Interior-point regenerating codes on graphs

Adway Patra Alexander Barg

***Abstract*—We consider the use of regenerating codes in distributed storage systems where connections between the nodes are constrained by a graph. In this setting the cost of node repair is determined by the graphical distance from the helper nodes to the failed node. In our recent work (arXiv:2108:00939) we considered the MSR case, showing that linear MSR codes are amenable to intermediate processing of the information, resulting in reduced repair bandwidth which also meets the lower bound on the minimum repair cost. Here we extend this study to the non-MSR case. We derive a *lower bound* on the repair bandwidth and formulate *repair procedures with intermediate processing* for several families of regenerating codes, with an emphasis on the recent constructions from multilinear algebra. We also consider intermediate processing for the problem of *partial node repair*.**

I. Introduction

A distributed storage system (DSS) is formed of a number of nodes connected by communication links which carry the information to accomplish the two basic tasks performed in the system, namely data recovery and node repair. The amount of information sent over the links is a key metric of the system efficiency. The problem of node repair has been widely studied in the literature in the last decade following its introduction in [1]. A DSS is modeled as n storage nodes each with a storage capacity of l units, used to store a file $\mathscr{F}$ of size M, such that the following two properties are met:

- *(Reconstruction)* The entire file can be recovered by accessing any $k < n$ nodes.
- *(Repair)* If a single node fails, data from d surviving, or helper, nodes is used to restore the lost data. We assume that each of the helper nodes contributes $\beta \le l$ units of data, and that $k \le d \le n-1$. The parameter β is called the *per-node repair bandwidth*.

We write the parameters of a regenerating code as (n, k, d, β, l, M). The fundamental tradeoff between the file size and the repair bandwidth is expressed by the bound of [1] which has the form

$$M \le \sum_{i=1}^{k} \min\{l, (d-i+1)\beta\}. \qquad (1)$$

In this work we focus on the repair problem, which exists in two versions, namely functional and exact repair. While for functional repair the entire bound (1) is achievable, for the more stringent exact repair requirement there is a gap between the achievable file size and the bound, first demonstrated in [14] in an example and then extended in [7], [11] to all sets of parameters (n, k, d). This bound can be attained for the two corner points of the storage/bandwidth curve (1), giving rise, respectively, to Minimum Storage Regenerating (MSR) codes and Minimum Bandwidth Regenerating (MBR) codes. The MSR point is the most widely studied in the literature, giving rise to a variety of interesting families of algebraic codes. Some of these families are mentioned below, and we refer to the survey [9] for a detailed overview. The main focus of this work is on interior points of the trade-off curve. Several interior-point code families are known, among them constructions in [5], [12], [4], and [2].

Problem statement and our results: We consider a variant of the node repair problem in which communication between the nodes is constrained by a (connected) graph $G(V, E)$ and the cost of sending a unit of information from v_i to v_j is determined by the graph distance $\rho(v_i, v_j)$ in G. This assumption results in a bias in the information cost of node repair in favor of the helper nodes closer to the failed node v_f, and suggests that the helper nodes located closer to the failed node v_f combine the information received from the outer extremes of the helper set before relaying it to the failed node. In our earlier work [8], which also introduced this problem, we called this approach *Intermediate Processing*, or IP (as opposed to direct relaying). In [8] we considered the MSR case of this problem, showing that linear MSR codes can be modified to implement IP, resulting in savings in the repair bandwidth over simple relaying.

Here we study a general version of this problem, considering regenerating codes on graphs for interior points on the tradeoff curve. In Sec. II we prove a general lower bound on the repair bandwidth which generalizes a result of [8]. Our main results are related to implementing the IP techniques for interior-point codes, and our main focus is *evaluation codes*[1]. We start by rephrasing the IP repair for product-matrix codes (which are still MSR codes) and their recent extension in [3], which sets the stage for later parts of the paper. Then in Sec. III-D we turn to interior point codes. Of the general interior-point code families, the moulin codes of [2] fall under the evaluation category, and we show that repair for these codes can be adjusted to support IP. The same claim applies to determinant codes, omitted from this extended abstact. We note however that, unlike the MSR case, there exists a gap between the lower bound on the minimum possible required information transmission and what is achievable using the constructions designed here.

II. Bounds on the Repair Bandwidth

For a finite field $F = \mathbb{F}_q$ we consider a code $\mathscr{C} \subset F^{nl}$ whose codewords are represented by $l \times n$ matrices over F. We assume that each coordinate (a vector in F^l) is written on a single storage node, and that a failed node amounts to

The authors are with Dept. of ECE and ISR, University of Maryland, College Park, MD 20742. Emails: {apatra,abarg}@umd.edu. A. Barg is also with IITP, Russian Academy of Sciences, 127051 Moscow, Russia. This research was supported by NSF grants CCF2110113 and CCF2104489.

[1] We use this term loosely to refer to a code whose encoding can be phrased as evaluation of a linear functional written in a convenient algebraic way.

