Electronic Colloquium on Computational Complexity, Report No. 145 (2021)

Revisiting a Lower Bound on the Redundancy of Linear Batch Codes

Omar Alrabiah Venkatesan Guruswami

Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213.
Email: {oalrabia,venkatg}@andrew.cmu.edu

Abstract

A recent work of [LW21] shows a redundancy lower bound of $\Omega(\sqrt{Nk})$ for systematic linear k-batch codes of block length N by looking at the $O(k)$ tensor power of the dual code. In this note, we present an alternate proof of their result via a linear independence argument on a collection of polynomials.

1 Result statement

Batch codes are a family of codes introduced by [IKOS04] for applications in load balancing. Following Definition 1.1 in [LW21], a linear batch code is formally defined as follows.

Definition 1 (Linear Batch Codes). *For a field $\mathbb{F}$, Let $C \leqslant \mathbb{F}^N$ be a linear code of dimension n. The code C is a systematic linear k-batch code if for any multiset of indices $\{i_1, \dots, i_k\} \subseteq [n]$, there exist k mutually disjoint sets $R_1, \dots, R_k \subseteq [N]$ and linear functions $g_1, \dots, g_k$ such that $g_j(\mathbf{c}|_{R_j}) = c_{i_j}$ for all codewords $\mathbf{c} \in C$ and $j \in [k]$.*

Recently in [LW21], they prove the following upper bound on the rate of systematic linear batch codes by looking at the $O(k)$'th tensor power of $C^{\perp}$.

Theorem 1 ([LW21]). *Given a systematic linear k-batch code $C \leqslant \mathbb{F}^N$, we have* $\dim(C) \leqslant N - \Omega(\sqrt{Nk})$.

In this note, we give an alternate presentation of the lower bound proved in [LW21] for systematic linear k-batch codes. Our approach uses polynomials in a fashion similar to the approach in [Woo16] (see also [RV16] for a related perspective using vector products). The proof proceeds in two steps. We first convert the definition of a systematic linear $3t$-batch code into something that we call a t-ordered-batch codes. We then work with t-ordered-batch codes to show the redundancy lower bound by constructing a collection of polynomials and then showing that they are linearly independent.

ISSN 1433-8092

2 Proof

As part of our proof, we define the notion of ordered-batch codes and then proceed to show a reduction from linear systematic batch codes to ordered-batch codes.

Definition 2 (Ordered-Batch Codes)**.** *For a field $\mathbb{F}$, Let $C \leqslant \mathbb{F}^N$ be a linear code of dimension n. The code C is a t-ordered-batch code if for any set of indices $S = \{i_1, \dots, i_t\} \subseteq [n]$, there exist $2t$ mutually disjoint sets $A_1 \dots, A_t, B_1, \dots, B_t \subseteq [N]$ and linear functions $g_1, \dots, g_t, h_1, \dots, h_t$ such that $g_j(\mathbf{c}|_{A_j}) = h_j(\mathbf{c}|_{B_j}) = c_{i_j}$ for all codewords $\mathbf{c} \in C$ and $j \in [t]$. Moreover, the repair groups satisfy the following additional property: consider a directed graph D_S with vertices S and edges $i_j \to i_k$ if $i_k \in A_j \cup B_j$. Then the graph D_S is a DAG.*

Proposition 2. *If a linear code $C \leqslant \mathbb{F}^N$ is a systematic linear $3t$-batch code, then it is also a t-ordered-batch code.*

Proof. Consider a systematic linear $3t$-batch code C. By applying the definition of systematic batch codes for the multiset $\{i_1, i_1, i_1, i_2, \ i_2, i_2, \dots, i_t, i_t, i_t\}$ (the multiset where each of the elements of the set $S = \{i_1, \dots, i_t\} \subseteq [n]$ occur exactly three times), each element i_j obtains three repair groups R_j^1, R_j^2, R_j^3, where all $3t$ repair groups are subsets of $[N]$ and are all mutually disjoint. Now, consider the directed graph D_S with S as its vertices, and the edges are $i_j \to i_k$ if $i_k \in R_j^\epsilon$ for some $\epsilon \in \{1, 2, 3\}$ (D_S might also have self-loops). Because the repair groups are mutually disjoint, the in-degree of every vertex in D_S is at most 1. Thus the directed cycles of D_S are vertex-disjoint. That's because if two cycles C_1 and C_2 have a common vertex v. then the incoming edges to v from the cycles C_1 and C_2 must be the same as the in-degree of v is at most 1. Thus the previous vertex of v in both C_1 and C_2 is the same, and call it u. We can repeat the argument for the vertex u, and by iteration, we would deduce that all the edges of the cycles C_1 and C_2 are the same. Thus $C_1 = C_2$. Now, because all the cycles of D_S are vertex-disjoint, then we can remove a collection E_0 of vertex-disjoint edges such that D_S becomes a DAG. Since each edge has a uniquely associated repair group and the collection E_0 is vertex-disjoint, then that means that we can remove at most one repair group from each $i_j \in S$ such that the new directed graph D_S is now a DAG. $\square$

