A note on mutual correlated agreement for Reed–Solomon codes

Ulrich Haböck
ulrich@starkware.co,

November 17, 2025

Abstract

We outline how to generalize the Guruswami-Sudan list decoder analysis from Ben–Sasson, Carmon, Ishai, Kopparty and Saraf [BCI+20] in order to obtain a "global" proximity gap, called mutual correlated agreement in [ACFY24], or strong correlated agreement in [Zei24].

1 Introduction

Let $\mathcal{C} = \mathsf{RS}[\mathbb{F}_q, D, k]$ be a Reed–Solomon code over a finite field $\mathbb{F}_q$, with domain of definition $D \subseteq \mathbb{F}_q$, and dimension $k+1$. Let $\mathcal{C}^2$ be the interleaved code of $\mathcal{C}$, consisting of all words $[p_0, p_1]$ with $p_0, p_1 \in \mathcal{C}$. We write $\Delta(f, g)$ for the relative Hamming distance between functions $f, g : D \to \mathbb{F}_q$, and we use the same notation for interleaved words. If two interleaved words $[f_0, f_1]$, $[g_0, g_1]$ agree on a set A of density $\alpha \in (0, 1)$, we say that they have *correlated α-agreement*. In terms of the relative distance, $\Delta([f_0, f_1], [g_0, g_1]) \leq 1 - \alpha$.

For any $f_0, f_1 : D \to \mathbb{F}_q$ and distance $\gamma \in (0, 1)$, we consider

$$\mathcal{L}([f_0, f_1], \gamma) = \{[p_0, p_1] \in \mathcal{C}^2 \ : \ \Delta([p_0, p_1], [f_0, f_1]) \leq \gamma\},$$

be the set of all γ-proximate codewords of $[f_0, f_1]$. This set is empty whenever $\Delta([f_0, f_1], \mathcal{C}^2) > \gamma$, and non-emptiness implies that $\Delta([f_0, f_1], \mathcal{C}^2) \leq \gamma$, in other words: the existence of some $[p_0, p_1] \in \mathcal{C}^2$, which has correlated $(1-\gamma)$-agreement with $[f_0, f_1]$.

Mutual correlated agreement. A *regular* proximity gap states that, if for sufficiently many $z \in \mathbb{F}_q$, the linear combination $f_z = f_0 + z f_1$ is γ-close to $\mathcal{C}$, then $\Delta([f_0, f_1], \mathcal{C}^2) \leq \gamma$ (and consequently $\Delta(f_z, \mathcal{C}) \leq \gamma$ for *every* z). A *global* proximity gap instead aspires the following rough statement: For all, but a few

exceptions in $z \in \mathbb{F}_q$, every proximate of $f_z = f_0 + zf_1$ is obtained as linear combination $p_0 + zp_1$ of some $[p_0, p_1] \in \mathcal{L}([f_0, f_1], \gamma)$. Concretely,

$$\begin{aligned}\mathcal{L}(f_z, \gamma) &= \{p \in \mathcal{C} \ : \ \Delta(p, f_0 + z \cdot f_1) \leq \gamma\} \\ &= \{p_0 + zp_1 \ : \ [p_0, p_1] \in \mathcal{L}([f_0, f_1], \gamma)\},\end{aligned}$$

for all $z \in \mathbb{F}_q \setminus E$, where E the set of exceptions is small. The crucial task is to bound the size of E.

We state the property in its most elementary form, restricting to the simplest form of combinations, i.e. the linear combination of two words.

