Freivalds Trick and Schwartz-Zippel Lemma

Lecture 12
October 8, 2019
Outline

- Freivald’s randomized algorithm to verify matrix multiplication
- Generalization to polynomial identity testing via Schwartz-Zippel Lemma
Part I

Freivalds Algorithm for checking Matrix Multiplication
Verifying Matrix Multiplication

Problem

Given three $n \times n$ matrices $A, B, C$, is $AB = C$?
Verifying Matrix Multiplication

Problem

Given three $n \times n$ matrices $A$, $B$, $C$, is $AB = C$?

Naive algorithm: compute $D = AB$ and check if $D = C$

Running time: $T(n) + n^2$ time where $T(n)$ is time to multiply $n \times n$ matrices. Current best bound on $T(n)$ is $n^{2.3728}$.

Question: Can we do better with randomization?
Freivald’s Algorithm

- Pick random vector \( r \in \{0, 1\}^n \) — each coordinate is independent and uniform over \( \{0, 1\} \).
- Output YES if \( ABr = Cr \) and NO otherwise.

Theorem

If \( AB = C \) the algorithm outputs YES with probability 1. If \( AB \neq C \) algorithm outputs YES with probability at most \( \frac{1}{2} \).

Repeating \( k \) times the probability of error is \( \leq \frac{1}{2^k} \) and running time is \( O(kn^2) \).
Freivald’s Algorithm

- Pick random vector \( r \in \{0, 1\}^n \) — each coordinate is independent and uniform over \( \{0, 1\} \).
- Output YES if \( ABr = Cr \) and NO otherwise.

Running time: \( O(n^2) \)
Freivald’s Algorithm

- Pick random vector $r \in \{0, 1\}^n$ — each coordinate is independent and uniform over $\{0, 1\}$.
- Output YES if $ABr = Cr$ and NO otherwise.

Running time: $O(n^2)$

**Theorem**

If $AB = C$ the algorithm outputs YES with probability 1. If $AB \neq C$ algorithm outputs YES with probability at most $1/2$.

Repeating $k$ times the probability of error is $\leq 1/2^k$ and running time is $O(kn^2)$. 
Proof

Lemma

Let \( x, y \) be two vectors in \( F^n \) for some field/ring \( F \). If \( r \in \{0, 1\}^n \) is a random vector then \( \Pr[x^t r = y^t r \mid x \neq y] \leq 1/2 \).
Proof

Lemma

Let \( x, y \) be two vectors in \( \mathbb{F}^n \) for some field/ring \( \mathbb{F} \). If \( r \in \{0, 1\}^n \) is a random vector then \( \Pr[x^t r = y^t r \mid x \neq y] \leq 1/2 \).

- Assume \( x \neq y \). Without loss of generality \( x_1 \neq y_1 \).
- Fix \( r_2, r_3, \ldots, r_n \). Let \( \alpha = x_2 r_2 + \ldots + x_n r_n \) and \( \beta = y_2 r_2 + \ldots + y_n r_n \).
- If \( r_1 \) is uniformly random in \( \{0, 1\} \) what is \( \Pr[x_1 r_1 + \alpha = y_1 r_1 + \beta] \)?
Proof

**Lemma**

Let $\mathbf{x}, \mathbf{y}$ be two vectors in $\mathbb{F}^n$ for some field/ring $\mathbb{F}$. If $r \in \{0, 1\}^n$ is a random vector then $\Pr[\mathbf{x}^t r = \mathbf{y}^t r \mid \mathbf{x} \neq \mathbf{y}] \leq 1/2$.

