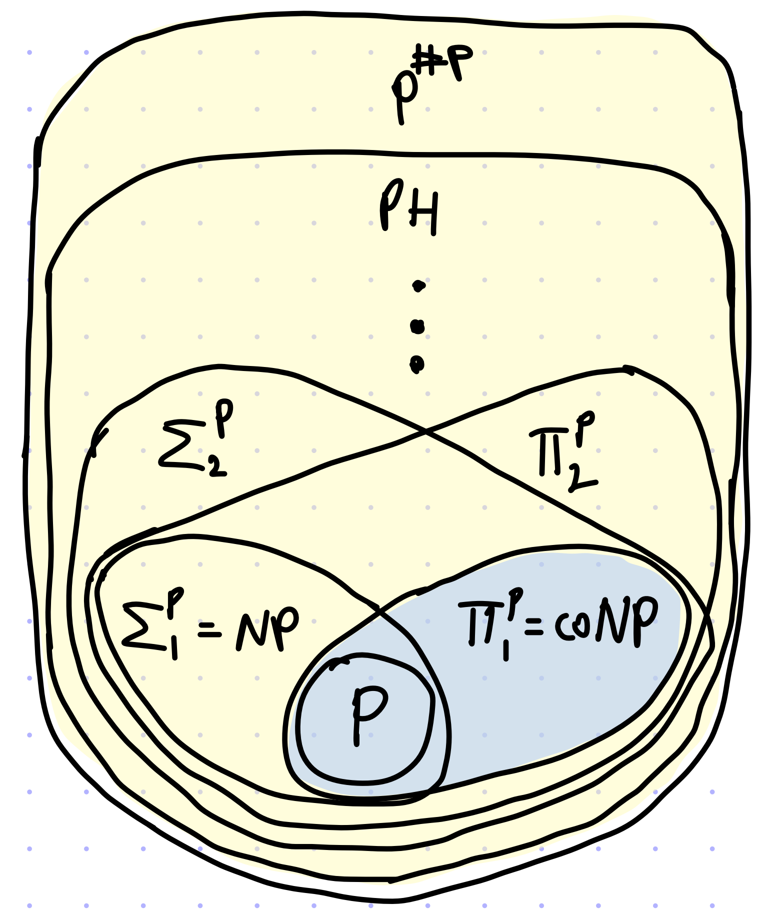
Lecture A.2

Sumcheck Protocol

Summer Graduate School on
Foundations and Frontiers of Probabilistic Proofs
2021.07.27

Interactive Proofs for Counting Problems

We saw an interactive proof for GNI, a problem in coNP not known to be in P.
Yet, GNI is not believed to be coNP-complete. [If so, PH collapses to 2nd level.]



theorem: $UNSAT \in IP$, so $coNP \subseteq IP$

theorem: $\#SAT \in IP$, so $P^{\#P} \subseteq IP$

These results should be surprising:

- many languages beyond NP!
- the interactive proof for GNI leveraged properties of graph isomorphisms, but UNSAT and #SAT do not seem to have similar properties

⇒ we will learn new ideas: arithmetization, sumcheck protocol

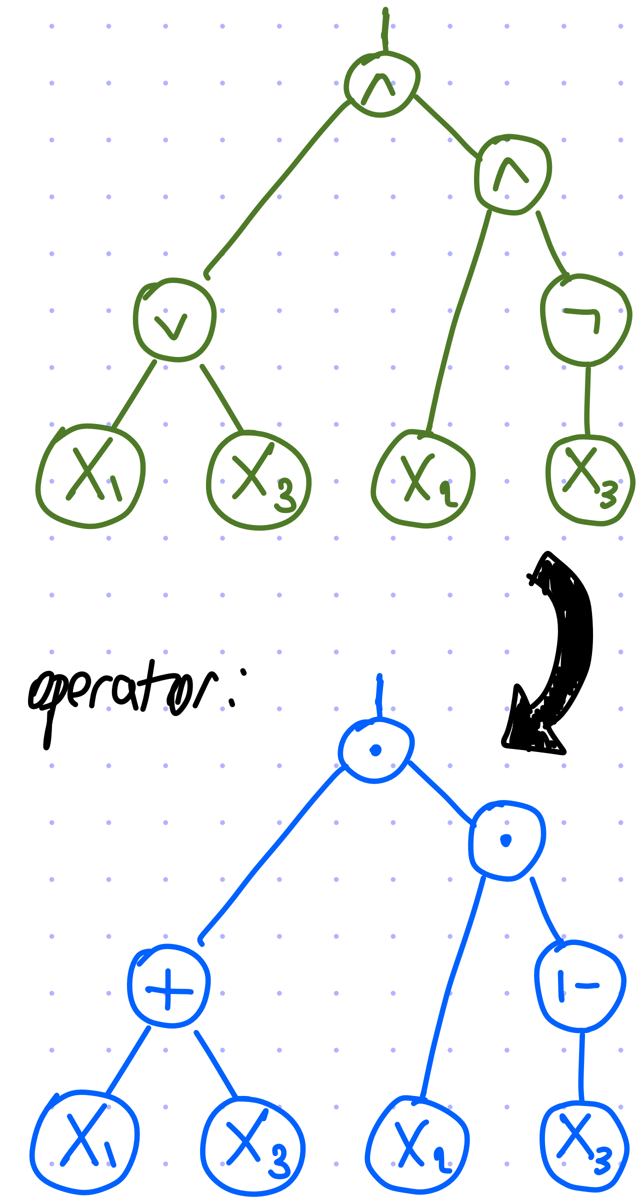
[$P^{\#P}$ = languages decidable in polynomial time via a machine with a #SAT oracle]

