Lecture 10
22 March 2016

A *refutation* of a statement P is a proof of the statement NOT P. For example, given a boolean formula ϕ, asking someone to refute that ϕ is satisfiable means asking him to prove that ϕ is not satisfiable. Asking him to refute that G_0 and G_1 are isomorphic means asking him to prove that G_0 and G_1 are not isomorphic.

Today we will study the existence of short and efficiently verifiable refutations for NP problems. We will explain why, for example, it is believed that there are no such refutations for general instances of SAT. On the other hand, we will show how the (interactive) refutations for graph isomorphism follow directly from the existence of interactive proofs for the same problem satisfying an additional property called *statistical zero-knowledge.*

1 Efficient refutations

The class coNP consists of those problems $(\mathit{YES}, \mathit{NO})$ such that $(\mathit{NO}, \mathit{YES})$ is in NP. These are the problems that have short and efficiently checkable refutations. For example, the following two problems are in coNP:

$\overline{\text{SAT}}$: Is boolean formula ϕ unsatisfiable?

$\overline{\text{PMATCH}}$: Does graph G have no perfect matching?

$\overline{\text{GI}}$: Are graphs G_0 and G_1 not isomorphic?

Let us compare the decision problems SAT and $\overline{\text{SAT}}$. For SAT, given any boolean formula, we can always provide a short and efficiently checkable proof that the formula is satisfiable: The certificate is simply the satisfying assignment. But what if the formula is not satisfiable? Do we still expect to have a proof that this is the case?

Consider, for instance, the formula:

$$(x_1 \text{ OR } \overline{x_2}) \text{ AND } (x_1 \text{ OR } x_3 \text{ OR } \overline{x_4}) \text{ AND } (\overline{x_1} \text{ OR } \overline{x_2}) \text{ AND } (x_2).$$

This formula is not satisfiable for the following reason: The clauses $(x_1 \text{ OR } \overline{x_2})$ and $(\overline{x_1} \text{ OR } \overline{x_2})$ can only be simultaneously satisfied if x_2 is false, while the clause (x_2) requires x_2 to be true. So no matter which assignment we choose, the formula will not be satisfied.

For this specific example we did manage to give a proof that the formula is unsatisfiable. Is it possible to provide such a certificate for *every* unsatisfiable formula? If we allow the certificates to have length exponential in the size of the formula, the answer is yes. However, it is not known whether we can do so with polynomial length certificates that are verifiable in polynomial time.

Now let's look at $\overline{\text{PMATCH}}$: Can we get a certificate that a graph does not have a perfect matching? Here the answer is yes: Tutte's theorem tells us that a graph has no perfect matching *if and only if* there exists a subset of vertices S such that after removing S and all its incident edges, the rest of the graph has more than $|S|$ connected components with an odd number of vertices. So the set of vertices S is a certificate that the graph has no perfect matching: This set is certainly of polynomial size, and once we have S the conclusion of Tutte's theorem can be verified in polynomial time.

Actually, there is a more brutal way to certify that a graph has no perfect matching: Run Edmonds' perfect matching algorithm on the graph. If the algorithm does not find a perfect matching, we can take this as a certificate that the perfect matching does not exist (if it did exist, the algorithm would have found it).

These example illustrate the relationship between the classes P, NP, and coNP. In general, if a promise problem $(\mathit{YES}, \mathit{NO})$ is in P then $(\mathit{NO}, \mathit{YES})$ is also in P and therefore in NP, so P is a subclass of coNP. On the other hand, we do not know if SAT is in coNP, giving a potential example of a problem that is in NP but not in coNP. If this example is correct, then it is in fact universal:

Theorem 1. *If* SAT $\in$ coNP*, then* NP $=$ coNP*.*

Proof. We showed that there is a polynomial-time reduction from every NP-search problem to SAT. This implies that there is a polynomial-time reduction between their decision versions. So for every NP decision problem $P = (\mathit{YES}, \mathit{NO})$ there is a reduction that maps YES instances of P to satisfiable formulas and NO instances of P to unsatisfiable formulas. If SAT is in coNP then there is an polynomial-time verifier that accepts unsatisfiable formulas (with a proof of unsatisfiability) and rejects satisfiable ones (with any "proof"). Therefore there is a polynomial-time verifier for $\overline{P} = (\mathit{NO}, \mathit{YES})$, so P is in coNP. □

To summarize, we know for sure that P is in the intersection of NP and coNP, but it appears plausible that NP and coNP are distinct. Does the intersection of NP and coNP contain problems other than the ones in P? Graph isomorphism is a potential example: This is a problem that has both polynomial-time proofs and polynomial-time (interactive) refutations, but no known polynomial-time algorithms.

