**CS789: Introduction to Probabilistic Proof Systems** Spring 2025

# Lecture 02

09/Jan/25

Instructor: Chethan Kamath Scribe: Priyanshu Singh

## 1 Recap

In the last lecture, we discussed the difference between traditional "NP" proof systems and probabilistic proof systems: traditional NP proofs are "perfect", deterministic verification systems, while probabilistic proof systems like IP, ZK, PCP, NIZK, and SNARKs relax this requirement. An example of power of such relaxation is demonstrated by Freivalds' algorithm: an efficient randomised checker for matrix multiplication, which runs in $O(n^2)$ time but with a one-sided error. Note that this is asymptotically faster than the fastest known algorithm for computing matrix multiplication [VXXZ24].

## 2 Plan for Today's Lecture

We will cover the following topics in today's lecture:

- Finite Fields and polynomials over finite fields
- A "Reed-Solomon" view of Freivalds' algorithm
- Schwartz-Zippel Lemma, a fact about multivariate polynomials that we will use throughout the course

## 3 Finite Fields and Polynomials over Finite Fields

Finite fields can be regarded as finite abstractions of real numbers. Thus they inherit most desirable properties of real number but – being finite – they are convenient to represent on computers.

**Definition 1** (Finite Field)**.** A field $\mathbb{F}$ is a set $\mathcal{F}$ with two operations $+$ and $\times$ satisfying the following properties:

1. $(\mathcal{F}, +)$ is an additive abelian group with identity 0
2. $(\mathcal{F} \setminus \{0\}, \times)$ is a multiplicative abelian group with identity 1
3. $\times$ is distributive over $+$, i.e., $a \times (b + c) = a \times b + a \times c$ for all $a, b, c \in \mathcal{F}$

For a finite field, the *order* $|\mathcal{F}|$ is finite.

- Examples of non-finite fields: $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Q}$
- $(\mathbb{Z}_p, +, \times)$, for a prime p, is a finite field

- $(\mathbb{Z}_n, +, \times)$, for a composite integer $n$, is not a finite field

**Fact 1.** *The order of a finite field is always of the form* $\mathrm{q} = \mathrm{p}^k$*, where* p *is a prime and* $k$ *is an integer* $\geq 1$*.*

**Notation 1.** We will denote a finite field of order $q$ by $\mathbb{F}_\mathrm{q}$

**Notation 2** (Polynomials over Finite Fields)**.** For a finite field $\mathbb{F}_\mathrm{q}$:

- Univariate polynomials over $\mathbb{F}_\mathrm{q}$ are denoted as $\mathbb{F}_\mathrm{q}[x]$, representing all polynomials with coefficients from $\mathbb{F}_\mathrm{q}$ and a single variable $x$.
- $m$-variate polynomials over $\mathbb{F}_\mathrm{q}$ are denoted as $\mathbb{F}_\mathrm{q}[x_1, x_2, \ldots, x_m]$, representing all polynomials with coefficients from $\mathbb{F}_\mathrm{q}$ and $m$ variables.

**Fact 2.** *For any two degree-d polynomials* $p_1(x)$ *and* $p_2(x) \in \mathbb{F}_\mathrm{q}[x]$*,* $p_1(a) = p_2(a)$ *for at most* $d$ *values of* $a \in \mathbb{F}_\mathrm{q}$*.*

**Fact 3.** *Any non-zero degree-d polynomial (univariate)* $p(x) \in \mathbb{F}_\mathrm{q}[x]$ *has at most* $d$ *distinct roots.*

**Homework 1** (Fact 3 $\Rightarrow$ Fact 2)**.** Show that if any non-zero degree-$d$ polynomial $p(x) \in \mathbb{F}_\mathrm{q}[x]$ has at most $d$ distinct roots (Fact 3), then for any two degree-$d$ polynomials $p_1(x)$ and $p_2(x) \in \mathbb{F}_\mathrm{q}[x]$, $p_1(a) = p_2(a)$ for at most $d$ values of $a \in \mathbb{F}_\mathrm{q}$ (Fact 2). (*Hint*: Consider the polynomial $p_1(x) - p_2(x)$.)

