STIR (blueprint)

LeastAuthority

April 11, 2025

Chapter 1

The Reed-Solomon code

Definition 1.1 (Error-Correcting Code)**.** *An* error-correcting code *of length n over an alphabet Σ is a subset $\mathcal{C} \subseteq \Sigma^n$. The code $\mathcal{C}$ is called a* linear code *if $\Sigma = \mathbb{F}$ is a finite field and $\mathcal{C}$ is a subspace of $\mathbb{F}^n$.*

Definition 1.2 (Reed-Solomon Code)**.** *The* Reed-Solomon code *over finite field $\mathbb{F}$, evaluation domain $\mathcal{L} \subseteq \mathbb{F}$ and degree $d \in \mathbb{N}$ is the set of evaluations (over $\mathcal{L}$) of univariate polynomials (over $\mathbb{F}$) of degree less than d:*

$$\mathrm{RS}[\mathbb{F}, \mathcal{L}, d] := \left\{ f : \mathcal{L} \to \mathbb{F} \;\middle|\; \exists\, \hat{f} \in \mathbb{F}^{<d}[X] \text{ such that } \forall x \in \mathcal{L},\ f(x) = \hat{f}(x) \right\}.$$

The rate of $\mathrm{RS}[\mathbb{F}, \mathcal{L}, d]$ is $\rho := \frac{d}{|\mathcal{L}|}$.

Given a code $\mathcal{C} := \mathrm{RS}[\mathbb{F}, \mathcal{L}, d]$ and a function $f : \mathcal{L} \to \mathbb{F}$, we sometimes use $\hat{f} \in \mathbb{F}^{<d}[X]$ to denote a nearest polynomial to f on $\mathcal{L}$ (breaking ties arbitrarily).

Remark 1.3. *Note that the evaluation domain $\mathcal{L} \subseteq \mathbb{F}$ is a non-empty set.*

Definition 1.4. *For a Reed-Solomon code $\mathcal{C} := \mathrm{RS}[\mathbb{F}, \mathcal{L}, d]$, parameter $\delta \in [0, 1]$, and a function $f : \mathcal{L} \to \mathbb{F}$, let $\mathsf{List}(f, d, \delta)$ denote the list of codewords in $\mathcal{C}$ whose relative Hamming distance from f is at most δ. We say that $\mathcal{C}$ is (δ, l)-*list decodable *if*

$$\left|\mathsf{List}(f, d, \delta)\right| \leq l \quad \text{for every function } f.$$

The Johnson bound provides an upper bound on the list size of this Reed-Solomon code:

Theorem 1.5 (Johnson bound)**.** *The Reed-Solomon code $\mathrm{RS}[\mathbb{F}, \mathcal{L}, d]$ is $(1 - \sqrt{\rho} - \eta, \frac{1}{2\eta\rho})$-list-decodable for every $\eta \in (0, 1 - \sqrt{\rho})$, where $\rho := \frac{d}{|\mathcal{L}|}$ is the rate of the code.*

Chapter 2

Tools for Reed-Solomon codes

2.1 Random linear combination as a proximity generator

Theorem 2.1. *Let $\mathcal{C} := \mathrm{RS}[\mathbb{F}, \mathcal{L}, d]$ be a Reed-Solomon code with rate $\rho := \frac{d}{|\mathcal{L}|}$ and let $B'(\rho) := \sqrt{\rho}$. For every $\delta \in (0, 1 - B'(\rho))$ and functions $f_1, \dots, f_m : \mathcal{L} \to \mathbb{F}$, if*

$$\Pr_{r \leftarrow \mathbb{F}}\Big[\Delta\Big(\sum_{j=1}^{m} r^{j-1} \cdot f_j, \mathrm{RS}[\mathbb{F}, \mathcal{L}, d]\Big) \leq \delta\Big] > \mathsf{err}'(d, \rho, \delta, m),$$

then there exists a subset $S \subseteq \mathcal{L}$ with

$$|S| \geq (1 - \delta) \cdot |L|,$$

and for every $i \in [m]$, there exists $u \in \mathrm{RS}[\mathbb{F}, \mathcal{L}, d]$ such that

$$f_i(S) = u(S).$$

Above, $\mathsf{err}'(d, \rho, \delta, m)$ is defined as follows:

- *if $\delta \in \left(0, \frac{1-\rho}{2}\right]$ then*
$$\mathsf{err}'(d, \rho, \delta, m) = \frac{(m-1) \cdot d}{\rho \cdot |\mathbb{F}|}$$
- *if $\delta \in \left(\frac{1-\rho}{2}, 1 - \sqrt{\rho}\right)$ then*
$$\mathsf{err}'(d, \rho, \delta, m) = \frac{(m-1) \cdot d^2}{|\mathbb{F}| \cdot \left(2 \cdot \min 1 - \sqrt{\rho} - \delta, \frac{\sqrt{\rho}}{20}\right)^7}$$

2.2 Univariate Function Quotienting

In the following, we start by defining the *quotient* of a univariate function.

