Notes on Complexity Theory Last updated: October, 2011

Lecture 16

Jonathan Katz

1 Interactive Proofs

Let us begin by re-examining our intuitive notion of what it means to "prove" a statement. Traditional mathematical proofs are *static* and are verified *deterministically*: the verifier checks the claimed proof of a given statement and is either convinced that the statement is true (if the proof is correct) or remains unconvinced (if the proof is flawed — note that the statement may possibly still be true in this case, it just means there was something wrong with the proof). A statement is true (in this traditional setting) iff there *exists* a valid proof that convinces a legitimate verifier.

Abstracting this process a bit, we may imagine a prover $\mathbf{P}$ and a verifier $\mathbf{V}$ such that the prover is trying to convince the verifier of the truth of some particular statement x; more concretely, let us say that $\mathbf{P}$ is trying to convince $\mathbf{V}$ that $x \in L$ for some fixed language L. We will require the verifier to run in polynomial time (in $|x|$), since we would like whatever proofs we come up with to be efficiently verifiable. A traditional mathematical proof can be cast in this framework by simply having $\mathbf{P}$ send a proof π to $\mathbf{V}$, who then deterministically checks whether π is a valid proof of x and outputs $\mathbf{V}(x, \pi)$ (with 1 denoting acceptance and 0 rejection). (Note that since $\mathbf{V}$ runs in polynomial time, we may assume that the length of the proof π is also polynomial.) The traditional mathematical notion of a proof is captured by requiring:

- If $x \in L$, then there *exists* a proof π such that $\mathbf{V}(x, \pi) = 1$.
- If $x \notin L$, then *no matter what proof* π the prover sends we have $\mathbf{V}(x, \pi) = 0$.

We refer to the above as a type of proof system, a term we will define more formally later. It should be obvious that L has a proof system of the above sort iff $L \in \mathcal{NP}$.

There are two ways the above can be generalized. First, we can allow the verifier to be *probabilistic*. Assume for a moment that we restrict the prover to sending an empty proof. If the verifier is deterministic, then a language L has a proof system of this sort only if $L \in \mathcal{P}$ (as the prover is no help here). But if the verifier is probabilistic then we can handle any $L \in \mathcal{BPP}$ (if we allow two-sided error). If we go back to allowing non-empty proofs, then we already gain something: we can eliminate the error when $x \in L$. To see this, recall the proof that $\mathcal{BPP} \in \Sigma_2$. The basic idea was that if a set $S \subset \{0,1\}^\ell$ is "small" then for any strings $z_1, \ldots, z_\ell \in \{0,1\}^\ell$, the set $\bigcup_i (S \oplus z_i)$ is still "small." To make this concrete, say $|S| \leq 2^\ell/4\ell$. Then for any $z_1, \ldots, z_\ell$ we have:

$$\left| \bigcup_{i=1}^{\ell} (S \oplus z_i) \right| \leq \ell \cdot |S| \leq 2^\ell/4. \tag{1}$$

On the other hand, if S is "large" (specifically, if $|S| \geq (1 - \frac{1}{4\ell}) \cdot 2^\ell$) then there exist $z_1, \ldots, z_m$ such that $\bigcup_i (S \oplus z_i) = \{0,1\}^\ell$.

The above leads to the following proof system for any $L \in \mathcal{BPP}$: Let M be a $\mathcal{BPP}$ algorithm deciding L, using a random tape of length ℓ, and having error at most $1/4\ell$ (for some polynomial ℓ).