**Homework 2** (Prove Fact 3)**.** Prove that any non-zero degree-$d$ polynomial $p(x) \in \mathbb{F}_\mathrm{q}[x]$ has at most $d$ distinct roots. (*Hints:*

- Use induction on the degree $d$
- For the base case, consider $d = 0$ (constant polynomials)
- For the inductive step:
  - If $a$ is a root of $p(x)$, then $(x - a)$ divides $p(x)$
  - Consider the polynomial $p(x)/(x - a)$ and its degree.)

# 4 Freivalds' Algorithm: A Reed-Solomon View

A slight variant of Frievalds' algorithm, tailored for prime-order field $\mathbb{F}_\mathrm{p}$, is described in Algorithm 1. Observe that the main difference from the algorithm discussed in Lecture 01 is that $\bar{x}$ is a structured vector constructed from a single $r \leftarrow \mathbb{F}_\mathrm{p}$. The intuition for why this works will become clear in Section 4.2, but let's first analyse its soundness.

**Algorithm 1** Freivalds' Algorithm

**Require:** Matrices $\bar{A}, \bar{B}, \bar{C} \in \mathbb{F}_{\mathrm{p}}^{n \times n}$

1: Verifier chooses $r \in \mathbb{F}_{\mathrm{p}}$ uniformly at random
2: Construct vector $\bar{x} = [1, r, r^2, \ldots, r^{n-1}]^\top$
3: Compute $\bar{y}_1 = \bar{B}x$
4: Compute $\bar{y}_2 = \bar{A}\bar{y}_1$ and $\bar{y}_3 = \bar{C}x$
5: Accept if $\bar{y}_2 = \bar{y}_3$; reject otherwise

## 4.1 Analysis

The completeness of Freivalds' algorithm is straightforward: if $\bar{C}^* = \bar{C} := \bar{A}\bar{B}$, then for any choice of $r \in \mathbb{F}_{\mathrm{p}}$ the following holds

$$\bar{A}(\bar{B}\bar{x}) = (\bar{A}\bar{B})\bar{x} = \bar{C}^*\bar{x}.$$

Therefore, the verifier always accepts. In the claim below, we focus on soundness.

**Claim 1.** *If $\bar{C}^* \neq \bar{A}\bar{B}$, then:*

$$\Pr_{r \leftarrow \mathbb{F}_{\mathrm{p}}}[\mathsf{V} \textit{ rejects}] \geq 1 - \frac{n-1}{\mathrm{p}}$$

*Proof.* Since $C^* \neq AB$, they must differ in at least one row. Let $i$ be such a row.

$$\bar{C}^*\bar{x} = \begin{bmatrix} \vdots & & & \\ c^*_{i1} & c^*_{i2} & \cdots & c^*_{in} \\ \vdots & & & \end{bmatrix} \begin{bmatrix} 1 \\ r \\ r^2 \\ \vdots \\ r^{n-1} \end{bmatrix}$$

$$\bar{C}\bar{x} = \begin{bmatrix} \vdots & & & \\ c_{i1} & c_{i2} & \cdots & c_{in} \\ \vdots & & & \end{bmatrix} \begin{bmatrix} 1 \\ r \\ r^2 \\ \vdots \\ r^{n-1} \end{bmatrix}$$

Looking at the $i$-th entries of $\bar{C}^*\bar{x}$ and $\bar{C}\bar{x}$, we get two polynomials:

$$p_1(x) = (\bar{C}^*_{i*})x = \sum_{k=1}^{n} c^*_{ik} x^{k-1}$$

$$p_2(x) = (\bar{C}_{i*})x = \sum_{k=1}^{n} c_{ik} x^{k-1}$$

Since $\bar{C}^* \neq \bar{C}$, we know that $p_1(x) \neq p_2(x)$. By Fact 2, $p_1(a) = p_2(a)$ for at most $n-1$ values of $a \in \mathbb{F}_{\mathrm{p}}$. Therefore:

$$\Pr_{r \leftarrow \mathbb{F}_{\mathrm{p}}}[\mathsf{V} \text{ accepts}] \leq \frac{n-1}{\mathrm{p}}$$

Hence:

$$\Pr_{r \leftarrow \mathbb{F}_\mathrm{p}}[\mathsf{V} \text{ rejects}] \geq 1 - \frac{n-1}{\mathrm{p}}$$

□

## 4.2 Reed-Solomon Encoding

The above test can be seen through the lens of Reed-Solomon encoding, defined below.