Definition 2.2. *Let $f : \mathcal{L} \to \mathbb{F}$ be a function, $S \subseteq \mathbb{F}$ be a set, and $\mathsf{Ans}, \mathsf{Fill} : S \to \mathbb{F}$ be functions. Let $\hat{\mathsf{Ans}} \in \mathbb{F}^{<|S|}[X]$ be the (unique) polynomial with $\hat{\mathsf{Ans}}(x) = \mathsf{Ans}(x)$ for every $x \in S$, and let*

$\hat{V}_S \in \mathbb{F}^{<|S|+1}[X]$ *be the unique non-zero polynomial with* $\hat{V}_S(x) = 0$ *for every* $x \in S$. *The* quotient function $\mathsf{Quotient}(f, S, \mathsf{Ans}, \mathsf{Fill}) : \mathcal{L} \to \mathbb{F}$ *is defined as follows:*

$$\forall x \in \mathcal{L}, \quad \mathsf{Quotient}(f, S, \mathsf{Ans}, \mathsf{Fill})(x) := \begin{cases} \mathsf{Fill}(x) & \text{if } x \in S \\ \dfrac{f(x) - \hat{\mathsf{Ans}}(x)}{V_S(x)} & \text{otherwise} \end{cases}$$

Next we define the polynomial quotient operator, which quotients a polynomial relative to its output on evaluation points. The polynomial quotient is a polynomial of lower degree.

Definition 2.3. *Let* $\hat{f} \in \mathbb{F}^{<d}[X]$ *be a polynomial and* $S \subseteq \mathbb{F}$ *be a set, let* $\hat{V}_S \in \mathbb{F}^{<|S|+1}[X]$ *be the unique non-zero polynomial with* $\hat{V}_S(x) = 0$ *for every* $x \in S$. *The* polynomial quotient $\mathsf{PolyQuotient}(\hat{f}, S) \in \mathbb{F}^{<d-|S|}[X]$ *is defined as follows:*

$$\mathsf{PolyQuotient}(\hat{f}, S)(X) := \frac{\hat{f}(X) - \hat{\mathsf{Ans}}(X)}{\hat{V}_S(X)}$$

The following lemma, implicit in prior works, shows that if the function is “quotiented by the wrong value”, then its quotient is far from low-degree.

Lemma 2.4. *Let* $f : \mathcal{L} \to \mathbb{F}$ *be a function,* $d \in \mathbb{N}$ *be the degree parameter,* $\delta \in (0, 1)$ *be a distance parameter,* $S \subseteq \mathbb{F}$ *be a set with* $|S| < d$, *and* $\mathsf{Ans}, \mathsf{Fill} : \mathcal{L} \to \mathbb{F}$ *are functions. Suppose that for every* $u \in \mathsf{List}(f, d, \delta)$ *there exists* $x \in S$ *with* $\hat{u}(x) \neq \mathsf{Ans}(x)$. *Then*

$$\Delta(\mathsf{Quotient}(f, S, \mathsf{Ans}, \mathsf{Fill}), \mathrm{RS}[\mathbb{F}, \mathcal{L}, d - |S|]) + \frac{|T|}{|\mathcal{L}|} > \delta,$$

where $T := \{x \in \mathcal{L} \cap S : \hat{\mathsf{Ans}}(x) \neq f(x)\}$.

2.3 Out of domain sampling

Lemma 2.5. *Let* $f : \mathcal{L} \to \mathbb{F}$ *be a function,* $d \in \mathbb{N}$ *be a degree parameter,* $s \in \mathbb{N}$ *be a repetition parameter, and* $\delta \in (0, 1)$ *be a distance parameter. If* $\mathrm{RS}[\mathbb{F}, \mathcal{L}, d]$ *be* (d, l)*-list decodable then*

$$\begin{aligned} \Pr_{r_1, \ldots, r_l \leftarrow \mathbb{F} \ \mathcal{L}}[\exists \text{ distinct } u, u' \in \mathsf{List}(f, d, \delta) : \forall i \in [s], \hat{u}(i) = \hat{u}'(i)] &\leq \binom{l}{2} \cdot \Big(\frac{d-1}{|\mathbb{F}| - |\mathcal{L}|}\Big)^s \\ &\leq \Big(\frac{l^2}{2}\Big) \cdot \Big(\frac{d}{|\mathbb{F}| - |\mathcal{L}|}\Big)^s \end{aligned}$$

2.4 Folding univariate functions

STIR relies on k-wise folding of functions and polynomials - this is similar to prior works, although presented in a slightly different form. As shown below, folding a function preserves proximity from the Reed-Solomon code with high probability.