Arithmetization of a Boolean Formula

A boolean formula $\phi(x_1,\dots,x_n)$ is a tree where:

- every leaf node is labeled by a variable x_i;
- every internal node is a logical operator on its children.

$(\vee, \wedge, \neg)$



Arithmetization replaces each logical operator with an arithmetic operator:

$\neg x \mapsto 1-x \qquad x \wedge y \mapsto x\cdot y \qquad x \vee y \mapsto x+y$

Thus a boolean formula $\phi(x_1,\dots,x_n)$ is mapped to a polynomial $p(x_1,\dots,x_n)$ such that $\deg_{tot}(p) \leq |\phi|$, evaluating p at a point takes $|\phi|$ operations, and:

claim: Let ϕ be a 3CNF with m clauses. Then:

- $\phi \in \text{UNSAT} \Rightarrow \sum_{a_1,\dots,a_n \in \{0,1\}} p(a_1,\dots,a_n) = 0$
- $\phi \notin \text{UNSAT} \Rightarrow 0 < \sum_{a_1,\dots,a_n \in \{0,1\}} p(a_1,\dots,a_n) \leq 2^n \cdot 3^m$

$\longrightarrow$

corollary:

$\forall$ prime $q > 2^n 3^m$

$$\phi \in \text{UNSAT} \Updownarrow \sum_{a_1,\dots,a_n \in \{0,1\}} p(a_1,\dots,a_n) = 0 \bmod q$$

Sumcheck Protocol

statement

$$\sum_{\alpha_1,\ldots,\alpha_n \in H} p(\alpha_1,\ldots,\alpha_n) = \gamma$$

d is a bound on the individual degree of p

$P(\mathbb{F}, H, n, \gamma, p)$ $\qquad$ $V^{p}(\mathbb{F}, H, n, \gamma, d)$

Prover		Verifier
$p_1(X) := \sum_{\alpha_2,\ldots,\alpha_n \in H} p(X,\alpha_2,\ldots,\alpha_n)$	$\xrightarrow{p_1 \in \mathbb{F}[X]}$	$\sum_{\alpha_1 \in H} p_1(\alpha_1) \stackrel{?}{=} \gamma$
	$\xleftarrow{\omega_1 \in \mathbb{F}}$	$w_1 \leftarrow \mathbb{F}$
$p_2(X) := \sum_{\alpha_3,\ldots,\alpha_n \in H} p(w_1,X,\alpha_3,\ldots,\alpha_n)$	$\xrightarrow{p_2 \in \mathbb{F}[X]}$	$\sum_{\alpha_2 \in H} p_2(\alpha_2) \stackrel{?}{=} p_1(w_1)$
	$\xleftarrow{w_2 \in \mathbb{F}}$	$w_2 \leftarrow \mathbb{F}$
	$\vdots$	$\vdots$
$p_n(X) := p(w_1,\ldots,w_{n-1},X)$	$\xrightarrow{p_n \in \mathbb{F}[X]}$	$\sum_{\alpha_n \in H} p_n(\alpha_n) = p_{n-1}(w_{n-1})$
	$\xleftarrow{\omega_n \in \mathbb{F}}$	$w_n \leftarrow \mathbb{F}$
		$p(w_1,\ldots,w_n) \stackrel{?}{=} p_n(w_n)$
	$O(n \cdot d)$ elements	$O(n \cdot \lvert H \rvert \cdot d)$ fops + 1 eval of p

claim: if statement is true then verifier accepts w.p. 1

Soundness of Sumcheck Protocol

can improve to $1-\left(1-\frac{d}{|\mathbb{F}|}\right)^n$

<u>claim</u>: if $\sum_{\alpha_1,\ldots,\alpha_n\in H} p(\alpha_1,\ldots,\alpha_n)\neq\gamma$ then $\Pr[\text{verifier accepts}]\leq\frac{n\cdot d}{|\mathbb{F}|}$

<u>proof</u>: Fix a malicious prover, described via n polynomials $\tilde{p}_1,\ldots,\tilde{p}_n\in\mathbb{F}[X]$ such that $\tilde{p}_i$ depends on the verifiers messages $\omega_1,\ldots,\omega_{i-1}\in\mathbb{F}$.