2 Statistical zero-knowledge

The interactive proof for graph non-isomorphism from the last lecture has one curious property: After interacting with the prover, the verifier does not learn anything about the graphs G_0 and G_1 beyond the fact that the two are not isomorphic. Recall that the verifier chooses a random bit $b \in \{0, 1\}$, sends a random graph isomorhphic to G_b to the prover and expects to receive b as an answer. So the verifier already knows the answer he is going to get (provided the graphs are indeed isomorphic and the prover is honest).

Contrast this with the standard proofs for SAT where the verifier does not merely find out that the formula is satisfiable, but also learns the satisfying assignment for it. Similarly, in a proof of graph isomorphism, the verifier learns not only that G_0 and G_1 are isomorphic, but also the isomorphism ϕ between the vertices of the two graphs. Is it possible to come up with alternate proofs that hide this additional information?

We will show how to do so in the case of graph isomorphism. Consider the following interactive proof for the statement "G_0 and G_1 are isomorphic:"

Interactive proof for graph isomorphism

On input (G_0, G_1):

P: Apply a random isomorphism to G_0 and send the resulting graph G to the Verifier.

V: Send a random bit $b \sim \{0, 1\}$ to the Prover.

P: Send an isomorphism π such that $\pi(G_b) = G$.

V: If $\pi(G_b) = G$, accept, otherwise reject.

This proof is clearly complete: If G_0 and G_1 are isomorphic then an isomorphism between G_b and G will exist regardless of the value of b, so the verifier accepts yes instances with probability one. On the other hand, the soundness (the probability that the verifier accepts when G_0 and G_1 are not isomorphic) is at most half: Regardless of the choice of G, G_b and G fail to be isomorphic with probability at least $1/2$ over the choice of b, in which case the verifier rejects. So this is a valid interactive proof for graph isomorphism.

Now let's see what the verifier learns when G_0 and G_1 are isomorphic (beyond the fact that they are isomorphic). The verifier observes a graph G obtained by applying a random isomorphism to G_0 (or G_1) together with an isomorphism π from G_b to G. This is "information" that the verifier could have generated on its own in the following way: First choose b and π at random and then set G to equal $\pi(G_b)$.

Proofs in which the verifier learns nothing beyond the validity of the statement to be proved are called zero-knowledge proofs. For the general definition we need the following concepts:

- The *statistical distance* between two distributions X and Y over the same sample space is the maximum over all events T of $\Pr[X \in T] - \Pr[Y \in T]$.
- The *view* of interactive Turing Machine A in an interaction with B consists of A's randomness and the sequence of messages exchanged between the two.
- A function f is *negligible* if for every polynomial p and all sufficiently large n, $f(n) \leq p(n)$.

Definition 2. An interactive proof (V, P) for promise problem $(\mathit{YES}, \mathit{NO})$ is *statistical zero-knowledge* if there exists a randomized polynomial-time Turing Machine S called *the simulator* such that for every $x \in \mathit{YES}$, the statistical distance between $S(x)$ and the view of V in the interaction with P on input x is negligible in $|x|$.

The class SZK consists of all (promise) problems that have statistical zero-knowledge interactive proofs, regardless of the number of rounds.

In the proof of graph non-isomorphism, when G_0 and G_1 are not isomorphic the verifier's view consists of a random bit b, a random permutation π, the graph $\pi(G_b)$ (sent to the prover) and the bit b' equal to b (sent by the prover). The simulator outputs $(b, \pi, \pi(G_b), b)$, which in this case is identically distributed to the verifier's view (i.e., the statistical distance is zero).

