IAS/PCMI Summer Session 2000
Clay Mathematics Undergraduate Program
Advanced Course on Computational Complexity

Lecture 11: Measuring the Complexity of Proofs

David Mix Barrington and Alexis Maciel
July 31, 2000

1. How Do We Prove a Tautology?

Consider formulas with Boolean variables, the connectives NOT, AND and OR (usually written $\neg$, $\wedge$ and $\vee$), but no quantifiers. For example,

$$F = x \wedge y \to x \vee y$$

is such a formula where the connective $A \to B$ is simply seen as an abbreviation for $\neg A \vee B$. (We are assuming the following precedence rule: $\neg$, then $\wedge$ and $\vee$, then $\to$.) How do you convince someone that this formula is always true, i.e., that it evaluates to true no matter what Boolean values are assigned to the variables? Formulas with this property are called *tautologies*.

One way would be to present that person with a *truth table*:

x	y	$x \wedge y$	$x \vee y$	F
0	0	0	0	1
0	1	0	1	1
1	0	0	1	1
1	1	1	1	1

Now what about the formula

$$F_n = x_1 \wedge x_2 \wedge \cdots \wedge x_n \to x_1 \vee x_2 \vee \cdots \vee x_n?$$

We could again present a truth table for this formula, but that table would have size at least $n2^n$, which is exponential in the size of F_n. So just to look at the entire table requires time exponential in the size of the formula.

Can we do better? For example, can we convince someone that F_n is a tautology by presenting a proof that has size polynomial in the size of F_n? The answer is yes and here is how. First, let us go back to F. Consider the following sequence of steps:

1. $x \to x$
2. $x \wedge y \to x$
3. $x \wedge y \to x \vee y$

It is clear that the first line is a tautology and that this implies that the second line is a tautology too. The third line follows from the second one since it is clear that if $x \wedge y \to x$ is a tautology, then $x \wedge y \to x \vee y$ is one too. What we have here is a more sophisticated type of proof, at least when compared to truth tables. We are relying on accepted *axioms* like $x \to x$ and on *inference rules* such as "if $x \to x$, then $x \wedge y \to x$".

Note that the above proof is convincing even if x and y are replaced by arbitrary formulas. This is because the axiom $A \to A$ and the inference rules "if $A \to B$, then $A \wedge C \to B$" and "if $A \to B$, then $A \to B \vee C$" all make sense even if A and B are replaced by arbitrary formulas. Therefore, what we have is a convincing proof of $A \wedge B \to A \vee B$, where A and B are arbitrary formulas.

Now let $A = x_1$ and $B = x_2 \wedge \cdots \wedge x_n$. Then the above gives us a proof of F_n. And this proof has size $O(n)$, which is much better than the $n2^n$ we got with truth tables.

Let us look a slightly more serious example. Let

$$G = A \vee (B \wedge C) \to (A \vee B) \wedge (A \vee C).$$

This states that $\vee$ distributes over $\wedge$. A proof of G could go as follows:

1. $A \to A$
2. $A \to A \vee B$
3. $B \to B$
4. $B \wedge C \to A \vee B$
5. $A \vee (B \wedge C) \to A \vee B$, from lines 2 and 4
6. $A \to A$
7. $A \to A \vee C$
8. $C \to C$

9. $B \wedge C \to A \vee C$

10. $A \vee (B \wedge C) \to A \vee C$, from lines 7 and 9

11. $A \vee (B \wedge C) \to (A \vee B) \wedge (A \vee C)$, from lines 5 and 10

Now consider the generalized form of this "Distributivity Law":

$$G_n = A \vee (B_1 \wedge \cdots \wedge B_n) \to (A \vee B_1) \wedge \cdots \wedge (A \vee B_n).$$

The following is a proof of G_n:

1. $A \vee (B_1 \wedge \cdots \wedge B_n) \to (A \vee B_1) \wedge (A \vee (B_2 \wedge \cdots \wedge B_n))$, by using the above proof with $B = B_1$ and $C = B_2 \wedge \cdots \wedge B_n$