Define: $\forall i\in[n]$ $E_i :=$ "event that $\tilde{p}_i\equiv p_i$", $W =$ "event that verifier accepts".

<u>lemma</u>: For $j=n,n-1,\ldots,1$: $\Pr[W]\leq\frac{(n-j+1)\cdot d}{|\mathbb{F}|}+\Pr[W\mid E_j\wedge\ldots\wedge E_n]$.

This suffices to prove the claim because if we set $j:=1$ then we get:

$$\Pr[W]\leq\frac{n\cdot d}{|\mathbb{F}|}+\underbrace{\Pr[W\mid E_1\wedge\ldots\wedge E_n]}_{\leq\Pr[W\mid E_1]=0\text{ because }\sum_{\alpha_1\in H}\tilde{p}_1(\alpha_1)=\sum_{\alpha_1\in H}p_1(\alpha_1)\neq\gamma}$$

$$=\frac{n\cdot d}{|\mathbb{F}|}+0$$

We are left to prove the lemma.

<u>Lemma</u>: For $j = n, n-1, \ldots, 1$: $\Pr[W] \leq \frac{(n-j+1)\cdot d}{|\mathbb{F}|} + \Pr[W \mid E_j \wedge \cdots \wedge E_n]$.

Proof is by induction on j.

<u>Base case</u>: $j = n$

$$\Pr[W] \leq \underbrace{\Pr[W \mid \overline{E_n}]}_{} + \Pr[W \mid E_n] \leq \frac{d}{|\mathbb{F}|} + \Pr[W \mid E_n]$$

$$= \Pr[V \text{ accepts} \mid \tilde{p}_n \neq p_n]$$

$$\leq \Pr[\tilde{p}_n(w_n) = p(w_1, \ldots, w_n) \mid \tilde{p}_n \neq p_n]$$

$$= \Pr[\tilde{p}_n(w_n) = p_n(w_n) \mid \tilde{p}_n \not\equiv p_n] \leq$$

Polynomial Identity Lemma

$\forall$ non-zero $f \in \mathbb{F}[X]$, $\Pr_{\alpha \leftarrow S}[f(\alpha) = 0] \leq \frac{\deg(f)}{|S|}$

<u>Inductive case</u>:

$$\Pr[W] \leq \frac{(n-j+1)\cdot d}{|\mathbb{F}|} + \Pr[W \mid E_j \wedge \cdots \wedge E_n]$$

assume for $j \in \{n, n-1, \ldots, 2\}$

$$\leq \frac{(n-j+1)\cdot d}{|\mathbb{F}|} + \Pr[W \mid \overline{E_{j-1}} \wedge E_j \wedge \cdots \wedge E_n] + \Pr[W \mid E_{j-1} \wedge E_j \wedge \cdots \wedge E_n]$$

$$\leq \Pr[\tilde{p}_{j-1}(w_{j-1}) = \sum_{\alpha_j \in H} \tilde{p}_j(\alpha_j) \mid \tilde{p}_{j-1} \neq p_{j-1}, \tilde{p}_j \equiv p_j]$$

$$\leq \Pr[\tilde{p}_{j-1}(w_{j-1}) = \sum_{\alpha_j \in H} p_j(\alpha_j) \mid \tilde{p}_{j-1} \not\equiv p_{j-1}]$$

$$= \Pr[\tilde{p}_{j-1}(w_{j-1}) = p_{j-1}(w_{j-1}) \mid \tilde{p}_{j-1} \not\equiv p_{j-1}]$$

$$\leq \frac{(n-j+1)\cdot d}{|\mathbb{F}|} + \frac{d}{|\mathbb{F}|} + \Pr[W \mid E_{j-1} \wedge \cdots \wedge E_n]$$

proved for $j-1$

$$\leq \frac{(n-(j-1)+1)\cdot d}{|\mathbb{F}|} + \Pr[W \mid E_{j-1} \wedge \cdots \wedge E_n]$$

