Luca Trevisan April 14, 2010

Today we prove the Valiant-Vazirani theorem.

Theorem 1 (Valiant-Vazirani) *Suppose there is a polynomial time algorithm that on input a CNF formula having exactly one satisfying assignment finds that assignment. (We make no assumption on the behaviour of the algorithm on other inputs.) Then* $\mathbf{NP} = \mathbf{RP}$.

1 The Valiant-Vazirani Theorem

As discussed in the last lecture, our approach is the following: given a satisfiable formula ϕ and a number k such that ϕ has roughly 2^k satisfying assignments, we pick a random hash function $h : \{0,1\}^n \to \{0,1\}^{k+2}$ from a family of pairwise independent hash functions, and we construct a formula $\psi(x)$ which is equivalent to $\phi(x) \wedge (h(x) = 0)$. With constant probability, ψ has precisely one satisfying assignment, and so we can pass it to our hypothetical algorithm, which finds a satisfying assignment for ψ and hence a satisfying assignment for ϕ.

If we are only given ϕ, we can try all possible values of k between 0 and n (where n is the number of variables in ϕ), and run the above procedure for each k. When the correct value of k is chosen, we have a constant probability of finding a satisfying assignment for ϕ.

Once we have a randomized algorithm that, given a satisfiable formula, finds a satisfying assignment with constant probability, we have an **RP** algorithm for 3SAT: run the assignment-finding algorithm, accept if it finds a satisfying assignment and reject otherwise. The existence of an **RP** algorithm for 3SAT implies that $\mathbf{NP} \subseteq \mathbf{RP}$ because **RP** is closed under many-to-one reductions, and so $\mathbf{RP} = \mathbf{NP}$ because we have $\mathbf{RP} \subseteq \mathbf{NP}$ by definition.

The main calculation that we need to perform is to show that if we have a set of size roughly 2^k, and we hash its elements pairwise independently to $\{0,1\}^{k+2}$, then there is a constant probability that exactly one element is hashed to $(0, \ldots, 0)$.

Lemma 2 *Let* $T \subseteq \{0,1\}^n$ *be a set such that* $2^k \leq |T| < 2^{k+1}$ *and let* H *be a family of pairwise independent hash functions of the form* $h : \{0,1\}^n \to \{0,1\}^{k+2}$. *Then if we pick* h *at random from* H, *there is a constant probability that there is a unique element* $x \in T$ *such that* $h(x) = \mathbf{0}$. *Precisely,*

$$\mathop{\mathbb{P}}_{h \in H}[|\{x \in T : h(x) = \mathbf{0}\}| = 1] \geq \frac{1}{8}$$

Proof: Let us fix an element $x \in T$. We want to compute the probability that x is the *unique* element of T mapped into $\mathbf{0}$ by h. Clearly,

$$\mathop{\mathbb{P}}_h[h(x) = \mathbf{0} \wedge \forall y \in T - \{x\}.h(y) \neq \mathbf{0}] = \mathop{\mathbb{P}}_h[h(x) = \mathbf{0}] \cdot \mathop{\mathbb{P}}_h[\forall y \in T - \{x\}.h(y) \neq \mathbf{0} | h(x) = \mathbf{0}]$$

and we know that

$$\mathop{\mathbb{P}}_h[h(x) = \mathbf{0}] = \frac{1}{2^{k+2}}$$

The difficult part is to estimate the other probability. First, we write

$$\mathop{\mathbb{P}}_h[\forall y \in T - \{x\}.h(y) \neq \mathbf{0} | h(x) = \mathbf{0}] = 1 - \mathop{\mathbb{P}}_h[\exists y \in T - \{x\}.h(y) = \mathbf{0} | h(x) = \mathbf{0}]$$

And then observe that

$$\begin{aligned} & \mathop{\mathbb{P}}_h[\exists y \in T - \{x\}.h(y) = \mathbf{0} | h(x) = \mathbf{0}] \\ \leq & \sum_{y \in |T| - \{x\}} \mathop{\mathbb{P}}_h[h(y) = \mathbf{0} | h(x) = \mathbf{0}] \\ = & \sum_{y \in |T| - \{x\}} \mathop{\mathbb{P}}_h[h(y) = \mathbf{0}] \\ = & \frac{|T| - 1}{2^{k+2}} \\ \leq & \frac{1}{2} \end{aligned}$$

Notice how we used the fact that the value of $h(y)$ is independent of the value of $h(x)$ when $x \neq y$.