2. $A \vee (B_1 \wedge \cdots \wedge B_n) \to (A \vee B_1)$, from line 1

3. $A \vee (B_1 \wedge \cdots \wedge B_n) \to A \vee (B_2 \wedge \cdots \wedge B_n)$, from line 1

4. $A \vee (B_2 \wedge \cdots \wedge B_n) \to (A \vee B_2) \wedge \cdots \wedge (A \vee B_n)$ (see below for an explanation of this one)

5. $A \vee (B_1 \wedge \cdots \wedge B_n) \to (A \vee B_2) \wedge \cdots \wedge (A \vee B_n)$, by combining lines 3 and 4

6. G_n, from lines 2 and 5

A proof for line 4 can be constructed recursively, by repeating the process that proved G_n given line 4. At each step in this recursion, line 4 gets simpler until it becomes $A \vee (B_{n-1} \wedge B_n) \to (A \vee B_{n-1}) \wedge (A \vee B_n)$, which we can prove directly. Or, in other words, the existence of a proof of the formula on line 4 can be established by induction on the number of B_i's. Either way, we get a convincing proof of G_n and its size is $O(n^2)$, which once again is much better than the size of the corresponding truth table.

2. The Sequent Calculus

What we have established in the previous two examples are the basic elements of a proof system called the *Sequent Calculus*. Lines in a Sequent Calculus proof are called *sequents* and are formulas of the form $(A_1 \wedge \cdots \wedge A_m) \to (B_1 \vee \cdots \vee B_n)$. Since they always have this general form, sequents are written as $A_1, \ldots, A_m \to B_1, \ldots, B_n$. By convention, the *empty sequent* "$\to$" is considered false, "$\to F$" is considered equivalent to F, and "$F \to$" is considered equivalent to $\neg F$. We will allow unbounded fan-in connectives in our formulas and these will be written as follows: $\wedge(A_1, \ldots, A_n)$ and $\vee(A_1, \ldots, A_n)$.

In the Sequent Calculus, the *axioms*, or *initial sequents*, are "$A \to A$", "$\to \wedge()$" and "$\vee() \to$". The *inference rules* are divided into three groups: the *structural* rules, the *logical* rules, and the *cut* rule. The structural rules are:

Weakening: if $\Gamma_1, \Gamma_2 \to \Delta$, then $\Gamma_1, A, \Gamma_2 \to \Delta$. Similarly on the right.

Permutation: if $\Gamma_1, A, B, \Gamma_2 \to \Delta$, then $\Gamma_1, B, A, \Gamma_2 \to \Delta$. Similarly on the right.

Contraction: if $\Gamma_1, A, A, \Gamma_2 \to \Delta$, then $\Gamma_1, A, \Gamma_2 \to \Delta$. Similarly on the right.

The logical rules are:

NOT-left: if $\Gamma \to A, \Delta$, then $\neg A, \Gamma \to \Delta$.

NOT-right: if $A, \Gamma \to \Delta$, then $\Gamma \to \neg A, \Delta$.

AND-left: if $A_1, \wedge(A_2, \ldots, A_n), \Gamma \to \Delta$, then $\wedge(A_1, \ldots, A_n), \Gamma \to \Delta$.

AND-right: if $\Gamma \to A_1, \Delta$ and $\Gamma \to \wedge(A_2, \ldots, A_n), \Delta$, then $\Gamma \to \wedge(A_1, \ldots, A_n), \Delta$.

OR-left: if $A_1, \Gamma \to \Delta$ and $\vee(A_2, \ldots, A_n), \Gamma \to \Delta$, then $\vee(A_1, \ldots, A_n), \Gamma \to \Delta$.

OR-right: if $\Gamma \to A_1, \vee(A_2, \ldots, A_n), \Delta$, then $\Gamma \to \vee(A_1, \ldots, A_n), \Delta$.

Finally, the cut rule is:

Cut rule: if $A, \Gamma \to \Delta$ and $\Gamma \to A, \Delta$, then $\Gamma \to \Delta$.

In each of these rules and axioms, A and the A_i's can be replaced by arbitrary formulas (in a consistent way). In the rules, the Γ's and Δ's stand for arbitrary sequences of formulas.