**Definition 2** (Reed-Solomon Encoding)**.** Let p be a prime. For a message vector $\bar{a} = (a_1, a_2, \ldots, a_n) \in \mathbb{F}_\mathrm{p}^n$, the Reed-Solomon encoding is defined as:

$$(p(0), p(1), \ldots, p(\mathrm{p}-1))$$

where $p(x)$ is the polynomial:

$$p(x) = \sum_{i=1}^{n} a_i x^{n-i}$$

This encoding transforms an $n$-dimensional vector into a $q$-dimensional vector by evaluating the associated polynomial at all points in the field $\mathbb{F}_\mathrm{p}$: see Figure 1 (which is adapted from [Tha22, Chapter 2]). Thus the encoding can be considered to be a very long Reed-Solomon codeword of $\bar{a}$.

Now, let's try to understand why Algorithm 1 works. Since $p(x)$ has degree at most $n-1$, the Reed-Solomon encodings $p(x)$ and $p'(x)$ of any two message vectors $\bar{a} \neq \bar{a}' \in \mathbb{F}_\mathrm{p}^n$ (even if they only differ at only *one* index), can agree at *at most* $n-1$ out of the p entries. This is due to Fact 2. Thus the verifier can catch a cheating prover by probing the encoding at a random index $r \leftarrow \mathbb{F}_\mathrm{p}$, who will be caught with overwhelming probability if $\mathrm{p} \gg n$.

# 5 The Schwartz-Zippel Lemma

Finally, we state and prove Schwartz-Zippel Lemma, a multivariate version of Fact 3, which will be useful in later lectures. It states that the number of roots of a non-zero $m$-variate polynomial of total degree $d$ over $\mathbb{F}_\mathrm{q}$ is at most $d \cdot \mathrm{q}^{m-1}$. Below, in Theorem 1, we state it in a slightly different way.

**Theorem 1** (Schwartz-Zippel Lemma)**.** *For any non-zero $m$-variate polynomial $p(x_1, \ldots, x_m) \in \mathbb{F}_\mathrm{q}[x_1, \ldots, x_m]$ of total degree at most $d$:*

$$\Pr_{(a_1,\ldots,a_m) \leftarrow \mathbb{F}_\mathrm{q}^m}[p(a_1, \ldots, a_m) = 0] \leq \frac{d}{\mathrm{q}}$$

*Proof.* We prove by induction on the number of variables $m$.

- *Base Case ($m = 1$)*: This reduces to the univariate case, which follows from Fact 3
- *Induction hypothesis*: For any non-zero $(m-1)$-variate polynomial $p'(x_2, \ldots, x_m)$ of total degree at most $d'$:

$$\Pr_{(a_2,\ldots,a_m) \leftarrow \mathbb{F}_\mathrm{q}^{m-1}}[p'(a_2, \ldots, a_m) = 0] \leq \frac{d'}{\mathrm{q}}$$