The prover sends a proof $\pi = (z_1, \ldots, z_\ell)$ to the verifier (where each $z_i \in \{0,1\}^\ell$); $\mathbf{V}$ then chooses a random $r \in \{0,1\}^\ell$ and accepts iff

$$\bigvee_{i=1}^{\ell} M(x; r \oplus z_i) = 1.$$

For common input x, let S_x be the set of random coins for which $M(x) = 1$. If $x \in L$, then $|S_x| \geq (1 - \frac{1}{4\ell}) \cdot 2^\ell$ and so there does indeed exist $\pi = (z_1, \ldots, z_\ell)$ such that $r \in \bigcup_i (S_x \oplus z_i)$ for *every* $r \in \{0,1\}^\ell$. Fixing such a π, this means that for every r there exists an index i for which $r \in S_x \oplus z_i$, and so $r \oplus z_i \in S_x$. Thus, if the prover sends this π the verifier will always accept. On the other hand, if $x \notin L$ then $|S_x| \leq 2^\ell/4\ell$ and so, using Eq. (1), we have

$$\Pr_{r \in \{0,1\}^\ell} \left[r \in \bigcup_{i=1}^{\ell} (S \oplus z_i) \right] \leq 1/4.$$

So $\mathbf{V}$ accepts in this case with probability at most 1/4.

To summarize, we have shown a proof system for any $L \in \mathcal{BPP}$ such that:

- If $x \in L$, then there *exists* a proof π such that $\Pr[\mathbf{V}(x, \pi) = 1] = 1$.
- If $x \notin L$, then *no matter what proof* π the prover sends we have $\Pr[\mathbf{V}(x, \pi) = 1] \leq 1/4$.

Thus, assuming $\mathcal{P} \neq \mathcal{BPP}$, we see that *randomization* helps. And assuming $\mathsf{co}\mathcal{RP} \neq \mathcal{BPP}$, allowing *communication from the prover to the verifier* helps.

We can further generalize proof systems by allowing *interaction* between the prover and verifier. (One can think of this as allowing the verifier to ask questions. In this sense, the notion of a proof becomes more like a lecture than a static proof written in a book.) Note that unless we also allow randomness, allowing interaction will not buy us anything: if the verifier is deterministic then the prover can predict all the verifier's questions in advance, and simply include all the corresponding answers as part of the (static) proof.

Before we explore the additional power of interaction, we introduce some formal definitions. For interactive algorithms $\mathbf{P}, \mathbf{V}$, let $\langle \mathbf{P}, \mathbf{V} \rangle (x)$ denote the output of $\mathbf{V}$ following an interaction of $\mathbf{P}$ with $\mathbf{V}$ on common input x.

Definition 1 *$L \in \mathcal{IP}$ if there exist a pair of interactive algorithms $(\mathbf{P}, \mathbf{V})$, with $\mathbf{V}$ running in probabilistic polynomial time (in the length of the common input x), such that*

1. *If $x \in L$, then $\Pr[\langle \mathbf{P}, \mathbf{V} \rangle (x) = 1] = 1$.*
2. *If $x \notin L$, then for* any *(even cheating) $\mathbf{P}^*$ we have $\Pr[\langle \mathbf{P}^*, \mathbf{V} \rangle (x) = 1] \leq 1/2$.*

(We stress that $\mathbf{P}$ and $\mathbf{P}^*$ are allowed to be computationally unbounded.) $(\mathbf{P}, \mathbf{V})$ satisfying the above are called a proof system *for L. We say $L \in \mathcal{IP}[\ell]$ if it has a proof system as above using $\ell = \ell(|x|)$ rounds of interaction (where each message sent by either party counts as a round).*

Using this notation, we have seen already that $\mathcal{NP} \cup \mathcal{BPP} \subseteq \mathcal{IP}[1]$.

Some comments on the definition are in order:

- One could relax the definition to allow for *two-sided* error, i.e., error even when $x \in L$. It is known, however, that this results in an equivalent definition [1] (although the round complexity increases by a constant). On the other hand, if the definition is "flipped" so that we allow error only when $x \in L$ (and require no error when $x \notin L$) we get a definition that is equivalent to $\mathcal{NP}$.

- As usual, the error probability of 1/2 is arbitrary, and can be made exponentially small by repeating the proof system suitably many times. (It is easy to see that sequential repetition works, and a more detailed proof shows that parallel repetition works also [2, Appendix C].)