A Sequent Calculus *proof* is a finite sequence of lines, each consisting of either an initial sequent or a sequent that can be derived from previous ones by one of the inference rules. A proof of formula F can be either

1. a proof whose last line is "$\to F$",

2. a proof whose last line is the sequent $A_1, \ldots, A_m \to B_1, \ldots, B_n$, if F is of the form $A_1 \wedge \cdots \wedge A_m \to B_1 \vee \cdots \vee B_n$, or

3. a proof that uses "$F \to$" as an additional axiom and whose last line is the empty sequent "$\to$".

This last form of proof is called a *refutation* of $\neg F$. The *size* of a proof is the total number of symbols in it.

It should be clear that Sequent Calculus proofs are convincing, in the sense that they can only prove tautologies. This property is called *soundness* and follows from the fact that each of the axioms and inference rules are themselves sound.

Now does every tautology have a Sequent Calculus proof? The answer is yes and can be illustrated by the proof we presented earlier for the formula $G = A \vee (B \wedge C) \to (A \vee B) \wedge (A \vee C)$. If we read the proof backwards, starting with the bottom line and following each application of the inference rules in the reverse order, towards the axioms, we realize that each inference rule is simplifying the current sequent by eliminating a connective. For example, $A \vee (B \wedge C) \to A \vee B$ is broken down into $A \to A \vee B$ and $B \wedge C \to A \vee B$. Now, except for weakening, each inference rule has a property we can call *reverse soundness*: if its conclusion is a tautology, then each of its hypotheses is a tautology. We can use this to recursively break down the formula we are trying to prove until we end up with sequents that must be axioms. (The details are left as an exercise.) Therefore, we say that the Sequent Calculus is *complete*.

From our first example, we get that the formula $F_n = A_1 \wedge A_2 \wedge \cdots \wedge A_n \to A_1 \vee A_2 \vee \cdots \vee A_n$ has a truth table of size at least $n2^n$ but a Sequent Calculus proof of size $O(n)$. More precisely, what we have here is a sequence of formulas $F_1, F_2, F_3, \ldots$ and a sequence of proofs $P_1, P_2, P_3, \ldots$ such that P_i is a proof of F_i. Equivalently, we can view these sequences as a single (parametrized) formula F_n and a single (parametrized) proof P_n. We measure the complexity of P_n as a function of n. For example, in the case of Sequent Calculus proofs, the *size* of P_n is the total number of symbols in P_n, expressed as a function of n.

The proof of completeness of the Sequent Calculus outlined above actually shows that every formula has a Sequent Calculus proof of size $O(n2^n)$. And we have two examples of tautologies, F_n and G_n, that have Sequent Calculus proofs that are much smaller than that: $O(n)$ and $O(n^2)$, respectively. So we can say that as proof systems, the Sequent Calculus is more powerful than truth tables since the truth table of a formula with n variables is always of size $\Theta(n2^n)$.

But does every tautology have a Sequent Calculus proof of polynomial size? The answer is not known and we will see in the next section that a positive answer would have very interesting consequences.

In the meantime, let us say that the Sequent Calculus is closely related to *Frege* proof systems. As in the Sequent Calculus, a proof in these systems is a sequence of formulas, each of which of is either an axiom or a formula that can be derived from previous ones by an inference rule. A particular Frege proof system is defined by a finite, sound and implicationally complete set of axioms and inference rules. A proof system is *implicationally complete* if whenever $A \to B$ is a tautology, there is a proof of B that uses A as an additional axiom. Note that neither A nor B need to be tautologies themselves. Since for any axiom A, $A \to B$ is a tautology whenever B is, implicational completeness implies completeness.

As in the Sequent Calculus, the axioms and rules of a Frege system are *schemas*, meaning that they can be used with arbitrary formulas. Any two Frege systems are equivalent to each other in the sense that a formula has a polynomial-size proof in one system if and only if it has a polynomial-size proof in the other. The idea is that the logical implication corresponding to a rule in one system can be proved in the other by using implicational completeness. (The details are left as an exercise.) The Sequent Calculus is also equivalent to any Frege system. In contrast, we know from the examples of the previous section that the Sequent Calculus and truth tables are not equivalent.