In the proof of graph isomorphism, when G_0 and G_1 are isomorphic the verifier's view consists of (G, b, π) where G is a random graph isomoprhic to G_0 and b, π are random conditioned on $\pi(G_b) = G$. The simulator outputs $(\pi(G_b), b, \pi)$ which is again identically distributed to the verifier's view. So both examples satisfy our definition of statistical zero-knowledge.

3 Statistical difference

A *sampler* is a circuit $C \colon \{0,1\}^m \to \{0,1\}^n$ that takes a uniformly random input r and outputs a sample $C(r) \in \{0,1\}^n$. The *statistical distance* between two samplers C_0 and C_1 with outputs in $\{0,1\}^n$ is the statistical distance between their output distributions. We consider the following promise problem:

SD (STATISTICAL DIFFERENCE):
Input: Two samplers C_0 and C_1.
Yes instances: The statistical distance between C_0 and C_1 is at least $2/3$.

No instances: The statistical distance between C_0 and C_1 is at most $1/3$.

We argue that SD has a statistical zero-knowledge proof. First, we show that this is the case when the quantities $2/3$ and $1/3$ are replaced by $1-\varepsilon$ and $1/3$, where ε is some negligible function of the input length (the sizes of C and D). We then design a reduction that affects this change of quantities.

The proof for statistical distance is very much like the one of graph non-isomorphism:

Interactive proof for statistical difference

On input (C_0, C_1):

V: Choose random $b \sim \{0,1\}$, random $r \sim \{0,1\}^m$ and send $C_b(r)$ to the prover.

P: Send $b' = 1$ if $y \in T$ and $b' = 0$ if not, where T is the event that maximizes $\Pr_r[C_1(r) \in T] - \Pr_r[C_0(r) \in T]$.

V: If $b' = b$, accept, otherwise reject.

If (C_0, C_1) is a yes instance of SD then $\Pr_r[C_1(r) \in T] - \Pr_r[C_0(r) \in T] \geq 1-\varepsilon$, so $\Pr[C_1(r) \notin T] \leq \varepsilon$ and $\Pr[C_0(r) \in T] \leq \varepsilon$. Regardless of the choice of b, the prover makes a mistake with probability at most ε. To analyze no instances we make use of the following alternative characterization of statistical distance.

Lemma 3. *The statistical distance between X and Y is δ if and only if there exists a joint distribution (X, Y) such that $X = Y$ with probability $1-\delta$.*

If (C_0, C_1) is a no instance of SD then the verifier's first message can equivalently be described like this: The verifier samples $b \sim \{0,1\}$ and (Y_0, Y_1) from the joint distribution on the outputs Y_0, Y_1 of C_0, C_1 from Lemma 3 then sends Y_b to the prover. Conditioned on $Y_0 = Y_1$, b' and b are independent and the prover suceeds with probability exactly half, so the probability that the verifier accepts can be at most $\frac{1}{2}(1-\frac{2}{3}) + \frac{1}{3} = \frac{2}{3}$. Therefore the described proof has completeness error ε and soundness error at most $2/3$.

It remains to argue that the proof is statistical zero-knowledge. If (C_0, C_1) is a yes instance, the verifier's view consists of b, r, $C_b(r)$, and b'. The simulator outputs b, r, $C_b(r)$, and b. Since $b' = b$ with probability $1-\varepsilon$, there is a joint distribution under which the two views are identically distributed with probability $1-\varepsilon$. By the other direction of Lemma 3, the statistical distance between the two views is at most ε, therefore negligible in the input size.