- Although the honest prover is allowed to be computationally unbounded, it suffices for it to be a PSPACE machine. In certain cases it may be possible to have **P** run in polynomial time (for example, if $L \in \mathcal{NP}$ and **P** is given a proof π as auxiliary information). In general, it remains an open question as to how powerful **P** needs to be in order to give a proof for some particular class of languages.[1]

1.1 Graph Non-Isomorphism is in $\mathcal{IP}$

It is possible to show that $\mathcal{IP} \subseteq$ PSPACE (since, fixing some **V** and some x, we can compute the optimal prover strategy in polynomial space). But does interaction buy us anything? Does $\mathcal{IP}$ contain anything more than $\mathcal{NP}$ and $\mathcal{BPP}$? We begin by showing the rather surprising result that graph *non*-isomorphism is in $\mathcal{IP}$.

If G is an n-vertex graph and π is a permutation on n elements, we let $\pi(G)$ be the n-vertex graph in which

$$(i,j) \text{ is an edge in } G \Leftrightarrow (\pi(i),\pi(j)) \text{ is an edge in } \pi(G).$$

Note that G_0 is isomorphic to G_1 (written $G_0 \cong G_1$) iff $G_0 = \pi(G_1)$ for some π. (We identify a graph with its adjacency matrix. So, there is a difference between two graphs being *equal* [i.e., having the *same* adjacency matrix] and being *isomorphic*.)

Let G_0, G_1 be two graphs. The proof system for graph non-isomorphism works as follows:

1. The verifier chooses a random bit b and a random permutation π, and sends $G' = \pi(G_b)$.

2. If $G' \cong G_0$, the prover replies with 0; if $G' \cong G_1$, it replies with 1.

3. The verifier accepts iff the prover replies with **V**'s original bit b.

Note that if $G_0 \ncong G_1$, then it cannot be the case that both of $G' \cong G_0$ and $G' \cong G_1$ hold; so, the prover always answers correctly. On the other hand, if $G_0 \cong G_1$ (so that (G_0, G_1) is *not* in the language) then the verifier's bit b is completely hidden to the prover (even though the prover is all-powerful!); this is because a random permuted copy of G_0 is in this case distributed identically to a random permuted copy of G_1. So when G_0, G_1 *are* isomorphic, even a cheating prover can only make the verifier accept with probability 1/2.

[1] For example, we will soon see that co$\mathcal{NP} \subseteq \mathcal{IP}$. By what we have just said, we know that if $L \in$ co$\mathcal{NP}$ then there exists a proof system for L with a prover running in PSPACE. But we do not know whether there exists a proof system for L with a prover running in, say, $\mathcal{P}^{\text{co}\mathcal{NP}} = \mathcal{P}^{\mathcal{NP}}$.

2 Public-Coin Proof Systems

Crucial to the above protocol for graph non-isomorphism is that the verifier's coins are *private*, i.e., hidden from the prover. At around the same time the class $\mathcal{IP}$ was proposed, a related class was proposed in which the verifier's coins are required to be *public* (still, the verifier does not toss coins until they are needed, so that the prover does not know what coins will be tossed in the future). These are called *Arthur-Merlin* proof systems, where Arthur represents the (polynomial-time) verifier and Merlin the (all-powerful) prover. We again require perfect completeness and bounded soundness (though see Theorems 1 and 2 below). As in the case of $\mathcal{IP}$ one can in general allow polynomially many rounds of interaction. Although it might appear that Arthur-Merlin proofs are (strictly) *weaker* than general interactive proofs, this is not the case [3]. We do not prove this, but an indication of the general technique will be given in Section **??**.