- Assume $\mathbf{x} \neq \mathbf{y}$. Without loss of generality $x_1 \neq y_1$.
- Fix $r_2, r_3, \ldots, r_n$. Let $\alpha = x_2 r_2 + \ldots + x_n r_n$ and $\beta = y_2 r_2 + \ldots + y_n r_n$.
- If $r_1$ is uniformly random in $\{0, 1\}$ what is $\Pr[x_1 r_1 + \alpha = y_1 r_1 + \beta]$? At most 1/2. If $\alpha \neq \beta$ then $r_1 = 0$ will distinguish and if $\alpha = \beta$ then $r_1 = 1$ will distinguish.
- Holds for any fixed $r_2, \ldots, r_n$ and $r_1$ random. Since $r_1$ is independent of $r_2, \ldots, r_n$, holds for random $r$. 
Proof of Theorem

**Theorem**

If $AB = C$ the algorithm outputs YES with probability 1. If $AB \neq C$ algorithm outputs YES with probability at most $1/2$.

Suppose $AB \neq C$. Let $D = AB$. Since $D \neq C$ there is some $i, j$ where $D_{i,j} \neq C_{i,j}$. Assume wlog that $i, j = 1, 1$.

Let $x$ be first row of $D$ and $y$ be first row of $C$. $x \neq y$. Apply preceding lemma.
Part II

Polynomial Identity Testing and Schwartz-Zippel Lemma
Polynomials

Definition

A (univariate) polynomial over a field $\mathbb{F}$ is a finite sum of terms of the form $a_i x^i$ where $a_i \in \mathbb{F}$ and $x$ is a variable.

$$p(x) = \sum_{j=0}^{n-1} a_j x^j$$

The numbers $a_0, a_1, \ldots, a_n$ are the coefficients of the polynomial. The degree is the highest power of $x$ with a non-zero coefficient.

Example

$$p(x) = 3 - 4x + 5x^3$$

$a_0 = 3, a_1 = -4, a_2 = 0, a_3 = 5$ and $\text{deg}(p) = 3$
A \textit{(univariate) polynomial} over a field $\mathbb{F}$ is a finite sum of terms of the form $a_i x^i$ where $a_i \in \mathbb{F}$ and $x$ is a variable.

$$p(x) = \sum_{j=0}^{n-1} a_j x^j$$

The numbers $a_0, a_1, \ldots, a_n$ are the \textit{coefficients} of the polynomial. The \textit{degree} is the highest power of $x$ with a non-zero coefficient.

Polynomials represented by vector $a = (a_0, a_1, \ldots, a_{n-1})$ of coefficients.
Polynomial Identity Testing

Definition

A polynomial $p(x)$ is identically 0 if $p(x) = 0$ for all $x$. Equivalently it corresponds to all coefficients being 0.
Definition
A polynomial $p(x)$ is identically 0 if $p(x) = 0$ for all $x$. Equivalently it corresponds to all coefficients being 0.

Question (PIT)
Given a polynomial $p$ in some implicit fashion, is $p$ identically 0?
Polynomial Identity Testing

**Definition**
A polynomial $p(x)$ is identically 0 if $p(x) = 0$ for all $x$. Equivalently it corresponds to all coefficients being 0.

**Question (PIT)**
Given a polynomial $p$ in some implicit fashion, is $p$ identically 0?

**Examples:**
- $p(x) = p_1(x)p_2(x) \cdots p_k(x) - q_1(x)q_2(x) \cdots q_\ell(x)$ for some complicated polynomials $p_1, p_2, \cdots, p_k, q_1, \cdots, q_\ell$.
- $p$ is given as a black box via only an evaluation oracle.
Randomized Algorithm for Univariate Case

- Pick a random element \( a \) from a finite subset \( S \subseteq F \) where \( F \) is the underlying field.
- Evaluate \( p(a) \). If \( p(a) = 0 \) output \( p \) is identically 0. Otherwise say no.
Randomized Algorithm for Univariate Case

- Pick a random element $a$ from a finite subset $S \subseteq F$ where $F$ is the underlying field.
- Evaluate $p(a)$. If $p(a) = 0$ output $p$ is identically 0. Otherwise say no.

Lemma

Suppose $p$ is not 0. Then the algorithm says YES with probability at most $d/|S|$ where $d$ is the degree of $p$. 