The folding operator is based on the following fact, decomposing univariate polynomials into bivariate ones.

Fact 2.6. *Given a polynomial* $\hat{q} \in \mathbb{F}[X]$*:*

- *For every univariate polynomial* $\hat{f} \in \mathbb{F}[X]$*, there exists a unique bivariate polynomial* $\hat{Q} \in \mathbb{F}[X, Y]$ *with* $\mathsf{deg}_X(\hat{Q}) := \lfloor \mathsf{deg}(\hat{f})/\mathsf{deg}(\hat{q}) \rfloor$ *and* $\mathsf{deg}_Y(\hat{Q}) < \mathsf{deg}(\hat{q})$ *such that* $\hat{f}(Z) = \hat{Q}(\hat{q}(Z), Z)$*. Moreover* $\hat{Q}$ *can be computed efficiently given* $\hat{f}$ *and* $\hat{q}$*. Observe that if* $\mathsf{deg}(\hat{f}) < t \cdot \mathsf{deg}(\hat{q})$ *then* $\mathsf{deg}(\hat{Q}) < t$*.*
- *For every* $\hat{Q}[X, Y]$ *with* $\mathsf{deg}_X(\hat{Q}) < t$ *and* $\mathsf{deg}_Y(\hat{Q}) < \mathsf{deg}(\hat{q})$*, the polynomial* $\hat{f}(Z) = \hat{Q}(\hat{q}(Z), Z)$ *has degree* $\mathsf{deg}(\hat{f}) < t \cdot \mathsf{deg}(\hat{q})$*.*

Below, we define folding of a polynomial followed by folding of a function.

Definition 2.7. *Given a polynomial* $\hat{f} \in \mathbb{F}^{<d}[X]$*, a folding parameter* $k \in \mathbb{N}$ *and* $r \in \mathbb{F}$*, we define a polynomial* $\mathsf{PolyFold}(\hat{f}, k, r) \in \mathbb{F}^{d/k}[X]$ *as follows. Let* $\hat{Q}[X, Y]$ *be the bivariate polynomial derived from* $\hat{f}$ *using Fact 2.6 with* $\hat{q}(X) := X^k$*. Then* $\mathsf{PolyFold}(\hat{f}, k, r)(X) := \hat{Q}(X, r)$*.*

Definition 2.8. *Let* $f : \mathcal{L} \to \mathbb{F}$ *be a function,* $k \in \mathbb{N}$ *a folding parameter and* $\alpha \in \mathbb{F}$*. For every* $x \in \mathcal{L}^k$*, let* $\hat{p}_x \in \mathbb{F}^{<k}[X]$ *be the polynomial where* $\hat{p}_x(y) = f(y)$ *for every* $y \in \mathcal{L}$ *such that* $y^k = x$*. We define* $\mathsf{Fold}(f, k, \alpha) : \mathcal{L} \to \mathbb{F}$ *as follows.*

$$\mathsf{Fold}(f, k, \alpha) := \hat{p}_x(\alpha).$$

In order to compute $\mathsf{Fold}(f, k, \alpha)(x)$ *it suffices to interpolate the* k *values* $\{f(y) : y \in \mathcal{L} \text{ s.t. } y^k = x\}$ *into the polynomial* $\hat{p}_x$ *and evaluate this polynomial at* α*.*

The following lemma shows that the distance of a function is preserved under folding. If a functions f has distance δ to a Reed-Solomon code then, with high probability over the choice of folding randomness, its folding also has a distance of δ to the "k-wise folded" Reed-Solomon code.

Lemma 2.9. *For every function* $f : \mathcal{L} \to \mathbb{F}$*, degree parameter* $d \in \mathbb{N}$*, folding parameter* $k \in \mathbb{N}$*, distance parameter* $\delta \in (0, \min\{\Delta(\mathsf{Fold}[f, k, r^{\mathsf{fold}}], \mathrm{RS}[\mathbb{F}, \mathcal{L}^k, d/k]), 1 - \mathsf{B}^*(\rho)\})$*, letting* $\rho := \frac{d}{|\mathcal{L}|}$*,*

$$\Pr_{r^{\mathsf{fold}} \leftarrow \mathbb{F}}[\Delta(\mathsf{Fold}[f, k, r^{\mathsf{fold}}], \mathrm{RS}[\mathbb{F}, \mathcal{L}^k, d/k]) < \delta] > \mathsf{err}^*(d/k, \rho, \delta, k).$$

Above, B^* *and* err^* *are the proximity bound and error (respectively) described in Section 2.1.*

2.5 Combine functions of varying degrees

We show a new method for combining functions of varying degrees with minimal proximity require- ments using geometric sums. We begin by recalling a fact about geometric sums.