We will consider for now only the Arthur-Merlin classes **MA** and **AM** where there are one or two rounds of interaction. For the class **MA** Merlin talks first, and then Arthur chooses random coins and tries to verify the "proof" that Merlin sent. (We have already seen this type of proof system before when we showed an interactive proof for $\mathcal{BPP}$.) For the class **AM** Arthur talks first but is limited to sending its random coins (so the previous proof of graph non-isomorphism does not satisfy this); then Merlin sends a proof that is supposed to "correspond" to these random coins, and Arthur verifies it. (Arthur does not choose any additional random coins after receiving Merlin's message, although it would not change the class if Arthur did; see Theorem 3, below.) One can also express these in the following definition, which is just a specialization of the general definition of Arthur-Merlin proofs to the above cases:

Definition 2 *$L \in \mathbf{MA}$ if there exists a deterministic algorithm $\mathbf{V}$ running in polynomial time (in the length of its first input) such that:*

- *If $x \in L$ then $\exists \pi$ such that for all r we have $\mathbf{V}(x, \pi, r) = 1$.*
- *If $x \notin L$ then $\forall \pi$ we have $\Pr_r[\mathbf{V}(x, \pi, r) = 1] \leq 1/2$.*

$L \in \mathbf{AM}$ if there exists a deterministic algorithm $\mathbf{V}$ running in polynomial time (in the length of its first input) such that:

- *If $x \in L$ then for all r there exists a π such that $\mathbf{V}(x, r, \pi) = 1$.*
- *If $x \notin L$ then $\Pr_r[\exists \pi : \mathbf{V}(x, r, \pi) = 1] \leq 1/2$.*

In the case of **MA** *the prover (Merlin) sends π and the verifier (Arthur) then chooses random coins r, while in the case of* **AM** *the verifier (Arthur) sends random coins r and then the prover (Merlin) responds with π.*

MA can be viewed as a randomized version of $\mathcal{NP}$ (since a fixed proof is verified using randomization) and so a language in **MA** is sometimes said to have "publishable proofs." It is clear that Arthur-Merlin proofs are not more powerful than the class $\mathcal{IP}$ (since an Arthur-Merlin proof system is a particular kind of proof system).

As we have said, **MA** and **AM** do not change if we allow error when $x \in L$. We now prove this. Let $\mathbf{MA}_\varepsilon$ and $\mathbf{AM}_\varepsilon$ denote the corresponding classes when (bounded) two-sided error is allowed.

Theorem 1 $\mathbf{MA}_\varepsilon = \mathbf{MA}$.

Proof Let $L \in \mathbf{MA}_\varepsilon$. Then there is a proof system such that if $x \in L$ then there exists a π (that Merlin can send) for which Arthur will accept with high probability (i.e., $\mathbf{V}(x, \pi, r) = 1$ with high probability over choice of r), while if $x \notin L$ then for any π Arthur will accept only with low probability (i.e., $\mathbf{V}(x, \pi, r) = 1$ with low probability over choice of r). For a given x *and* π, let $S_{x,\pi}$ denote the set of coins r for which $\mathbf{V}(x, \pi, r) = 1$. So if $x \in L$ there exists a π for which $S_{x,\pi}$ is "large," while if $x \notin L$ then for *every* π the set $S_{x,\pi}$ is "small." Having Merlin send π along with a proof that $S_{x,\pi}$ is "large" (exactly as in the $\mathcal{BPP}$ case) gives the desired result. ■

Theorem 2 $\mathbf{AM}_\varepsilon = \mathbf{AM}$.

Proof Say $L \in \mathbf{AM}_\varepsilon$. Using standard error reduction, we thus have a proof system for L in which Arthur sends a random string r of (polynomial) length ℓ and the error is less than $1/4\ell$. For a common input x, let S_x denote the set of challenges r (that Arthur can send) for which there exists a π (that Merlin can send) such that $\mathbf{V}(x, r, \pi) = 1$. By definition of $\mathbf{AM}_\varepsilon$, if $x \in L$ then $|S_x| \geq (1 - \frac{1}{4\ell}) \cdot 2^\ell$ while if $x \notin L$ then $|S_x| \leq 2^\ell/4\ell$. Exactly as in the proof system for $\mathcal{BPP}$ shown previously, this means that we have the following proof system for L:

1. Merlin sends $z_1, \ldots, z_\ell \in \{0,1\}^\ell$.
2. Arthur sends random $r' \in \{0,1\}^\ell$.
3. Merlin proves that $r' \in \bigcup_i (S_x \oplus z_i)$ by finding an i such that $r' \oplus z_i \in S_x$, setting $r = r' \oplus z_i$, and then computing the appropriate response π to the "challenge" r. So Merlin's response is (i, π).
4. Arthur runs $\mathbf{V}(x, r' \oplus z_i, \pi)$ and outputs the result.

