Note on Soundness Proof of Basefold under List Decoding

- Jade Xie jade@secbit.io
- Yu Guo yu.guo@secbit.io

In the previous article "Overview of Basefold's Soundness Proof under List Decoding", we outlined the approach to the soundness proof in the [H24] paper. This article will delve deeper into the proof details following this approach, focusing mainly on the proof of [H24, Lemma 1], which demonstrates the soundness error of the Basefold protocol in the commit phase.

Lemma 1 [H24, Lemma 1] (Soundness commit phase). Take a proximity parameter $\theta = 1 - \left(1 + \frac{1}{2 \cdot m}\right) \cdot \sqrt{\rho}$, with $m \geq 3$. Suppose that a (possibly computationally unbounded) algorithm P^* succeeds the commitment phase with $r \geq 0$ rounds with probability larger than

$$\varepsilon_C = \varepsilon_0 + \varepsilon_1 + \ldots + \varepsilon_r, \tag{1}$$

where $\varepsilon_0 = \varepsilon(\mathcal{C}_i, M, \theta)$ is the soundness error from Theorem 3, and

$$\varepsilon_i := \varepsilon(\mathcal{C}_i, 1, B_i, \theta) + \frac{1}{|F|}, \tag{2}$$

with $\varepsilon(\mathcal{C}_i, 1, B_i, \theta)$ being the soundness error from Theorem 4, where $B_i = \frac{|D|}{|D_i|} = 2^i$. Then $(g_0, \ldots, g_M)$ belongs to $\mathcal{R}$.

[H24, Theorem 3] mentioned in the lemma is the correlated agreement theorem for subcodes under list decoding, while [H24, Theorem 4] is the weighted version of [H24, Theorem 3].

The relation $\mathcal{R}$ implies that P^* has not cheated, indicating that the committed polynomials $(g_0, \ldots, g_M)$ are both within distance θ from the corresponding encoding space and consistent with the committed values $v_0, \ldots, v_M$ at the query point $\vec{\omega} = (\omega_1, \ldots, \omega_n)$, i.e.,

$$\mathcal{R} = \left\{ \begin{array}{c} \exists p_0, \ldots, p_M \in \mathscr{F}[X]^{<2^n} \text{ s.t.} \\ (g_0, \ldots, g_M) : d\left((g_0, \ldots, g_M), (p_0, \ldots, p_M)\right) < \theta \\ \wedge \bigwedge_{k=0}^{M} P_k\left(\omega_1, \ldots, \omega_n\right) = v_k \end{array} \right\}. \tag{3}$$

Lemma 1 states that if P^*'s success probability in the commit phase exceeds ε_C, we can trust that P^* has not cheated, and the claimed relation $\mathcal{R}$ holds.

Here, we need to mathematically define what it means for P^* to succeed in the $0 \leq r \leq n$ round of the commit phase. This is the concept of α-good given in the [H24] paper. From the protocol itself, P^*'s success means that the verifier receives $f_0, \Lambda_0, f_1, \Lambda_1, f_2, \Lambda_2, \ldots, \Lambda_{r-1}, f_r$ from P^*, then performs checks: one is the sumcheck, and the other is randomly selecting x in D_0 to verify that the FRI folding is correct. First, the parameter $\alpha = 1 - \theta \in (0, 1)$, i.e.,

$$\alpha = \left(1 + \frac{1}{2 \cdot m}\right) \cdot \sqrt{\rho} \tag{4}$$

Let $\mathcal{F}_i$ represent the polynomial space corresponding to the Reed-Solomon code $\mathcal{C}_i = \mathrm{RS}_{2^{n-i}}[F, D_i]$, where D_i is the result of applying the mapping π to D i times, i.e., $D_i = \pi^i(D), i = 0, \ldots, n$. Therefore, the polynomial subspace corresponding to $\mathcal{C}_i' \subseteq \mathcal{C}_i$ is defined as

$$\mathcal{F}_i' = \{p(X) \in \mathcal{F}_i : P(\omega_{i+1}, \ldots, \omega_n) = 0\}. \tag{5}$$

1. The sumcheck is correct. This means there exists $p_r(X) \in \mathcal{F}_r$, with corresponding multivariate polynomial P_r satisfying

   $$L((\omega_1, \ldots, \omega_r), (\lambda_1, \ldots, \lambda_r)) \cdot P_r(\omega_1, \ldots, \omega_n) = q_{r-1}(\lambda_r) \tag{6}$$