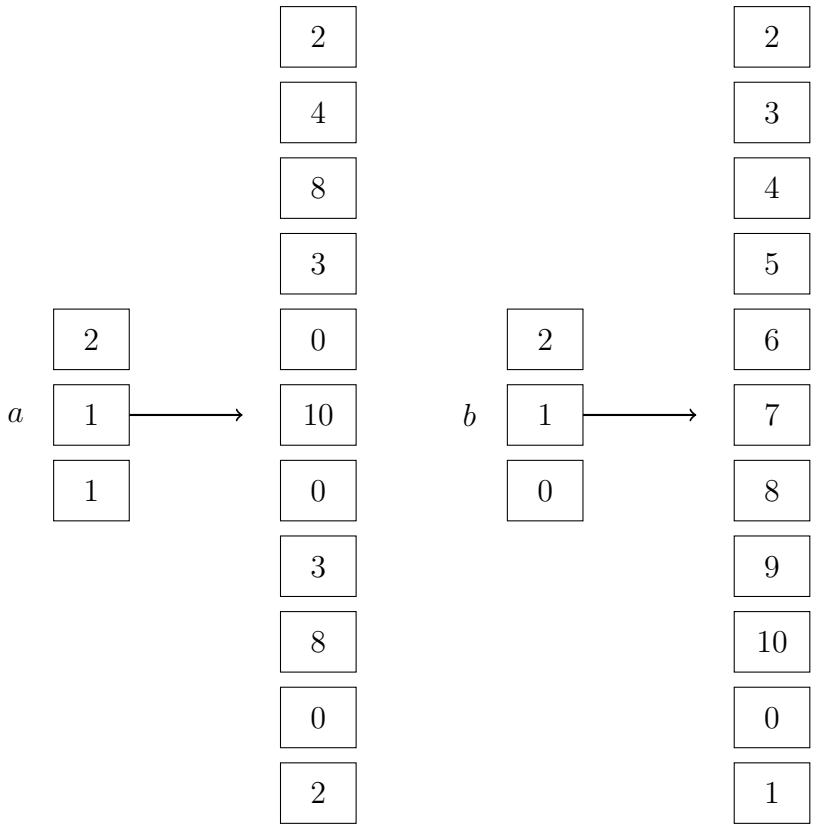


Figure 1: Reed-Solomon encoding example showing the encoding of vectors $a = (2, 1, 1)$ and $b = (2, 1, 0)$ over $\mathbb{F}_{11}$.

- *Induction*: Write $p(x_1, \ldots, x_m)$ as:

$$p(x_1, \ldots, x_m) = \sum_{i=0}^{d} x_1^i \cdot p_i'(x_2, \ldots, x_m)$$

Since $P \neq 0$, there exists some $p_i' \neq 0$. Let $i^*$ be the largest such $i \in [0, d]$.

  - Degree of $p_{i^*}'(x_2, \ldots, x_m) \leq d - i^*$
  - By inductive hypothesis:

$$\Pr_{(a_2,\ldots,a_m)}[p_{i^*}'(a_2, \ldots, a_m) = 0] \leq \frac{d - i^*}{\mathrm{q}}$$

  - When $p_{i^*}'(a_2, \ldots, a_m) \neq 0$, $p(x_1, a_2, \ldots, a_m)$ is a non-zero univariate polynomial of degree $i^*$
  - By Fact 3:

$$\Pr_{a_1}[p(a_1, \ldots, a_m) = 0 \mid p_{i^*}'(a_2, \ldots, a_m) \neq 0] \leq \frac{i^*}{\mathrm{q}}$$

Let $A$ denote the event $p(a_1, \ldots, a_m) = 0$ and $B$ denote $p_{i^*}'(a_2, \ldots, a_m) = 0$. Then:

$$\begin{aligned}
\Pr[A] &= \Pr[A, B] + \Pr[A, \bar{B}] \\
&= \Pr[A|B] \Pr[B] + \Pr[A|\bar{B}] \Pr[\bar{B}] \\
&\leq \Pr[B] + \Pr[A|\bar{B}] \\
&\leq \frac{d - i^*}{\mathrm{q}} + \frac{i^*}{\mathrm{q}} = \frac{d}{\mathrm{q}}
\end{aligned}$$

□

**Corollary 1.** *The number of roots of a non-zero $m$-variate polynomial of total degree $d$ over $\mathbb{F}_\mathrm{q}$ is at most $d \cdot \mathrm{q}^{m-1}$.*

# 6 Solutions to Homework Problems

## 6.1 Preliminaries: Factor Theorem for Fields

**Theorem 2** (Factor Theorem for Fields)**.** *Let $p(x) \in \mathbb{F}_\mathrm{q}[x]$. An element $a \in \mathbb{F}$ is a root of $p(x)$ if and only if $(x - a)$ divides $p(x)$.*

*Proof.* Let $p(x) \in \mathbb{F}_\mathrm{q}[x]$ and $a \in \mathbb{F}$.