$p(a) = 0$ only if $a$ is a root of $p$, or if $p = 0$ identically. $p$ has at most $d$ roots over $F$ via fundamental theorem of algebra.
Randomized Algorithm for Univariate Case

- Pick a random element \( a \) from a finite subset \( S \subseteq F \) where \( F \) is the underlying field.
- Evaluate \( p(a) \). If \( p(a) = 0 \) output \( p \) is identically \( 0 \). Otherwise say no.

**Lemma**

Suppose \( p \) is not \( 0 \). Then the algorithm says YES with probability at most \( d/|S| \) where \( d \) is the degree of \( p \).

- \( p(a) = 0 \) only if \( a \) is a root of \( p \), or if \( p = 0 \) identically.
- \( p \) has at most \( d \) roots over \( F \) via fundamental theorem of algebra.
Lemma

Suppose \( p \) is not 0. Then the algorithm says YES with probability at most \( \frac{d}{|S|} \) where \( d \) is the degree of \( p \).

Why restrict to \( S \)? Isn’t probability 0 when \( F = \mathbb{R} \) if we pick a random real/integer?
Lemma

Suppose $p$ is not 0. Then the algorithm says YES with probability at most $d/|S|$ where $d$ is the degree of $p$.

Why restrict to $S$? Isn't probability 0 when $F = \mathbb{R}$ if we pick a random real/integer?
Restricting $S$ is for computational purposes since evaluation $p(a)$ depends on both $p$ and bit representation of $a$ so picking an arbitrary integer/real would require large precision.
Lemma

Suppose $p$ is not 0. Then the algorithm says YES with probability at most $d/|S|$ where $d$ is the degree of $p$.

Why restrict to $S$? Isn’t probability 0 when $F = \mathbb{R}$ if we pick a random real/integer?
Restricting $S$ is for computational purposes since evaluation $p(a)$ depends on both $p$ and bit representation of $a$ so picking an arbitrary integer/real would require large precision.

Derandomization: Evaluate $p$ on any distinct $d + 1$ values and if $p(a) = 0$ for all of them output $p = 0$. 
**Definition**

A multivariate polynomial over a field $\mathbb{F}$ with $n$ variables $x_1, x_2, \ldots, x_n$ is a finite sum of terms of the form $ax_1^{i_1}x_2^{i_2} \ldots x_n^{i_n}$ where $a \in \mathbb{F}$ and $i_1, i_2, \ldots, i_n \in \mathbb{Z}^+$. The degree of the term $x_1^{i_1}x_2^{i_2} \ldots x_n^{i_n}$ is $\sum_{j=1}^{n} i_j$ and the total degree of $p$ is the maximum degree over all terms in $p$.

**Example**

$p(x_1, x_2, x_3) = 1 + x_1^2 + x_1x_2x_3 + x_2^3x_3^5$  
$\text{deg}(p) = 8$. 
Definition
A polynomial \( p(x_1, \ldots, x_n) \) is identically \( 0 \) if \( p \) is equal \( 0 \) for all \( x_1, \ldots, x_n \in \mathbb{F} \).

Question (PIT)
Given a polynomial \( p \) in some **implicit** fashion, is \( p \) identically \( 0 \)?

Multivariate polynomials are very powerful and can model many problems. PIT is a fundamental algorithmic tool.
Randomized Algorithm for PIT

- Pick independent random elements $a_1, a_2, \ldots, a_n$ from a finite subset $S \subseteq F$ where $F$ is the underlying field.
- If $p(a_1, a_2, \ldots, a_n) = 0$ output $p$ is identically 0. Otherwise say no.

Theorem (Schwartz-Zippel Lemma)

Suppose $p$ is a multivariate polynomial with degree $d$ that is not identically 0. Then the algorithm says YES with probability at most $d / |S|$.

The zero-set of multivariate polynomials is very complex, nevertheless the simple lemma for univariate case generalizes relatively easily and has numerous powerful applications.
Randomized Algorithm for PIT

- Pick independent random elements $a_1, a_2, \ldots, a_n$ from a finite subset $S \subseteq F$ where $F$ is the underlying field.
- If $p(a_1, a_2, \ldots, a_n) = 0$ output $p$ is identically 0. Otherwise say no.