The above has perfect completeness and soundness error at most 1/4 (we do not go through the analysis since it is the same as in the $\mathcal{BPP}$ case).

The problem is that the above is a three-round proof system (notationally, it shows that $L \in \mathbf{MAM}$)! But we show below that an "**MA**" step can be replaced by an "**AM**" step (while preserving perfect completeness), and so if we apply that here and then combine Merlin's last two messages we get an $\mathbf{AMM} = \mathbf{AM}$ protocol. ■

As promised, we now show that $\mathbf{MA} \subseteq \mathbf{AM}$. More generally, the proof shows that an "**MA**" step can be replaced by an "**AM**" step.

Theorem 3 $\mathbf{MA} \subseteq \mathbf{AM}$.

Proof Say $L \in \mathbf{MA}$. Then we have an **MA** proof system with perfect completeness and soundness error at most 1/2. Say the message π sent by Merlin has length $p(|x|)$ for some polynomial p. Using error reduction, we can obtain a proof system with perfect completeness and soundness error at most $1/2^{p+1}$; note that the lengths of the messages sent by Merlin do not change (only the lengths of the random coins r used by Arthur increase). So, when $x \in L$ there exists a π (call it π^*) for which $\mathbf{V}(x, \pi^*, r) = 1$ for all r chosen by Arthur, while if $x \notin L$ then for any π sent by Merlin the fraction of r for which Arthur accepts is at most $1/2^{p+1}$. Now simply flip the order of messages: first Arthur will choose r and send it to Merlin, and then Merlin replies with a π and Arthur verifies exactly as before. If $x \in L$ then Merlin has no problem, and can simply send π^*. On

the other hand, if $x \notin L$ then what is the probability that there *exists* a π that will cause Arthur to accept? Well, for any *fixed* π the probability that π will work is at most $1/2^{p+1}$. Taking a union bound over *all* π, we see that the probability that there exists one that works is at most $1/2$. We conclude that $L \in \mathbf{AM}$. ■

As we have said, the same proof shows that an "**MA**" step can be replaced by an "**AM**" step in general. So, $\mathbf{AMA} = \mathbf{AAM} = \mathbf{AM}$ and[2] $\mathbf{MAM} = \mathbf{AMM} = \mathbf{AM}$, and so on. In fact, the above proof technique shows that any Arthur-Merlin proof system with a *constant* number of rounds collapses to exactly **AM** (except for **MA** which may be strictly contained in **AM**). Note that the proof does not extend to proof systems with an arbitrary (non-constant) number of rounds since the communication complexity increases by a multiplicative factor each time an "**MA**" step is replaced by an "**AM**" step (and so if we perform this switch too many times, the communication will no longer be polynomial).

References

[1] M. Furer, O. Goldreich, Y. Mansour, M. Sipser, and S. Zachos. On Completeness and Soundness in Interactive Proof Systems. In *Advances in Computing Research: A Research Annual*, vol. 5 (Randomness and Computation, S. Micali, ed.), 1989. Available at http://www.wisdom.weizmann.ac.il/~oded/papers.html

[2] O. Goldreich. *Modern Cryptography, Probabilistic Proofs, and Pseudorandomness*. Springer-Verlag, 1998.

[3] S. Goldwasser and M. Sipser. Private Coins vs. Public Coins in Interactive Proof Systems. STOC '86.

[2] The theorem shows that $\mathbf{AMA} \subseteq \mathbf{AAM} = \mathbf{AM}$, but the inclusion $\mathbf{AM} \subseteq \mathbf{AMA}$ is trivial.