Interactive Proof for UNSAT (which shows that $coNP \subseteq IP$)

$P(\phi)$

the 3CNF ϕ is unsatisfiable

$V(\phi)$

$\xrightarrow{q \in \mathbb{N}}$

$2^n 3^m < q < 2^{poly(m,n)}$

$q \in$ PRIMES [probabilistic test suffices]

$p := \mathrm{ARITH}(\phi, \mathbb{F}_q)$

$p := \mathrm{ARITH}(\phi, \mathbb{F}_q)$

$P_{SC}(\mathbb{F}_q, \{0,1\}, n, 0, p)$

sumcheck protocol

$$\sum_{\alpha_1,\dots,\alpha_n \in \{0,1\}} p(\alpha_1,\dots,\alpha_n) \stackrel{?}{=} 0$$

$V_{SC}^{p}(\mathbb{F}_q, \{0,1\}, n, 0)$

$(\omega_1,\dots,\omega_n) \in \mathbb{F}_q^n$

$p(\omega_1,\dots,\omega_n) \in \mathbb{F}_q$

evaluate p at $(\omega_1,\dots,\omega_n)$ in $poly(m,n)$ time

The soundness error is:

$$\frac{n \cdot \deg_{ind}(p)}{q} \le \frac{n \cdot \deg_{tot}(p)}{q} \le \frac{n \cdot m}{q} < \frac{n \cdot m}{2^n \cdot 3^m} \le \frac{1}{2}$$

Arithmetization for #SAT

The arithmetization we used for UNSAT was coarse:

$$\forall (a_1,\ldots,a_n)\in\{0,1\}^n \quad \begin{array}{l} \phi(a_1,\ldots,a_n)=0 \;\rightarrow\; p(a_1,\ldots,a_n)=0 \\ \phi(a_1,\ldots,a_n)=1 \;\rightarrow\; 0<p(a_1,\ldots,a_n)\le 3^m \end{array}$$

We can modify the arithmetization to be more precise:

$$\neg X \mapsto 1-X \qquad X\wedge Y \mapsto X\cdot Y \qquad X\vee Y \mapsto X+Y-X\cdot Y$$

The new arithmetization satisfies:

<u>claim</u>: $\forall (a_1,\ldots,a_n)\in\{0,1\}^n \quad \begin{array}{l} \phi(a_1,\ldots,a_n)=0 \rightarrow p(a_1,\ldots,a_n)=0 \\ \phi(a_1,\ldots,a_n)=1 \rightarrow p(a_1,\ldots,a_n)=1 \end{array}$

p is a <u>low-degree extension</u> of ϕ because $p|_{\{0,1\}^n}\equiv\phi$

We can now reduce #SAT to a sumcheck problem:

<u>corollary</u>: $\forall$ prime $q>2^n$ $\quad \#\phi=c \iff \sum_{a_1,\ldots,a_n\in\{0,1\}} p(a_1,\ldots,a_n)=c \mod q$

Interactive Proof for #SAT

(which shows that $P^{\#P} \subseteq IP$)

$P(\phi, c)$ $\quad$ $\#\phi = c$ $\quad$ $V(\phi, c)$

$\xrightarrow{q \in \mathbb{N}}$

$2^n < q < 2^{poly(m,n)}$

$q \in PRIMES$

$p := ARITH^{\#}(\phi, \mathbb{F}_q)$ $\quad$ $p := ARITH^{\#}(\phi, \mathbb{F}_q)$

$P_{sc}(\mathbb{F}_q, \{0,1\}, n, c, p)$

sumcheck protocol

$$\sum_{\alpha_1, \dots, \alpha_n \in \{0,1\}} p(\alpha_1, \dots, \alpha_n) \stackrel{?}{=} c$$

$V_{sc}^{p}(\mathbb{F}_q, \{0,1\}, n, c)$

$(w_1, \dots, w_n) \in \mathbb{F}_q^n$ $\quad$ $p(w_1, \dots, w_n) \in \mathbb{F}_q$

evaluate p at $(w_1, \dots, w_n)$ in $poly(m,n)$ time

⊛ If M makes N oracle calls then error of each sumcheck must be $O(\frac{1}{N})$ so we choose q such that $\frac{n \cdot 3m}{q} \le \frac{1}{2} \cdot \frac{1}{N}$. (Else repeat IP $O(\log N)$ times.)

Let $L \in P^{\#P}$, and let M be a machine that decides L with a #SAT oracle.

Here is the IP for L

$V(x)$

simulate M on x and ask prover for help on #SAT calls

ϕ $\quad$ c

IP for $\#\phi = c$

$P(x)$ ⊛