Manipulating statistical difference We now show how to enlarge the statistical distance gap between yes instances and no instances from $2/3$ versus $1/3$ to $1-\exp(s^{-\Omega(1)})$ versus $1/3$, where s is the instance size. We will apply two different transformations given in the following lemmas:

Lemma 4. *Given two distributions X_0 and X_1, let X_0' and X_1' consists of two independent samples of X_a and X_b where a and b are random bits conditioned on $a \oplus b = 0$ and $a \oplus b = 1$, respectively. Then the statistical distance between X_0' and X_1' is the square of the statistical distance between X_0 and X_1.*

Proof. Given an event T that distinguishes X_1 from X_0, let T' be the event that exactly one of X_a and X_b are in T. Then it can be calculated that

$$\Pr[X_1' \in T'] - \Pr[X_0' \in T'] = (\Pr[X_1 \in T] - \Pr[X_0 \in T])^2$$

so the statistical distance between X'_0 and X'_1 is at least δ^2. To show it is at most δ^2 we apply Lemma 3. Let $Z = X_0 = X_1$ be the marginal distribution conditioned on $X_0 = X_1$ and X'_0, X'_1 be the marginal distributions conditioned on $X_0 \neq X_1$. Then the joint distribution of (X_a, X_b) is of the type (Z, Z) with probability $(1-\delta)^2$, (X'_a, Z) and (Z, X'_b) with probability $\delta(1-\delta)$ each and (X'_a, X'_b) with probability δ^2, where the two samples are independent. Now compare the four types of samples conditioned on $a \oplus b = 0$ and on $a \oplus b = 1$. The first three are identical, while the last one is disjoint. It follows that X'_0, X'_1 are identical with probability at least $1 - \delta^2$. By the other direction of Lemma 3 the statistical distance between X'_0 and X'_1 is at most δ^2. □

Lemma 5. *Let X'_0 and X'_1 consist of k independent copies of X_0 and X_1, respectively. If the statistical distance between X_0 and X_1 is δ then the statistical distance between X'_0 and X'_1 is at most $k\delta$ and at least $1 - 2\exp(-k\delta^2/2)$.*

Proof. By Lemma 3 X_0 and X_1 are identical with probability $1 - \delta$. The probability that all k samples are identical is then at least $1 - (1-\delta)^k = 1 - k\delta$. For the other inequality, if T is an event such that $\Pr[X_1 \in T] - \Pr[X_0 \in T] \geq \delta$ and $(\Pr[X_1 \in T] + \Pr[X_0 \in T])/2 = p$ then by a Chernoff bound the probability that at least kp of the copies of X_b are in T is at least $1 - \exp(-k\delta^2/2)$ if $b = 1$ and at most $\exp(-k\delta^2/2)$ if $b = 0$. So the statistical distance must be at least $1 - 2\exp(-k\delta^2/2)$. □

Given samplers C_0 and C_1 of size s, Lemma 4 produces samplers of size $2s + O(1)$ whose statistical distance is the square of the original one. If we apply this lemma $\log \ell$ times, we obtain samplers (C'_0, C'_1) of size $2^{O(\ell)} \cdot s$ whose statistical distance is at least $(2/3)^\ell$ for yes instances (C_0, C_1) and at most $(1/3)^\ell$ for no instances. Now applying Lemma 5 with $k = 3^{\ell-1}$ to C'_0 and C'_1 we end up with a pair of samplers of size $2^{O(\ell)}s$ whose statistical distance is at most $1/3$ for no instances and at least

$$1 - 2\exp(-k(2/3)^{2\ell}/2) \geq 1 - \exp(-(4/3)^\ell/6)$$

for yes instances. Choosing $\ell = \log s$ gives a polynomial-time reduction with a negligible error for yes instances as desired.

4 Completeness of statistical difference

The *complement* $\overline{\text{SD}}$ of SD is hard for statistical zero-knowledge, namely:

Theorem 6. *For every promise problem (YES, NO) there is a polynomial-time reduction that on input x outputs a pair of samplers C_0, C_1 such that*

$$\begin{aligned} &\textit{If } x \in \textit{YES}, \quad \textit{then } C_0 \textit{ and } C_1 \textit{ have statistical distance at most } 1/3, \\ &\textit{If } x \in \textit{NO}, \quad \textit{then } C_0 \textit{ and } C_1 \textit{ have statistical distance at least } 2/3. \end{aligned}$$

Corollary 7. *If a promise problem (YES, NO) is in* SZK *then its complement (NO, YES) is also in* SZK.

