > 0$, the following estimate holds.*

$$\pi(x) = \mathrm{li}(x) + \mathrm{O}\left(x^{1/2+\varepsilon}\right),$$

where

$$\mathrm{li}(x) = \int_2^x \frac{\mathrm{d}t}{\log t}.$$

So there is an intimate connection between RH and the distribution of prime numbers. The Riemann Hypothesis sharpens the result of the Prime Number Theorem. There is also a stronger conjecture called the Extended Riemann Hypothesis. To understand it, we require a little more background information.

Dirichlet's Theorem states that every arithmetic progression $\{a + nd\}_{n=0}^{\infty}$ contains infinitely many primes. We denote by $\pi(x, d, a)$ the number of primes in the progression $\{a + nd\}$ which are less than or equal to x. By ignoring the sequences which have no primes at all, we can also make the estimate

$$\pi(x, d, a) = \frac{(1 + \mathrm{o}(1))\, x}{\phi(d) \ln x},$$

where ϕ is the Euler phi function which enumerates the integers which are relatively prime to its argument. More carefully,

$$\phi(n) = |\mathbb{Z}_n^*|.$$

ERH generalizes Dirichlet's Theorem, just as RH generalizes the Prime Number Theorem.

Conjecture 4.2 (Extended Riemann Hypothesis) *For every $\varepsilon > 0$, the following estimate holds.*

$$\pi(x, d, a) = \frac{1}{\phi(d)} \mathrm{li}(x) + \mathrm{O}\left(x^{1/2+\varepsilon}\right).$$

It is clear that ERH implies RH.