   Based on the relationship between $q_i(X)$ and $\Lambda_i(X)$, we can deduce that P_r needs to satisfy

$$
\begin{aligned}
L((\omega_1,\ldots,\omega_r),(\lambda_1,\ldots,\lambda_r))\cdot P_r(\omega_{r+1},\ldots,\omega_n) &= q_{r-1}(\lambda_r) \\
&= L((\omega_1,\ldots,\omega_r),(\lambda_1,\ldots,\lambda_r))\cdot \Lambda_{r-1}(\lambda_r)
\end{aligned} \tag{1}
$$

2. The folding is correct. It needs to satisfy

$$
\left|\left\{ x \in D_0 : \begin{array}{c} (f_0,\ldots,f_r) \text{ satisfy all folding checks along } x \\ \wedge f_r(\pi^r(x)) = p_r(\pi^r(x)) \end{array} \right\}\right| \geq \alpha \cdot |D_0| \tag{2}
$$

Here, only when the proportion of x in D_0 satisfying the folding check is greater than α, after mapping through π^r, will the verifier pass in the end.

When conditions 1 and 2 are met, we say that such $(f_0, \Lambda_0, f_1, \Lambda_1, f_2, \Lambda_2, \ldots, \Lambda_{r-1}, f_r)$ is α-good for $(\lambda_0, \ldots, \lambda_r)$.

Proof of Lemma 1

The proof of Lemma 1 uses mathematical induction. First, it proves that the conclusion holds when $r = 0$, using [H24, Theorem 3]. Then, assuming Lemma 1 holds for $0 \leq r < n$, it proves that the conclusion also holds for $r+1$. This process uses the weighted [H24, Theorem 4], following a similar approach to the one introduced in the previous article. For example, in the $r+1$ round, starting with the conditions satisfied by f_{r+1} obtained after folding with the random number λ_{r+1}, which is close to the corresponding encoding space and satisfies the sumcheck constraint, we first deduce that the corresponding f'_{r+1} satisfies some conditions. This allows us to use the correlated agreement theorem for subcodes. Applying the theorem's conclusion, we can then derive the properties satisfied by $f_{r,0}$ and $f_{r,1}$ before folding, and from this, deduce the properties satisfied by f_r. At this point, applying the induction hypothesis, we can obtain that the conditions of the lemma are satisfied in the r-th round, thus proving that the conclusion holds in the r-th round, which in turn proves that the lemma holds in the $(r+1)$-th round.

Proof: First, prove that the lemma holds when $r = 0$. The given condition is that P^*'s success probability in the commit phase is greater than $\varepsilon(\mathcal{C}_0, M, \theta)$, and we want to prove that $(g_1, \ldots, g_M) \in \mathcal{R}$. According to the condition and the definition of α-good, we can deduce that with a probability greater than $\varepsilon(\mathcal{C}_0, M, \theta)$, the f_0 provided by P^* is α-good for λ_0. Then, considering the polynomials $g'_k = g_k - v_k$ before folding, the probability that they are within distance θ from the corresponding subcode $\mathcal{C}'_0 \subseteq \mathcal{C}_0$ (which means the consistent part is greater than α) is

$$
\Pr\left[\lambda_0 : \exists p'_0 \in \mathcal{F}'_0 \text{ s.t. agree}\left(\sum_{k=0}^{M} g'_k \cdot \lambda_0^k, p'_0(X)\right) \geq \alpha\right] > \varepsilon(\mathcal{C}_0, M, \theta) \tag{7}
$$

The purpose of considering polynomials $g'_k = g_k - v_k$ instead of g_k is to allow our analysis to enter the scope of the linear subcode $\mathcal{C}'_0$, so we can use [H24, Theorem 3] to obtain polynomials

$$
p'_0(X), \ldots, p'_M(X) \in \mathcal{F}'_0 \tag{8}
$$

and a set $D'_0 \subseteq D$, satisfying

1. $|D'_0|/|D| \geq \alpha$
2. $p'_k(X)|_{D'_0} = g'_k(X)|_{D'_0}$

Now that we have found polynomials $p'_0(X), \ldots, p'_M(X)$, for polynomials

$$
p'_0(X) + v_0, \ldots, p'_M(X) + v_M \in \mathcal{F}_0 \tag{9}
$$

they satisfy

$$
(p'_k(X) + v_k)|_{D'_0} = (g'_k(X) + v_k)|_{D'_0} = g_k(X)|_{D'_0} \quad 0 \leq k \leq M \tag{10}
$$

The multilinear polynomial $P_k \in F[X_1, \ldots, X_n]$ corresponding to $p'_0(X) + v_0$ also satisfies $P_k(\vec{\omega}) = v_k$, therefore $(g_1, \ldots, g_M) \in \mathcal{R}$.