**Theorem (Schwartz-Zippel Lemma)**

*Suppose $p$ is a multivariate polynomial with degree $d$ that is not identically 0. Then the algorithm says YES with probability at most $d/|S|$.***
Randomized Algorithm for PIT

- Pick independent random elements $a_1, a_2, \ldots, a_n$ from a finite subset $S \subseteq F$ where $F$ is the underlying field.
- If $p(a_1, a_2, \ldots, a_n) = 0$ output $p$ is identically 0. Otherwise say no.

Theorem (Schwartz-Zippel Lemma)

Suppose $p$ is a multivariate polynomial with degree $d$ that is not identically 0. Then the algorithm says YES with probability at most $d/|S|$.

The zero-set of multivariate polynomials is very complex, nevertheless the simple lemma for univariate case generalizes relatively easily and has numerous powerful applications.
Proof of Schwartz-Zippel Lemma

Proof based on induction on \( n \).

**Base case:** \( n = 1 \) then follows from univariate case.

**Induction step:** Assume theorem holds if num variables < \( n \) and consider case with \( n \) variables. Let \( p \) be non-zero polynomial. All variables occur in \( p \) with non-zero degree. Hence

\[
p(x_1, \ldots, x_n) = \sum_{j=0}^{d} x_1^j p_j(x_2, \ldots, x_n).
\]
Proof of Schwartz-Zippel Lemma

Proof based on induction on $n$.

Base case: $n = 1$ then follows from univariate case.

Induction step: Assume theorem holds if num variables $< n$ and consider case with $n$ variables. Let $p$ be non-zero polynomial. All variables occur in $p$ with non-zero degree. Hence

$$p(x_1, \ldots, x_n) = \sum_{j=0}^{d} x_1^j p_j(x_2, \ldots, x_n).$$

Since $p \neq 0$ let $t$ be largest $j$ such that $p_j(x_2, \ldots, x_n) \neq 0$. 

Proof of Schwartz-Zippel Lemma

Proof based on induction on $n$.

**Base case:** $n = 1$ then follows from univariate case.

**Induction step:** Assume theorem holds if num variables $< n$ and consider case with $n$ variables. Let $p$ be non-zero polynomial. All variables occur in $p$ with non-zero degree. Hence

$$p(x_1, \ldots, x_n) = \sum_{j=0}^{d} x_1^j p_j(x_2, \ldots, x_n).$$

Since $p \neq 0$ let $t$ be largest $j$ such that $p_j(x_2, \ldots, x_n) \neq 0$. $\deg(p_t) \leq d - t$. 
Think of algorithm as picking $a_2, \ldots, a_n$ first and creating univariate polynomial

$$q(x_1) = \sum_{j=0}^{t} x_1 p_j(a_2, \ldots, a_n)$$

and then picking $a_1$ independently and evaluating $q(a_1)$. 
Proof continued

Think of algorithm as picking $a_2, \ldots, a_n$ first and creating univariate polynomial

$$q(x_1) = \sum_{j=0}^{t} x_1 p_j(a_2, \ldots, a_n)$$

and then picking $a_1$ independently and evaluating $q(a_1)$.

\[
\Pr[\text{Alg is correct}] \geq \Pr[p_t(a_2, \ldots, a_n) \neq 0] \Pr[q(a_1) \neq 0 \mid p_t(a_2, \ldots, a_n) \neq 0] \\
\geq \Pr[p_t(a_2, \ldots, a_n) \neq 0] (1 - t/|S|) \quad \text{(since } q \text{ is a deg } t \text{ polynomial)} \\
\geq (1 - (d - t)/|S|)(1 - t/|S|) \quad \text{(by induction)} \\
\geq (1 - d/|S|).
\]
**Question:** To derandomize algorithm in the naive way one would need to evaluate \( p \) on \((d + 1)^n\) tuples. Exponential when \( n \) is large.

Whether PIT has a deterministic polynomial-time algorithm is a major open problem in complexity theory and algorithms.