Fact 2.10. *Let* $\mathbb{F}$ *be a field,* $r \in \mathbb{F}$ *be a field element,* $a \in \mathbb{N}$ *be a natural number. Then*

$$\sum_{i=0}^{a} r^i := \begin{cases} \left(\frac{1-r^{a+1}}{1-r}\right) & r \neq 1 \\ a+1 & r = 1 \end{cases}$$

Definition 2.11. *Given target degree* $d \in \mathbb{N}$*, shifting parameter* r*, functions* $f_1, \dots, f_m : \mathcal{L} \to \mathbb{F}$*, and degrees* $0 \le d_1, \dots, d_m \le d^*$*, we define* $\mathsf{Combine}(d^*, r, (f_1, d_1), \dots, (f_m, d_m)) : \mathcal{L} \to \mathbb{F}$ *as follows:*

$$\begin{aligned}\mathsf{Combine}(d^*, r, (f_1, d_1), \dots, (f_m, d_m))(x) &:= \sum_{i=1}^{m} r_i \cdot f_i(x) \cdot \Big(\sum_{l=0}^{d^*-d_i} (r \cdot x)^l \Big) \\ &= \begin{cases} \sum_{i=1}^{m} r_i \cdot f_i(x) \cdot \left(\frac{1-(xr)^{d^*-d_i+1}}{1-xr}\right) & x \cdot r \neq 1 \\ \sum_{i=1}^{m} r_i \cdot f_i(x) \cdot (d^* - d_i + 1) & x \cdot r = 1 \end{cases}\end{aligned}$$

Above, $r_1 := 1$*,* $r_i := r^{i-1+\sum_{j<i}(d^*-d_i)}$ *for* $i > 1$*.*

Definition 2.12. *Given target degree* $d \in \mathbb{N}$*, shifting parameter* r*, function* $f : \mathcal{L} \to \mathbb{F}$*, and degree* $0 \le d \le d^*$*, we define* $\mathsf{DegCor}(d^*, r, f, d)$ *as follows.*

$$\mathsf{DegCor}(d^*, r, f, d)(x) := f(x) \cdot \left(\sum_{i=0}^{m} (r \cdot x)^l \right) = \begin{cases} f(x) \cdot \frac{1-(xr)^{d^*-d_i+1}}{1-xr} & x \cdot r \neq 1 \\ f(x) \cdot (d^* - d_i + 1) & x \cdot r = 1 \end{cases}$$

(Observe that $\mathsf{DegCor}(d^*, r, f, d) = \mathsf{Combine}(d^*, r, (f, d))$*.)*

Below it is shown that combining multiple polynomials of varying degrees can be done as long as the proximity error is bounded by $(\min\{1 - \mathsf{B}^*(\rho), 1 - \rho - 1/|\mathcal{L}|\})$.

Lemma 2.13. *Let* d^* *be a target degree,* $f_1, \ldots, f_m : \mathcal{L} \to \mathbb{F}$ *be functions,* $0 \le d_1, \ldots, d_m \le d^*$ *be degrees,* $\delta \in \min\{1 - \mathsf{B}^*(\rho), 1 - \rho - 1/|\mathcal{L}|\}$ *be a distance parameter, where* $\rho = d^*/|\mathcal{L}|$*. If*

$$\Pr_{r \leftarrow \mathbb{F}}[\Delta(\mathsf{Combine}(d^*, r, (f_1, d_1), \ldots, (f_m, d_m)), \mathrm{RS}[\mathbb{F}, \mathcal{L}, d^*])] > \mathsf{err}^*(d^*, \rho, \delta, m \cdot (d^* + 1) - \sum_{i=1}^{m} d_i),$$

then there exists $S \subseteq \mathcal{L}$ *with* $|S| \ge (1 - \delta) \cdot |\mathcal{L}|$*, and*

$$\forall i \in [m], \exists u \in \mathrm{RS}[\mathbb{F}, \mathcal{L}, d_i], f_i(S) = u(S).$$

Note that this implies $\Delta(f_i, \mathrm{RS}[\mathbb{F}, \mathcal{L}, d_i]) < \delta$ *for every* i*. Above,* B^* *and* err^* *are the proximity bound and error (respectively) described in Section 2.1.*