Now assume the lemma holds for $0 \leq r < n$, and we want to prove that it still holds for $r+1$. According to the conditions of the lemma, in the $(r+1)$-th round, P^*'s success probability in the commit phase exceeds $(\varepsilon_0 + \varepsilon_1 + \ldots + \varepsilon_r) + \varepsilon_{r+1}$. Let $\mathfrak{T}$ be the set composed of $\mathrm{tr}_r = (\lambda_0, f_0, \Lambda_0, \ldots, \lambda_r, f_r, \Lambda_r)$. Therefore, under the condition

$$
\Pr[\mathfrak{T}] > \varepsilon_0 + \ldots + \varepsilon_r \tag{11}
$$

P^*'s success probability is greater than ε_{r+1}, i.e.,

$$\Pr\left[\lambda_{r+1} : \begin{array}{c}\exists f_{r+1} \text{ s.t. } (\lambda_0, f_0, \Lambda_0, \ldots, \lambda_r, f_r, \Lambda_r, f_{r+1}) \\ \text{is } \alpha\text{-good for } (\lambda_0, \ldots, \lambda_{r+1})\end{array}\right] > \varepsilon_{r+1} \tag{12}$$

From the definition of α-good, we can deduce that for λ_{r+1} satisfying α-good, there exists a polynomial $p_{r+1} \in \mathcal{F}_{r+1}$ satisfying the sumcheck constraint, such that

$$\mathrm{agree}_{\nu_r}((1-\lambda_{r+1}) \cdot f_{r,0} + \lambda_{r+1} \cdot f_{r,1}, p_{r+1}) \geq \alpha \tag{3}$$

Here, ν_r is a sub-probability measure with density function defined as, for $y \in D_{r+1}$

$$\delta_r(y) := \frac{|\{x \in \pi^{-(r+1)}(y) : (f_0, \ldots, f_r) \text{ satisfies all folding checks along } x\}|}{|\pi^{-(r+1)}(y)|} \tag{13}$$

Here's an explanation of what equation (3) essentially represents: it's equivalent to equation (2) in the definition of α-good. According to the definition of the agree function, equation (3) is equivalent to

$$\frac{\nu_r(\{y \in D_{r+1} : ((1-\lambda_{r+1}) \cdot f_{r,0} + \lambda_{r+1} \cdot f_{r,1})(y) = p_{r+1}(y)\})}{|D_{r+1}|} \geq \alpha \tag{14}$$

First, let's form a set S_{r+1} consisting of y in D_{r+1} that satisfy the folding relation, then calculate this set using the ν_r function.

$$\begin{aligned}
\nu_r(S_{r+1}) &= \sum_{y \in S_{r+1}} \delta_r(y) \\
&= \sum_{y \in S_{r+1}} \frac{|\{x \in \pi^{-(r+1)}(y) : (f_0, \ldots, f_r) \text{ satisfies all folding checks along } x\}|}{|\pi^{-(r+1)}(y)|} \\
&= \sum_{y \in S_{r+1}} \frac{|\{x \in \pi^{-(r+1)}(y) : (f_0, \ldots, f_r) \text{ satisfies all folding checks along } x\}|}{2^{r+1}} \\
&:= \sum_{y \in S_{r+1}} \frac{|S_{y,0}|}{2^{r+1}} \\
&= \frac{\sum_{y \in S_{r+1}} |S_{y,0}|}{2^{r+1}}
\end{aligned} \tag{15}$$

Therefore

$$\begin{aligned}
\mathrm{agree}_{\nu_r}((1-\lambda_{r+1}) \cdot f_{r,0} + \lambda_{r+1} \cdot f_{r,1}, p_{r+1}) &= \frac{\nu_r(S_{r+1})}{|D_{r+1}|} \\
&= \frac{\sum_{y \in S_{r+1}} |S_{y,0}|}{2^{r+1} \cdot |D_{r+1}|} \\
&= \frac{\sum_{y \in S_{r+1}} |S_{y,0}|}{|D_0|}
\end{aligned} \tag{16}$$