Definition 1 ([ACFY24])**.** We say that $\mathcal{C} = \mathrm{RS}[\mathbb{F}_q, D, k]$ satisfies *mutual correlated agreement* up to distance θ, and with soundness error $\varepsilon : (0, \theta) \to [0, \infty)$, if: For every $0 < \gamma < \theta$, and words $f_0, f_1 : D \to \mathbb{F}_q$,

$$\Pr\left[z \in \mathbb{F}_q \ : \ \begin{array}{c}\exists A \subset D, \\ |A| \geq (1-\gamma) \cdot n,\end{array} \text{ s.t. } \begin{array}{c} f_0 + z \cdot f_1|_A \in \mathcal{C}|_A, \text{ but} \\ [f_0, f_1]\big|_A \notin \mathcal{C}^2\big|_A \end{array}\right] \leq \varepsilon(\gamma).$$

Mutual correlated agreement is a recent notion which is often used to simplify the security analysis of interactive oracle proofs (or, proofs of proximities), in particular those in which the proximate codewords are subject to certain constraints. The notion was first considered useful in the soundness analysis of Basefold [ZCF23], for which Zeilberger, Chen and Fish observed that it holds for general linear codes up to $\theta = \delta/3$. An early version of [Zei24] proves mutual correlated agreement up to $\theta = 1 - \sqrt[4]{1-\delta}$ (the "double" Johnson radius), while concurrently Gao, Kan, and Li [GKL24] showed its validity even up to $\theta = 1 - \sqrt[3]{1-\delta}$ (the "one-and-a-half" Johnson bound).[1] In the special case of Reed-Solomon codes, Anon, Chiesa, Fenzi and Yogev [ACFY24] proved mutual correlated agreement up to $\delta/2$, and they conjectured it to hold up to the Johnson radius $\theta = 1 - \sqrt{1-\delta}$.

In this note we demonstrate that Reed-Solomon codes satisfy Definition 1 up to the Johnson radius $\gamma = 1 - \sqrt{1-\delta}$ of the code, where $\delta = 1 - \frac{k}{|D|}$ is its minimum distance. We note that restricting to linear combinations is for the sake of simplicity. General combinations of more than two functions, in a multilinear manner, or using powers of z, can be proven similarly, without additional technical obstacles.

Remark. We have let the proof circulate in the community for about a year. Since time flies, and in particular in the light of the recent progress [GG25] and [BCH+25], we decided to eventually share it publicly now, and in an essentially unchanged form (yet somewhat polished). A more verbose update incorporating the improvements from [BCH+25] will be posted thereafter.

[1] In a later version, [Zei24] upgraded to the 1.5 Johnson bound as well.

2 From collinearity to correlated agreement

The following lemma, which goes back to Ames, Hazay, Ishai, Venkitasubramaniam [AHIV17] is used as the final step in many correlated agreement theorems, for example [BKS18, BCI+20, Hab24, GKL24, Zei24, BCH+25]. It states that, given a family of proximate polynomials which lie on a line, then we can infer correlated agreement whenever the family is large enough. We present the lemma in the lossless form, and for general linear codes.

Lemma 1 ([AHIV17, BKS18])**.** *Given an arbitrary linear code $\mathcal{C}$ represented by words over the domain D, and functions $f_0, f_1 : D \to \mathbb{F}_q$. Fix $\gamma \in (0,1)$, and assume two codewords $p_0, p_1 \in \mathcal{C}$, such that $\Delta(p_0 + zp_1, f_0 + zf_1) \leq \gamma$ for all z contained in a set $S \subseteq \mathbb{F}_q$. Then, if*

$$|S| > \lceil \gamma \cdot n \rceil + 1,$$

it holds that $\Delta([a,b],[f_0,f_1]) \leq \gamma$. In other words, $[f_0, f_1]$ agrees with $[a,b] \in \mathcal{C}^2$ over a set of density $1-\gamma$.

Proof. Without loss of generality, we assume that $\gamma = e/n$ for some integer $e \in [0,n]$, where n is the size of the domain D. We show that $\Delta([a,b],[f_1,f_2]) > \gamma$ leads to a contradiction, proving the claim of the lemma.