Thus we have shown that a systematic linear $3t$-batch code implies a t-ordered-batch code. Next, we are going to show a lower bound on the redundancy of a t-ordered-batch code, which by Proposition 2 yields us Theorem 1.

Theorem 3. *For a t-ordered-batch code $C \leqslant \mathbb{F}^N$ of dimension n and redundancy r (so $N = n+r$), we have the inequality $\binom{r+2t-1}{2t} \geqslant \binom{n}{t}$. As such, $r = \Omega(\sqrt{tn})$.*

Proof. First, let us setup the viewpoint for the dual code $C^\perp$ that we shall follow in this proof. Let $G^\perp \in \mathbb{F}^{N \times r}$ denote the generator matrix for $C^\perp$. Let ω_i denote the i'th row of $G^\perp$. Then by those definitions, we see that for any dual codeword $c^\perp \in C^\perp$, we can find an $\alpha \in \mathbb{F}^r$ such that $c^\perp = G^\perp \alpha = (\langle \alpha, \omega_1 \rangle, \dots, \langle \alpha, \omega_N \rangle)^\top$.

Now, for any t pairwise distinct elements $S = \{i_1, ..., i_t\} \subseteq [n]$, by applying the t-ordered-batch code property to the set $\{i_1, i_2, \dots, i_t\}$, we can find pairwise disjoint repair groups $\{A_1, \dots, A_t\} \cup \{B_1, \dots, B_t\}$ contained in $[N]$ such that their associated directed graph D_S is a DAG. Moreover, we can find dual codewords $\{a_j\}_{j=1}^t \cup \{b_j\}_{j=1}^t \subseteq C^\perp$ satisfying $i_j \in \mathrm{Supp}(\ell_j) \subseteq L_j \cup \{i_j\}$ for all $j \in [t]$ and $(\ell, L) \in \{(a, A), (b, B)\}$. By our argument in the beginning, this means that there are $V_S \coloneqq \{\alpha_j\}_{j=1}^t \cup \{\beta_j\}_{j=1}^t \subseteq \mathbb{F}^r$ such that $\langle \lambda_j, w_k \rangle \neq 0$ if and only if $k \in L_j \cup \{i_j\}$ for $j \in [t]$ and $(\lambda, L) \in \{(\alpha, A), (\beta, B)\}$.

Now, for $X = (x_1, \ldots, x_r)$ with x_i being an indeterminate over $\mathbb{F}$, define the polynomial

$$p_S(X) \coloneqq \prod_{j=1}^{t} \langle \alpha_j, X \rangle \langle \beta_j, X \rangle$$

We claim that the collection of polynomials $\{p_S \mid S \subseteq [n], |S| = t\}$ are linearly independent. The inequality then follows as there are $\binom{n}{t}$ such polynomials. On the other hand, the polynomials p_S are homogeneous polynomials of degree $2t$ over r variables, and so the dimension of their span is at most $\binom{r+2t-1}{2t}$.

Consider variables $z_1, \ldots z_N$ over $\mathbb{F}$. Plug in $X = \sum_{k=1}^{N} z_k \omega_k$ in p_S to obtain the homogeneous polynomial

$$\begin{aligned} q_S(z_1, \ldots, z_N) \coloneqq p_S\left(\sum_{k=1}^{N} z_k \omega_k\right) &= \prod_{j=1}^{t} \left\langle \alpha_j, \sum_{k=1}^{N} z_k \omega_k \right\rangle \left\langle \beta_j, \sum_{k=1}^{N} z_k \omega_k \right\rangle \\ &= \prod_{j=1}^{t} \left(\sum_{k=1}^{N} z_k \langle \alpha_j, \omega_k \rangle\right) \left(\sum_{k=1}^{N} z_k \langle \beta_j, \omega_k \rangle\right) \end{aligned}$$

To show that the p_S's are linearly independent, it suffices for us to show that the q_S's are linearly independent. This follows by the fact that the map $p(X) \mapsto p\left(\sum_{k=1}^{N} z_k \omega_k\right)$ is a linear map, and the images of the p_S's are the q_S's. Now, to show that the q_S's are linearly independent, we will show that for any set $T \subseteq [N]$ of size t, the monomial $\prod_{i \in T} z_i^2$ has a nonzero coefficient in q_S if and only if $T = S$. From this claim, the linear independence of $\{q_S \mid S \subseteq [n], |S| = t\}$ then follows.