The numerator $\sum_{y \in S_{r+1}} |S_{y,0}|$ in the above equation represents the number of points in D_0 that satisfy the $(r+1)$-th folding correctly, and also pass the folding checks for $(f_0, \ldots, f_r)$. Equation (3) becomes

$$\sum_{y \in S_{r+1}} |S_{y,0}| \geq \alpha \cdot |D_0| \tag{17}$$

This is completely consistent with equation (2) in the definition of α-good. Next, following the soundness proof approach introduced in the previous article, since the multilinear polynomial P_{r+1} corresponding to $p_{r+1}(X)$ satisfies the sumcheck constraint, it satisfies

$$\begin{aligned}
L((\omega_1, \ldots, \omega_{r+1}), (\lambda_1, \ldots, \lambda_{r+1})) \cdot P_{r+1}(\omega_{r+2}, \ldots, \omega_n) &= q_r(\lambda_{r+1}) \\
&= L((\omega_1, \ldots, \omega_{r+1}), (\lambda_1, \ldots, \lambda_{r+1})) \cdot \Lambda_r(\lambda_{r+1})
\end{aligned} \tag{18}$$

This leads to

$$
\begin{aligned}
L((\omega_1,\ldots,\omega_r),(\lambda_1,\ldots,\lambda_r)) \cdot L(\omega_{r+1},\lambda_{r+1}) \cdot & P_{r+1}(\omega_{r+2},\ldots,\omega_n) \\
&= L((\omega_1,\ldots,\omega_r),(\lambda_1,\ldots,\lambda_r)) \cdot L(\omega_{r+1},\lambda_{r+1}) \cdot \Lambda_r(\lambda_{r+1})
\end{aligned} \tag{19}
$$

For the choice of λ_{r+1}, there is a $1/|F|$ probability that $L(\omega_{r+1},\lambda_{r+1})=0$, making the above equation hold. Therefore, with a probability exceeding

$$
\varepsilon_{r+1} - \frac{1}{|F|} = \varepsilon(\mathcal{C}_{i+1}, 1, B_{r+1}, \theta) \tag{20}
$$

polynomials $p'_{r+1} = p_{r+1} - \Lambda_r(\lambda_{r+1}) \in \mathcal{F}'_{r+1}$, and $f'_{r,0} = f_{r,0} - \Lambda_r(0)$, $f'_{r,1} = f_{r,1} - \Lambda_r(1)$ satisfy

$$
\text{agree}_{\nu_r}((1-\lambda_{r+1}) \cdot f'_{r,0} + \lambda_{r+1} \cdot f'_{r,1}, p'_{r+1}) \geq \alpha \tag{21}
$$

The above satisfied condition can be written as

$$
\Pr\left[\lambda_{r+1} : \begin{array}{c} \exists p'_{r+1} \in \mathcal{F}'_{r+1} \text{ s.t.} \\ \text{agree}_{\nu_r}((1-\lambda_{r+1}) \cdot f'_{r,0} + \lambda_{r+1} \cdot f'_{r,1}, p'_{r+1}) \geq \alpha \end{array}\right] > \varepsilon(\mathcal{C}_{i+1}, 1, B_{r+1}, \theta) \tag{22}
$$

This also satisfies the conditions of the weighted correlated agreement theorem [H24, Theorem 4], so we can obtain polynomials $p'_{r,0}(X), p'_{r,1}(X) \in \mathcal{F}'_{r+1}$, and a set $A_{r+1} \subseteq D_{r+1}$ satisfying:

1. $\nu_r(A_{r+1}) \geq 1-\theta$
2. $p'_{r,0}(X)|_{A_{r+1}} = f'_{r,0}(X)|_{A_{r+1}}$, $p'_{r,1}(X)|_{A_{r+1}} = f'_{r,1}(X)|_{A_{r+1}}$

Now that we have found polynomials $p'_{r,0}(X), p'_{r,1}(X)$, there exist polynomials

$$
p_{r,0}(X) = p'_{r,0}(X) + \Lambda_r(0), \quad p_{r,1}(X) = p'_{r,1}(X) + \Lambda_r(1) \in \mathcal{F}_{r+1} \tag{23}
$$

and

$$
f_{r,0}(X) = f'_{r,0}(X) + \Lambda_r(0), \quad f_{r,1}(X) = f'_{r,1}(X) + \Lambda_r(1) \tag{24}
$$