3. A Formal Notion of Proof System

Now that we have defined the Sequent Calculus, and explored its relationship to truth tables and other Frege systems, let us step back and ask what a proof system is after all. In a proof system, given any string x and a formula F, it should be possible to determine whether x is a proof of F. So the predicate "x is a proof of F" should be computable. But we normally also want proofs to be easy to verify. So we will also insist that this predicate be computable in time polynomial in the length of $\langle x, F \rangle$. For example, whether a table is the correct truth table of a given formula can be determined in time linear in the size of the table. And determining whether a list of sequents constitutes a Sequent Calculus proof of a given formula can be done in time polynomial in the size of the list.

So our formal, abstract notion of a *proof system* will be this: a polynomial-time computable predicate $R(x, F)$, where x is any string and F is any formula. The meaning of R is that x is a proof of F, but it is R that defines what a proof is. The proof system is *complete* if for every tautology F, there is x such that $R(x, F) = 1$

(and then we say that x is a proof of F). The proof system is *sound* if F is tautology whenever there is x such that $R(x, F)$.

In the previous section, we said that any two Frege systems are equivalent in the sense that a formula has a polynomial-size proof in one system if and only if it has a polynomial-size proof in the other. Now we add the condition that proofs in one system can be efficiently translated into proofs in the other. We say that a proof system $\mathcal{S}$ *polynomially simulates* (P-simulates) a proof system $\mathcal{T}$ if there is a polynomial-time computable function f such that for every pair $\langle x, F\rangle$, x is a proof of F in $\mathcal{T}$ if and only if $f(x)$ is a proof of F in $\mathcal{S}$. (Notice the similarity to the notion of polynomial-time reducibility.) We then say that two systems are polynomially equivalent (P-equivalent) if they P-simulate each other. The Sequent Calculus P-simulates truth tables, but truth tables do not P-simulate the Sequent Calculus. All Frege systems are P-equivalent to each other.

Now let us go back to a question we asked earlier: does every tautology have a Sequent Calculus proof of polynomial size? We can define the *complexity* of a proof system as the following function of n: the maximum over all tautologies F of size n of the minimum length of a proof of F,

$$\max_{F:|F|=n} \; \min_{P:P \text{ is a proof of } F} |P|.$$

For example, the truth tables proof system has $\Theta(n2^n)$ complexity. Proof systems of complexity $n^{O(1)}$ are called *polynomially bounded*. Asking whether there is a polynomial-size Sequent Calculus proof of every tautology is the same as asking whether the Sequent Calculus is polynomially bounded.

There is currently no proof system known to be polynomially bounded. And if one was found, then this would imply that NP is closed under complement. One way to show that P $\neq$ NP is to show that NP is *not* closed under complement. This is because we know that P *is* closed under complement, so P and NP could not possibly be the same class. Therefore, if a single polynomially bounded proof system was found, then this approach to showing that P $\neq$ NP would be invalidated.

Theorem 1 *There is a polynomially bounded (sound and complete) proof system if and only if* NP *is closed under complement.*

Proof Suppose that NP is closed under complement. Consider the language TAUT consisting of all tautologies: TAUT $= \{F : F \text{ is a tautology}\}$. The complement of TAUT is in NP. Therefore, by hypothesis, TAUT is in NP. This implies that TAUT has a polynomial-time verifier V. Consider the predicate R computed by V. If $F \notin$

TAUT, then $R(P, F) = 0$ for every P. If $F \in$ TAUT, then there is P of length at most $|F|^{O(1)}$ such that $R(P, F) = 1$. Therefore, R is a polynomially bounded proof system.