Indeed, now, for any $k \notin S$, the degree of z_k in p_S is at most 1. This follows from the fact that the repair groups $\{A_j\}_{j=1}^{t} \cup \{B_j\}_{j=1}^{t}$ are mutually disjoint, meaning that z_k appears at most once in the repair groups $\{A_j\}_{j=1}^{t} \cup \{B_j\}_{j=1}^{t}$ and thus once in the product-form of q_S. This then means that if the monomial $\prod_{i \in T} z_i^2$ has a nonzero coefficient, then $i \in S$ for all $i \in T$. By homogeneity, we must have $T = S$.

Now, to show that the monomial $\prod_{i \in S} z_i^2$ has a nonzero coefficient, it suffices for us to show that in the expansion of q_S, the monomial $\prod_{i \in S} z_i^2$ occurs only once, and so it must have a nonzero coefficient. We have

$$\begin{aligned} q_S(z_1, \ldots, z_N) &= \prod_{j=1}^{t} \left(\sum_{k=1}^{N} z_k \langle \alpha_j, \omega_k \rangle\right) \left(\sum_{k=1}^{N} z_k \langle \beta_j, \omega_k \rangle\right) \\ &= \sum_{\substack{(u_1, \ldots u_t) \in [N]^t \\ (v_1, \ldots v_t) \in [N]^t}} \left(\prod_{j=1}^{t} \langle \alpha_j, \omega_{u_j} \rangle \langle \beta_j, \omega_{v_j} \rangle\right) \prod_{j=1}^{t} z_{u_j} z_{v_j} \end{aligned}$$

Notice that the coefficient of the monomial is nonzero if and only if $u_j \in A_j \cup \{i_j\}$ and $v_j \in B_j \cup \{i_j\}$ for all $j \in [t]$. If the multiset $\{u_j\}_{j=1}^{t} \cup \{v_j\}_{j=1}^{t}$ is the same as the multiset $S \cup S$, then consider the directed graph G on S with edges $i_j \to u_j$ if $u_j \neq i_j$ and edges $i_j \to v_j$ if $v_j \neq i_j$. In this directed graph G, there are no self-loops. Moreover, the in-degree of every vertex is equal to its out-degree for the following reasoning: if we include the edges $i_j \to u_j$ if $u_j = i_j$ and $i_j \to v_j$ if $v_j = i_j$, then every vertex in D_S will have an out-degree of 2, and since the multiset $\{u_j\}_{j=1}^{t} \cup \{v_j\}_{j=1}^{t}$ is the same as the multiset $S \cup S$, then the in-degree of every vertex is 2. Thus every vertex in this new graph has equal in-degree and out-degree. Since the edges that we added are self-loops, then removing them won't affect the equality between the in-degree and out-degree.

This means that G can be decomposed into a disjoint union of cycles, but since the edges of G are a subcollection of the edges of D_S, and the graph D_S has no directed cycles, then G must be the empty graph, which means $u_j = v_j = i_j$ for all $j \in [t]$. Thus the monomial $\prod_{i \in S} z_i^2$ occurs exactly once in the expansion of q_S. $\square$

Acknowledgments

We thank Ray Li and Mary Wootters for useful discussions and feedback.

References

[IKOS04] Yuval Ishai, Eyal Kushilevitz, Rafail Ostrovsky, and Amit Sahai. Batch codes and their applications. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 262–271, 2004.

[LW21] Ray Li and Mary Wootters. Improved batch code lower bounds. *arXiv preprint arXiv:2106.02163*, 2021.

[RV16] Sankeerth Rao and Alexander Vardy. Lower bound on the redundancy of PIR codes. *arXiv preprint arXiv:1605.01869*, 2016.

[Woo16] Mary Wootters. Linear codes with disjoint repair groups. `https://web.stanford.edu/~marykw/files/disjoint_repair_groups.pdf`, 2016.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il