- ($\Rightarrow$) If $(x - a) \mid p(x)$, then $\exists q(x) \in \mathbb{F}_\mathrm{q}[x]$ such that:

$$p(x) = (x - a) \cdot q(x) \Rightarrow p(a) = (a - a) \cdot q(a) = 0$$

  Therefore, $a$ is a root of $p(x)$.

- ($\Leftarrow$) If $a$ is a root of $p(x)$, then:

- $p(a) = 0$
- $(x - a)$ is a monic polynomial
- By polynomial long division, $\exists q(x), r(x)$ such that:

$$p(x) = (x - a) \cdot q(x) + r(x)$$

  where degree of $r(x) <$ degree of $(x - a)$
- Therefore, $r(x)$ is constant
- $0 = p(a) = 0 + r(a) = r(a) = r(x)$
- Thus, $p(x) = q(x) \cdot (x - a)$ and $(x - a) \mid p(x)$

□

## 6.2 Solution to Homework 1

*Proof.* Let $p_1(x) \neq p_2(x) \in \mathbb{F}_q[x]$. Define $q(x) = p_1(x) - p_2(x)$. Since the degree of both $p_1$ and $p_2(x)$ is $d$, the degree of $q(x)$ is also $d$. By Fact 3, $q(x)$ has at most $d$ roots. For any $a \in \mathbb{F}_q$:

$$\begin{aligned} a \text{ is a root of } q(x) &\Leftrightarrow q(a) = p_1(a) - p_2(a) = 0 \\ &\Leftrightarrow p_1(a) = p_2(a) \end{aligned}$$

Therefore, $p_1(a) = p_2(a)$ for at most $d$ values of $a \in \mathbb{F}_q$. □

## 6.3 Solution to Homework 2

*Proof.* Let $\mathbb{F}_q$ be a field and $p(x) \in \mathbb{F}_q[x]$ be a non-zero polynomial of degree $n \geq 0$. We prove by induction over $n$.

- *Base Case:* For $n = 0$, $p(x)$ is a nonzero constant and thus has zero roots.
- *Inductive Hypothesis:* For any non-zero $(m-1)$-variate polynomial $p'(x_2, \ldots, x_m)$ of total degree at most $d'$:

$$\Pr_{(a_2,\ldots,a_m) \leftarrow \mathbb{F}_q^{m-1}} [p'(a_2, \ldots, a_m) = 0] \leq \frac{d'}{q}$$

- *Induction:* Let $p(x) \in \mathbb{F}_q[x]$ be of degree $(n + 1)$.
  - If $p(x)$ has 0 roots, the result holds trivially.
  - If $p(x)$ has a root $a \in \mathbb{F}_q$, then by the Factor Theorem, $\exists q(x) \in \mathbb{F}_q[x]$ such that:

$$p(x) = q(x) \cdot (x - a)$$

  - $q(x)$ is of degree $\leq n$, so by our hypothesis, $q(x)$ has $\leq n$ roots
  - Any root of $q(x)$ is a root of $p(x)$, and any $b \in \mathbb{F}_q \neq a$ that is a root of $p(x)$ must be a root of $q(x)$
  - Therefore, the number of roots in $p(x)$ is $\leq (n + 1)$

  By the principle of mathematical induction, the statement holds for all $n \geq 0$.

□

# References

[Tha22] Justin Thaler. *Proofs, Arguments, and Zero-Knowledge*. https://people.cs.georgetown.edu/jthaler/ProofsArgsAndZK.html, 2022. 4

[VXXZ24] Virginia Vassilevska Williams, Yinzhan Xu, Zixuan Xu, and Renfei Zhou. New bounds for matrix multiplication: from alpha to omega. In *SODA*, pages 3792–3835. SIAM, 2024. 1