Now suppose that there is a polynomially bounded proof system R. Then the polynomial-time machine that decides R constitutes a polynomial-time verifier for TAUT. Therefore, TAUT is in NP. If a formula F is satisfiable, then its negation $\neg F$ is not a tautology, and vice-versa. Therefore, SAT $\leq_{\mathrm{P}}$ $\overline{\text{TAUT}}$, which implies that $\overline{\text{TAUT}}$ is NP-complete. Now suppose that $A \in$ NP. Then $A \leq_{\mathrm{P}}$ $\overline{\text{TAUT}}$, which implies that $\overline{A} \leq_{\mathrm{P}}$ TAUT. Therefore, $\overline{A} \in$ NP since TAUT $\in$ NP. This shows that NP is closed under complement. □

In the previous proof, we showed that TAUT $\in$ NP if and only if there is a polynomially bounded proof system. This leads to a more general idea. Say that R is a *proof system for a language* L if R is computable in polynomial time and if $w \in L$ if and only if there is a string P such that $R(P, w) = 1$. Then NP is exactly the class of languages that have polynomially bounded proof systems.

Why study the complexity proof systems? Here are three reasons. First, to answer some very natural and fundamental questions such as "What is a proof?", "What is an efficient proof system?" and "Do efficient proofs systems exist?".

Second, because of the connections to the P versus NP question we just explained. If we could show that there are no polynomially bounded proof systems, then we would have shown that P $\neq$ NP. This gives us the following approach for proving that P $\neq$ NP: start with proof systems known not to be polynomially bounded (like truth tables) and prove that increasingly more powerful proof systems are not polynomially bounded.

Third, consider the problem of finding a proof for a given tautology. Typically, the weaker the proof system, the more efficient the proof search algorithm. So relatively weak proof systems are studied in the context of automated theorem proving and understanding their exact power (i.e., the class of tautologies that have polynomial-size proofs in these systems) is therefore important.

4. Exercises

1. Consider a Frege system with only one rule "if P and $P \to Q$, then Q" (modus ponens) and with the following ten axioms:

   (a) $(P \wedge Q) \to P$

(b) $(P \wedge Q) \rightarrow Q$

(c) $P \rightarrow (P \vee Q)$

(d) $Q \rightarrow (P \vee Q)$

(e) $\neg\neg P \rightarrow P$

(f) $(P \rightarrow Q) \rightarrow [(P \rightarrow \neg Q) \rightarrow \neg P]$

(g) $P \rightarrow (Q \rightarrow P \wedge Q)$

(h) $(P \rightarrow R) \rightarrow [(Q \rightarrow R) \rightarrow (P \vee Q \rightarrow R)]$

(i) $P \rightarrow (Q \rightarrow P)$

(j) $(P \rightarrow Q) \rightarrow [(P \rightarrow [Q \rightarrow R]) \rightarrow (P \rightarrow R)]$

Give Sequent Calculus proofs of each of these axioms. Then give a sequent calculus proof that simulates the modus ponens rule, in the sense that it proves the sequent Q using the sequents $\rightarrow P$ and $\rightarrow (P \rightarrow Q)$ as additional axioms.

2. Provide the details of a proof that the Sequent Calculus is complete.

3. Show that the Sequent Calculus is implicationally complete.

4. In the version of the Sequent Calculus we defined, there are three possible ways to prove a formula F. Show that they are interchangeable by proving the following:

   (a) For any tautology F, a direct proof of F can be transformed in polynomial time into a refutation of $\neg F$, and vice-versa.

   (b) If F is of the form $A_1 \wedge \cdots \wedge A_m \rightarrow B_1 \vee \cdots \vee B_n$, then a proof whose last line is the sequent $\rightarrow F$ can be transformed in polynomial time into a proof whose last line is the sequent $A_1, \ldots, A_m \rightarrow B_1, \ldots, B_n$.

5. Show that any two Frege systems are P-equivalent.

6. Show that the Sequent Calculus is P-equivalent to any Frege system.

7. Show that a proof system for TAUT is also a proof system for the complement of SAT.

8. Show that a language is in NP if and only if it has a polynomially bounded proof system.

9. Suppose that $\mathcal{S}$ P-simulates $\mathcal{T}$. Show that the class of tautologies $\mathcal{T}$ can prove in polynomial size is contained in the class of tautologies $\mathcal{S}$ can prove in polynomial size. Show that if $\mathcal{T}$ is polynomially bounded, then so is $\mathcal{S}$.