Assume $e < n \cdot \Delta([a,b],[f_1,f_2])$. Then the disagreement set

$$E = \{x \in D \,:\, (p_0(x), p_1(x)) \neq (f_0(x), f_1(x))\}$$

contains a subset E' of size $e+1$. Denote $f_z = f_0 + zf_1$ and $c_z = p_0 + zp_1$. For any scalar $z \in S$ we have $\Delta(f_z, c_z) \leq e/n$, hence there must be at least one $x \in E'$ for which $c_z(x) = f_z(x)$. On the other hand, each point x from the disagreement set E has at most one $z \in S$ for which $c_z(x) = f_z(x)$. In fact, by linearity of the selection,

$$c_z(x) - f_z(x) = p_0(x) - f_0(x) + z \cdot (p_1(x) - f_1(x)),$$

is a non-trivial linear functional, since $(p_0(x), p_1(x)) \neq (f_0(x), f_1(x))$. Therefore,

$$|S| \leq |E'| = e + 1,$$

in contradiction to the size assumption on S. □

3 Generalizing the analysis from [BCI+20]

The goal of this section is to outline the proof of the following generalization of [BCI+20, Theorem 5.1].

Theorem 2. *Let $\mathcal{C} = \mathrm{RS}[\mathbb{F}_q, D, k]$ be the Reed-Solomon code over F with domain of definition D of size $|D| = n$ and miminum distance $\delta = 1 - \frac{k}{n}$ and let $\rho = \frac{k}{n}$. Take $\gamma = 1 - (1 + \frac{1}{2m}) \cdot \sqrt{\rho}$ for some integer $m \geq 3$. Then for any $f_0, f_1 \in \mathbb{F}_q^D$ the size of*

$$E = \left\{ z \in \mathbb{F}_q \,:\, \begin{array}{c} \exists A \subset D, \\ |A| \geq (1-\gamma) \cdot n, \end{array} \; s.t. \; \begin{array}{l} f_z|_A \in \mathcal{C}|_A, \; but \\ [f_0, f_1]\big|_A \notin \mathcal{C}^2\big|_A \end{array} \right\}$$

is bounded as

$$|E| \leq \frac{\ell^7}{3} \cdot (\rho \cdot n)^2,$$

where $\ell = (m + 1/2)/\sqrt{\rho}$.

Remark 3. This is essentially the same bound as for ordinary correlated agreement in [BCI+20, Theorem 5.1], with the minimal difference that in our notation ρ is the slightly reduced rate of the code.

The key idea in [BCI+20] is to run the Guruswami–Sudan list decoder on the word $f = f_0 + Zf_1$, where Z is a formal variable, and draw conclusions from its known behavior for those $Z = z$, for which $f_z = f_0 + zf_1$ is γ-proximate to the code $\mathcal{C}$. This amounts to studying the Guruswami–Sudan decoder over $\mathbb{K} = \mathbb{F}_q(Z)$ the field of rational functions over $\mathbb{F}_q$, and its analysis requires certain familiarity with algebraic function fields.

Our proof of Theorem 2 requires merely a careful inspection of the steps in [BCI+20], Section 5. Given multiplicity parameter $m \geq 1$, consider the Guruswami-Sudan interpolating polynomial of the $\mathbb{K}$-valued word $f_0 + Zf_1$ as constructed in Section 5.2.1 therein, decomposed into polynomials

$$Q(X, Y, Z) = C(X, Z) \cdot \prod_i R_i(X, Y^{p^{f_i}}, Z)^{e_i},$$

where each $R_i(X, Y, Z)$ is irreducible and separable, and p is the characteristic of F. The degrees of Q w.r.t. X, Y and YZ, are

$$\begin{aligned} D_Y &< \ell, \\ D_X &< \ell \cdot \rho \cdot n, \\ D_{YZ} &\leq \frac{\ell^3}{6} \cdot \rho \cdot n, \end{aligned}$$

see [BCI+20, Claim 5.4]. The discriminant in Y has $\deg_X \mathrm{disc}_Y(Q) < \ell^2 \cdot \rho \cdot n$, if it is not the zero polynomial, and this observation is used in [BCI+20, Section 5.2.3] to argue the existence of a point $x_0 \in F$ for which the discriminant of each irreducible factor

$$\mathrm{disc}_Y R_i(x_0, Y, Z) \neq 0,$$

for every i, as long as $|F| > \ell^2 \cdot \rho \cdot n$. (For $|F| \leq \ell^2 \cdot \rho \cdot n$ the statement of Theorem 2 is trivial.) This point x_0 serves as the starting point for the power series construction, the Hensel lift.