According to conclusion 2 given by correlated agreement, we can get

$$
p_{r,0}(X)|_{A_{r+1}} = f_{r,0}(X)|_{A_{r+1}}, \quad p_{r,1}(X)|_{A_{r+1}} = f_{r,1}(X)|_{A_{r+1}} \tag{25}
$$

For the multilinear polynomials $P_{r,0}$ and $P_{r,1}$ corresponding to $p_{r,0}(X), p_{r,1}(X)$, according to the definition of $\mathcal{F}'_r$, we can get

$$
\begin{aligned}
P_{r,0}(\omega_{r+2},\ldots,\omega_n) = \Lambda_r(0) \\
P_{r,1}(\omega_{r+2},\ldots,\omega_n) = \Lambda_r(1)
\end{aligned} \tag{26}
$$

Obtain $A_r = \pi^{-1}(A_{r+1}) \subseteq D_r$ by inverse mapping the points in set A_{r+1} through π. At these points, f_r must be consistent with

$$
p_r(X) = p_{r,0}(X^2) + X \cdot p_{r,1}(X^2) \in \mathcal{F}_r \tag{27}
$$

For the multilinear polynomial P_r corresponding to $p_r(X)$, it satisfies

$$
\begin{aligned}
P_r(\omega_{r+1},\omega_{r+2},\ldots,\omega_n) &= (1-\omega_{r+1}) \cdot P_{r,0}(\omega_{r+2},\ldots,\omega_n) + \omega_{r+1} \cdot P_{r,1}(\omega_{r+2},\ldots,\omega_n) \\
&= (1-\omega_{r+1}) \cdot \Lambda_r(0) + \omega_{r+1} \cdot \Lambda_r(1) \\
&= L(\omega_{r+1}, 0) \cdot \Lambda_r(0) + L(\omega_{r+1}, 1) \cdot \Lambda_r(1)
\end{aligned} \tag{28}
$$

From this, we can conclude that the sumcheck in the r-th round is satisfied:

$$
\begin{aligned}
L(\omega_1,\ldots,\omega_r,\lambda_1,\ldots,\lambda_r) \cdot & P_r(\omega_{r+1},\omega_{r+2},\ldots,\omega_n) \\
&= L(\omega_1,\ldots,\omega_r,\lambda_1,\ldots,\lambda_r) \cdot L(\omega_{r+1},0) \cdot \Lambda_r(0) \\
&\quad + L(\omega_1,\ldots,\omega_r,\lambda_1,\ldots,\lambda_r) \cdot L(\omega_{r+1},1) \cdot \Lambda_r(1) \\
&= q_r(0) + q_r(1) \\
&= q_{r-1}(\lambda_r)
\end{aligned} \tag{29}
$$

Now that we have obtained that the sumcheck in the r-th round is satisfied, we need to consider whether the folding relation is satisfied. Consider $x \in \pi^{-1}(A_r)$, we have

$$
\begin{aligned}
&\frac{|\{x \in \pi^{-r}(A_r) : \text{all folding checks hold for } f_0, \ldots, f_r\}|}{|D_0|} \\
&= \frac{1}{|D_0|} \cdot \sum_{y \in A_{r+1}} \delta(y) \cdot |\pi^{-(r+1)}(y)| \\
&= \frac{2^{r+1}}{|D_0|} \cdot \sum_{y \in A_{r+1}} \delta(y) \\
&= \frac{1}{|D_{r+1}|} \cdot \sum_{y \in A_{r+1}} \delta(y) \\
&= \nu_r(A_{r+1})
\end{aligned} \tag{30}
$$

We have already obtained $\nu_r(A_{r+1}) \geq \alpha$ through the correlated agreement theorem, so the proportion of x in D_0 that can satisfy the folding check exceeds α. Combining the sumcheck constraint and folding relation in the r-th round, we get that $(f_0, \Lambda_0, \ldots, f_r, \Lambda_r)$ is α-good for $(\lambda_0, \ldots, \lambda_r)$. Since the probability of generating such a trace set is

$$
\Pr[\mathfrak{T}] > \varepsilon_0 + \ldots + \varepsilon_r \tag{31}
$$

it satisfies the conditions of the lemma. By the induction hypothesis, the lemma holds in the r-th round, so we can conclude that $(g_0, \ldots, g_M) \in \mathcal{R}$. This proves that the lemma also holds in the $(r+1)$-th round. Thus, the lemma is proved. □

References

- [H24] Ulrich Haböck. "Basefold in the List Decoding Regime." *Cryptology ePrint Archive*(2024).