Putting everything together, we have

$$\mathop{\mathbb{P}}_h[\forall y \in T - \{x\}.h(y) \neq \mathbf{0} | h(x) = \mathbf{0}] \geq \frac{1}{2}$$

and so

$$\mathop{\mathbb{P}}_h[h(x) = \mathbf{0} \wedge \forall y \in T - \{x\}.h(y) \neq \mathbf{0}] \geq \frac{1}{2^{k+3}}$$

To conclude the argument, we observe that the probability that there is a unique element of T mapped into $\mathbf{0}$ is given by the sum over $x \in T$ of the probability that x is the unique element mapped into $\mathbf{0}$ (all this events are disjoint, so the probability of their union is the sum of the probabilities). The probability of a unique element mapped into $\mathbf{0}$ is then at least $|T|/2^{k+3} > 1/8$. $\Box$

Lemma 3 *There is a probabilistic polynomial time algorithm that on input a CNF formula ϕ and an integer k outputs a formula ψ such that*

- *If ϕ is unsatisfiable then ψ is unsatisfiable.*
- *If ϕ has at least 2^k and less than 2^{k+1} satifying assignments, then there is a probability at least $1/8$ then the formula ψ has exactly one satisfying assignment.*

PROOF: Say that ϕ is a formula over n variables. The algorithm picks at random vectors $a_1, \ldots, a_{k+2} \in \{0,1\}^n$ and bits $b_1, \ldots, b_{k+2}$ and produces a formula ψ that is equivalent to the expression $\phi(x) \wedge (a_1 \cdot x + b_1 = 0) \wedge \ldots \wedge (a_{k+2} \cdot x + b_{k+2} = 0)$. Indeed, there is no compact CNF expression to compute $a \cdot x$ if a has a lot of ones, but we can proceed as follows: for each i we add auxiliary variables $y_1^i, \ldots, y_n^i$ and then write a CNF condition equivalent to $(y_1^i = x_1 \wedge a_i[1]) \wedge \cdots \wedge (y_n^i = y_{n-1}^i \oplus (x_n \wedge a_i[n] \oplus b_i)))$. Then ψ is the AND of the clauses in ϕ plus all the above expressions for $i = 1, 2, \ldots, k+2$.

By construction, the number of satisfying assignments of ψ is equal to the number of satisfying assignments x of ϕ such that $h_{a_1,\ldots,a_{k+2},b_1,\ldots,b_{k+2}}(x) = \mathbf{0}$. If ϕ is unsatisfiable, then, for every possible choice of the a_i, ψ is also unsatisfiable.

If ϕ has between 2^k and 2^{k+1} assignments, then Lemma 2 implies that with probability at least $1/8$ there is exactly one satisfying assignment for ψ. $\Box$

We can now prove the Valiant-Vazirani theorem.

PROOF:[Of Theorem 1] It is enough to show that, under the assumption of the Theorem, 3SAT has an **RP** algorithm.

On input a formula ϕ, we construct formulae $\psi_0, \ldots, \psi_n$ by using the algorithm of Lemma 3 with parameters $k = 0, \ldots, n$. We submit all formulae $\psi_0, \ldots, \psi_n$ to the algorithm in the assumption of the Theorem, and accept if the algorithm can find a satisfying assignment for at least one of the formulae. If ϕ is unsatisfiable, then all the formulae are always unsatisfiable, and so the algorithm has a probability zero of accepting. If ϕ is satisfiable, then for some k it has between 2^k and 2^{k+1} satisfying assignments, and there is a probability at least $1/8$ that ψ_k has exactly one satisfying assignment and that the algorithm accepts. If we repeat the above procedure t times, and accept if at least one iteration accepts, then if ϕ is unsatisfiable we still have probability zero of accepting, otherwise we have probability at least $1 - (7/8)^t$ of accepting, which is more than $1/2$ already for $t = 6$. $\Box$