In analogy[2]to the decomposition used by Zeilberger [Zei24] and Gao, Kan and Li [GKL24], we decompose the folding proximates with respect to the irreducible factor they belong to. That is, we write $E = \bigcup_i E_i$, where

$$E_i = \left\{ z \in \mathbb{F}_q \ : \ \begin{array}{c} \exists A \subset D, \\ |A| \geq (1-\gamma)\cdot n, \end{array} \text{ s.t. } \begin{array}{c} \exists p(X) \in \mathcal{L}, \text{ with } f_z|_A = p|_A, \\ R_i(X, p(X)^{p^{f_i}}, z) = 0, \text{ but} \\ [f_0, f_1]\big|_A \notin \mathcal{C}^2\big|_A \end{array} \right\},$$

where $\mathcal{L} = F[X]^{<k}$. We then refine each E_i w.r.t. the irreducible components $R_i(x_0, Y, Z) = \prod_j H_{i,j}(Y, Z)$, as required for the potentially different starting values of the Hensel lift. That is we consider

$$E_{i,j} = \left\{ z \in F \ : \ \begin{array}{c} \exists A \subset D, \\ |A| \geq (1-\gamma)\cdot n, \end{array} \text{ s.t. } \begin{array}{c} \exists p(X) \in \mathcal{L}, \text{ with } f_\lambda|_A = p|_A, \\ R_i(X, p(X)^{p^{f_i}}, z) = 0, \\ H_{i,j}(p(x_0)^{p^{f_i}}, z) = 0, \\ \text{but } [f_0, f_1]\big|_A \notin \mathcal{C}^2\big|_A \end{array} \right\}.$$

Still, the number of (i, j)'s is bounded by $D_Y < \ell$, and Theorem 2 immediately follows from the following

Claim 1. *For each (i, j),*

$$|E_{i,j}| \leq \frac{\ell^6}{3} \cdot (\rho \cdot n)^2.$$

Proof. We prove the claim by contradiction. Suppose that $|E_{i,j}| > \frac{\ell^6}{3} \cdot (\rho \cdot n)^2$ for some (i, j), and for each $z \in E_{i,j}$ choose any $p_z(X) \in \mathcal{L}$ satisfying the claims in the definition of the set. Then $R = R_i(X, Y, Z)$ is a "uselful" factor in the sense of [BCI+20], having an irreducible factor $H = H_{i,j}(Y, Z)$ of $R_i(x_0, Y, Z)$ for which

$$S_{x_0,R,H} = \{z \in E \ : \ R(X, p_z(X)^{p^{f_i}}, z) \ \wedge \ H(p_z(x_0)^{p^{f_i}}, z) = 0\}$$

is of size

$$|S_{x_0,R,H}| > \frac{\ell^6}{3} \cdot (\rho \cdot n)^2 > 2 \cdot D_Y^2 \cdot D_X \cdot D_{YZ},$$

as stated by [BCI+20, Claim 5.7]. This is sufficient to let Step 5 – Step 7 of the analysis go through, and their generalization [BCI+20, Appendix C] for the inseparable case $f_i > 0$, showing that there exist polynomials $a(X), b(X) \in \mathcal{L}$, so that

$$\begin{aligned} R_i(X, Y^{p_i^{f_i}}, Z) &= Y^{p^{f_i}} - (a(X) + Z \cdot b(X))^{p^{f_i}} \\ &= (Y - (a(X) + Z \cdot b(X)))^{p^{f_i}}, \end{aligned}$$

[2]Their proofs rely on set intersection games, and the decomposition of E is done with respect to the various lines of polynomials which are constructed from agreement set intersections. In the algebraic world, and for $f_i = 0$, each factor $R_i(X, Y, Z)$ is expected to be of the form $Y - p_0(X) + Zp_1(X)$ with polynomials $p_0(X), p_1(X)$ of degree at most k, representing lines.