Proof. By Theorem 6, (YES, NO) reduces to $\overline{\text{SD}}$. Therefore (NO, YES) reduces to SD. In the last section we argued that SD is in SZK, so (NO, YES) is also in SZK. □

In particular, $\overline{\text{SD}}$ itself is in statistical zero-knowledge. By reversing the role of the yes and no instances we also obtain that SD is complete for SZK. Since SD has a two-round interactive proof (regardless of its zero-knowledge-ness) it is in AM. Combining all these observations we get the complexity class containment

$$\boxed{\text{SZK} \subseteq \text{AM} \cap \text{coAM}}$$

where coAM is the class of problems (NO, YES) such that (YES, NO) is in AM. In particular, this makes it unlikely that SAT has statistical zero-knowledge proofs because it would then have efficient refutations.

Before we sketch the proof of Theorem 6 let us explain how a restricted variant of graph non-isomorphism reduces to $\overline{\text{SD}}$. As in the example we gave in the last lecture, we will work under the promise that G_0 and G_1 have no automorphisms. We want a reduction that maps pairs of graphs (G_0, G_1) to pairs of circuits (C_0, C_1) so that if G_0 and G_1 are not isomorphic then C_0 and C_1 are statistically close, while if G_0 and G_1 are isomorphic then C_0 and C_1 should be statistically far.

Let us consider the distribution $X = \pi(G_b)$ where π is a random isomorphism and b is a random bit. If there are no automorphisms, and n is the number of vertices, then X is a flat distribution over a set of size $2n!$ when G_0 and G_1 are not isomorphic and $2n!$ when they are. If we take a sequence of six independent copies of X, the resulting distribution X^6 is also flat over support size $2^6(n!)^6$ and $(n!)^6$ for yes and no instances respectively. We now apply the following lemma to X^6:

Lemma 8. *Let Z be a random variable taking values in $\{0,1\}^n$. If Z is a flat distribution over a set of size at least 2^m and $H\colon \{0,1\}^n \to \{0,1\}^{m-2\ell}$ is a pairwise-independent hash function then the statistical distance between $(H, H(Z))$ and (H, U) is at most $2^{-\ell}$, where U is a uniform random variable independent of H.*

We apply Lemma 8 to $Z = X^6$ with $m = 6\log(n!) + 6$ and $\ell = 2$. If G_0 and G_1 are not isomorphic by Lemma 8 the statistical distance between $(H, H(X^k))$ and (H, U) is at most $1/4$. If G_0 and G_1 are isomorphic then for every H there are at most $(n!)^6$ possible outputs of $H(X^k)$ and $4(n!)^6$ possible outputs of U. If T is the event consisting of the possible outputs of $(H, H(X^k))$ then $\Pr[(H, H(X^6)) \in T] = 1$ while $\Pr[(H, U) \in T]$ is at most $1/4$, so the statistical distance between the two is at least $3/4$.

Proof sketch of Theorem 6 The first step in the proof of Theorem 6 is a transformation of the statistical zero-knowledge proof for (YES, NO) into one in which the verifier uses public coins, like the one for graph isomorphism. We will not show this part of the proof and will assume that each verifier message consists of a sequence of public coins.

We will assume that the completeness and soundness gaps of the proof system are negligible. This can be arranged by repeating the proof *sequentially* sufficiently many times. Let $2r(n)$ be a bound on the number of rounds of the proof system on inputs of size n, assuming the verifier asks first. Finally, instead of proving a $1/3$ versus $2/3$ gap we will prove a negligible versus $\Omega(1/r(n))$ gap. This can then be amplified to $1/3$ versus $2/3$ using Lemma 5. We will freely make use of properties of statistical distance given in the Appendix.

On input x of (YES, NO), the output distributions of C_0 and C_1 will consists of two parts. The first part is a single bit: In C_0 first the verifier's view is sampled from $S(x)$ then the verifier's decision given this view is output. In C_1 this bit is always 1.