and therefore

$$p_z(X) = a(X) + z \cdot b(X).$$

In particular, the folding proximate $p(X) \in \mathcal{L}$ in the definition of $E_{i,j}$ is unique. Hence

$$E_{i,j} = \left\{ z \in F \ : \ \begin{array}{c} \exists A \subset D, \\ |A| \geq (1-\theta) \cdot n, \end{array} \text{ s.t. } \begin{array}{c} f_z|_A = a(X) + z \cdot b(X)|_A, \\ \text{but } [f_0, f_1]\big|_A \notin \mathcal{C}^2\big|_A \end{array} \right\},$$

meaning that every $z \in E_{i,j}$ must improve agreement beyond the correlated agreement set $A^\circ = \{x \in D \ : \ (a(x), b(x)) = (f_0(x), f_1(x))\}$. From the proof of Lemma 1 we understand that the number of scalars which improve agreement beyond A° is bounded by $|D \setminus A^\circ| \leq n$, in contradiction to the size of $E_{i,j}$, since $\frac{\ell^6}{3} \cdot (\rho \cdot n)^2 = \frac{(m+1/2)^6}{3 \cdot \rho} \cdot n^2 > n$ regardless of the concrete choice of parameters. $\square$

References

[ACFY24] Gal Arnon, Alessandro Chiesa, Giacomo Fenzi, and Eylon Yogev. WHIR: Reed–solomon proximity testing with super-fast verification. In *IACR preprint archive 2024/1586*, 2024. `https://eprint.iacr.org/2024/1586`.

[AHIV17] Scott Ames, Carmit Hazay, Yuval Ishai, and Muthuramakrishnan Venkitasubramaniam. Ligero: Lightweight sublinear arguments without a trusted setup. In *CCS 2017*, 2017. Full paper: `https://eprint.iacr.org/2022/1608`.

[BCH+25] Eli Ben-Sasson, Dan Carmon, Ulrich Haböck, Swastik Kopparty, and Shubhangi Saraf. On proximity gaps for Reed–Solomon codes. Cryptology ePrint Archive, Paper 2025/2055, 2025. `https://eprint.iacr.org/2025/2055`.

[BCI+20] Eli Ben-Sasson, Dan Carmon, Yuval Ishai, Swastik Kopparty, and Shubhangi Saraf. Proximity gaps for Reed–Solomon codes. In *FOCS 2020*, 2020. Full paper: `https://eprint.iacr.org/2020/654`.

[BKS18] Eli Ben-Sasson, Swastik Kopparty, and Shubhangi Saraf. Worst-case to average case reductions for the distance to a code. In *CCC'18*, 2018. Full paper: `https://eccc.weizmann.ac.il/report/2018/090/`.

[GG25] Rohan Goyal and Venkatesan Guruswami. Optimal proximity gaps for subspace-design codes and (random) Reed–Solomon codes. Cryptology ePrint Archive, Paper 2025/2054, 2025. `https://eprint.iacr.org/2025/2054`.

[GKL24] Yiwen Gao, Haibin Kan, and Yuan Li. Linear proximity gap for reed–solomon codes within the 1.5 johnson bound. In *IACR ePrint Archive 2024/1810*, 2024. https://eprint.iacr.org/2024/1810.

[Hab24] Ulrich Haböck. Basefold in the list decoding regime. Cryptology ePrint Archive, Paper 2024/1571, 2024. https://eprint.iacr.org/2024/1571.

[ZCF23] Hadas Zeilberger, Binyi Chen, and Ben Fisch. BaseFold: efficient field-agnostic polynomial commitment schemes from foldable codes. In *IACR preprint archive 2023/1705*, 2023. https://eprint.iacr.org/2023/1705.

[Zei24] Hadas Zeilberger. Khatam: Reducing the communication complexity of code-based SNARKs. In *IACR preprint archive 2024/1843*, 2024. https://eprint.iacr.org/2024/1843.