The second part is a partial interaction sampled independently as follows: In both C_0 and C_1 a random number i between 0 and $r(|x|) - 1$ is chosen. In C_0, the first $2I + 1$ messages $S(x)$ of the simulator are output. In C_1, the first $2I$ messages of $S(x)$ are output followed by an independent random string corresponding to the next message of the verifier. To summarize:

C_0 samples (V's output on $S(x)$, first $2I+1$ rounds of $S(x)$),
C_1 samples (1, first $2I$ rounds of $S(x)$ followed by a random question).

We show that if x is a yes instance then C_0 and C_1 are statistically close. First, the verifier must almost always accept the view provided by $S(x)$, for otherwise it would distinguish the output of the simulator from its view of the actual interaction with the prover, violating the zero-knowledge property. So the first bit is almost always 1 in both distributions. For the second part, by the zero-knowledge property for every i the first $2i+1$ messages of the simulator are ε-close to the same messages in the actual interaction for some negligible ε. In the actual interaction, these consist of the first $2i$ messages followed by an independent random question asked by the verifier. Applying zero-knowledge again, they are therefore 2ε-close to the first $2i$ messages of the interaction followed by a random verifier message. We conclude that both parts of C_0 and C_1 are within negligible statistical distance.

Now suppose x is a no instance. Consider the following distribution $T(x)$ of verifier's views: First the verifier asks a random question q_1. Then the prover samples an answer a_1 by running $S(x)$ conditioned on the first message of $S(x)$ being equal to q_1. Then the verifier answers a random question q_2. Then the prover samples an answer a_2 by running $S(x)$ conditioned on the first three messages being equal to q_1, a_1, and q_2 respectively, and so on. We consider two cases.

If the statistical distance between $S(x)$ and $T(x)$ is at most $1/3$ then the verifier rejects the view $S(x)$ with probability at least $1/2$ because $T(x)$ represents an actual interaction between a verifier and a prover, so the probability that the verifier accepts this interaction is negligible. By statistical closeness, the probability that the verifier accepts $S(x)$ can be at most $1/2$. Then the first bit of C_0 is one with probability at most $1/2$, so the statistical distance between C_0 and C_1, even restricted on the first bit, is at least $1/2$.

If, on the other hand, the statistical distance between $S(x)$ and $T(x)$ exceeds $1/3$ we claim that the statistical distance between the second part of C_0 and C_1 is at least $1/3r$, where $r = r(|x|)$. To see this consider the following hybrid distributions $H_0, \dots, H_r$: In distribution H_i the first $2i-1$ rounds are sampled as in $T(x)$, but the remaining rounds are sampled from $S(x)$ conditioned on these first $2i-1$ messages. Then $H_0 = S(x)$ and $H_r = T(x)$, so the statistical distance between H_0 and H_r is at least $1/3$. It follows that for a random I, the statistical distance between H_I and H_{I+1} is at least $1/3r$. This remains true if we truncate H_I and H_{I+1} after the first $2I+1$ rounds because the remaining rounds are sampled from the same conditional distribution. But then H_I and H_{I+1} become identical to the second part of C_0 and C_1 so the statistical distance between these two is at least $1/3r$.

References

Zero-knowledge was first defined and studied by Goldwasser, Micali, and Rackoff. The containments SZK $\in$ coAM and SZK $\in$ AM were shown separately by Fortnow and by Aiello and Håstad. The equivalence of public and private coins and the closure under complement (Corollary 7) were proved by Okamoto. The completeness of statistical distance and our proof of Theorem 6 follows the work of Sahai and Vadhan. Lemma 8 is usually called the "leftover hash lemma" and is due to Håstad, Impagliazzo, Levin, and Luby.

Properties of statistical distance

In Section 4 we made use of the following properties of statistical distance sd. These can be derived from the definition and Lemma 3.

Claim 9. *For any three random variables* X, Y, Z, $\mathrm{sd}(X, Z) \leq \mathrm{sd}(X, Y) + \mathrm{sd}(Y, Z)$.

Claim 10. *For any pair of random variables* X, Y *and randomized procedure* A, $\mathrm{sd}(A(X), A(Y)) \leq \mathrm{sd}(X, Y)$. *In particular, projection on a subset of coordinates cannot increase statistical distance.*