978-1-6654-2159-1/22/$31.00 ©2022 IEEE

having its coordinate erased. Suppose that each coordinate of a codeword $C \in \mathscr{C}$ is written on a vertex of a graph $G(V, E)$ with $|V| = n$. Suppose further that the coordinate C_f for some $f \in [n]$ is erased, i.e., that the node $f \in [n]$ has failed. Let $D \subset V \backslash \{v_f\}, |D| = d$ be the set of nodes in the graph G that are closest to v_f in terms of graph distance. This set can be found by a breadth-first search with v_f as the root node. Let $G_{f,D} = (V_{f,D}, E_{f,D})$ be the subgraph spanned by $D \cup \{v_f\}$. To repair the failed node, the helper nodes provide information which is communicated to v_f over the edges in $E_{f,D}$. Each helper node in the graph, starting from the nodes farthest from the failed node, sends its repair data (β symbols each) to the next node along the shortest path towards v_f. An intermediate node can simply collect this data, supplement it with its own information, and forward it along the path to v_f (Accumulate-and-Forward, or AF). The AF technique can be wasteful in high-depth repair graphs since the same data gets transmitted multiple times. This gives rise to the problem of attaining savings by processing the information in the intermediate nodes relying on the IP approach. This idea has been already validated for MSR codes in [8].

A. *Lower bounds on the repair bandwidth*

In this section we derive a lower bound on the minimum required transmission for a set of helper nodes for repair of a failed node. Suppose that the information stored at the vertices is given by random variables $W_i, i \in [n]$ that have some joint distribution on $(F^l)^n$ and satisfy $H(W_i) = l$ for all i, where $H(\cdot)$ is the entropy. For a subset $A \subset V$ we write $W_A = \{W_i, i \in A\}$. Denote by S_i^f the information provided to v_f by the ith helper node in the traditional fully connected repair scheme, and let $S_D^f = \{S_i^f, i \in D\}$. By definition we have

$$H(S_i^f) = \beta, \quad H(S_i^f|W_i) = 0, \quad H(W_f|S_D^f) = 0.$$

We also assume that $H(\mathscr{F}|W_B) = 0$ for any $B \subset [n], |B| = k$, which supports the data reconstruction property. The following result was proved in [13]:

Lemma II.1. *For any $A \subset [n], |A| \le d$ and $i \notin A$*

$$H(W_i|W_A) \le \min(l, (d - |A|)\beta).$$

The next lemma forms a simple extension of our Lemma II.1 in [8], generalizing it to all exact regenerating codes.

Lemma II.2. *Let $v_f, f \in [n]$ be the failed node. For a subset of the helper nodes $E \subset D$ let R_E^f be a function of S_E^f such that*

$$H(W_f|R_E^f, S_{D\backslash E}^f) = 0. \tag{2}$$

If $|E| \ge d-k+1$, then $H(R_E^f) \ge (d-k+1)\beta$. In particular, at the MSR point we have $H(R_E^f) \ge l$.

Proof: By the assumption (2), given the contents of all the nodes in $D\backslash E$, the information contained in R_E^f is sufficient to repair v_f, i.e.,

$$H(W_f|R_E^f, W_{D\backslash E}) = 0. \tag{3}$$

We have $|D\backslash E| \le k - 1$. Consider a set $A \subset E$ with $|A| = k - 1 - |D\backslash E|$. Now, by (3)

$$H(R_E^f, W_{D\backslash E}, W_A) = H(R_E^f, W_{D\backslash E}, W_f, W_A) = M, \tag{4}$$

where the first equality in (4) follows from (3) and the chain rule, and the second follows from reconstruction property because $|D\backslash E| + |A| + 1 = k$. Next observe that

$$H(R_E^f, W_{D\backslash E}, W_A) \le H(R_E^f) + H(W_{D\backslash E}, W_A),$$

and so

$$\begin{aligned} H(R_E^f) &\ge M - H(W_{D\backslash E}, W_A) \\ &\ge M - \sum_{i=1}^{k-1} \min\{l, (d-i+1)\beta\} \end{aligned}$$

where the last inequality follows from Lemma II.1. The largest value of M is given in (1), implying the claim of the lemma. □

Note that at the MSR point $(d - k + 1)\beta = l$ and we recover Lemma II.1 from [8]. In that work we also showed that $H(R_E^f) = l$ is achievable at the MSR point. At the same time for any other point on the tradeoff curve, $(d-k+1)\beta < l$. Below in this paper we show that the value $H(R_E^f) = l$ can be achieved by some code families, and hence there is a possibility for improvement of the bound. The following lemma from [7] shows that this bound can indeed be sometimes improved.

Lemma II.3 ([7], Lemma 2)**.** *For any pair of disjoints sets $E, B \subseteq D$ with $i \notin E \cup B$, we have*

$$H(S_E^f|W_B) \ge \frac{|E|}{d - |B|} H(S_D^f|W_B).$$

Taking $B = \emptyset$ and noting that $H(S_D^f) \ge H(W_f) \ge l$, we obtain

Corollary II.4. *For any $E \subset D$, $H(R_E^f) \ge \frac{|E|l}{d}$.*

If $|E| > (d - k + 1)\beta(d/l)$ then this result is better than the claim of the lemma at the interior points.

The constructions presented below do not reach the bounds proved here, leaving an open question of the optimal repair bandwidth for the IP repair technique.

III. Intermediate Processing for Evaluation Codes

In this section we show that F-linear regenerating codes support repair on graphs with lower communication complexity compared to the AF strategy. We begin with an alternative description of node repair using the IP strategy at the MSR point for product-matrix codes and their generalization [3] (the latter result is formally new). The main purpose of the MSR part of this section is to develop intuition which will enable us to define node repair for interior-point codes in Sec. III-D.

A. *Product-matrix (PM) codes*

In this section we rewrite the IP repair of PM codes originally introduced in [8, Sec.II.A] to fit the evaluation code paradigm. PM codes, constructed in [10], form a family of MSR codes with parameters $[n, k, d = 2(k-1), l = k-1, \beta = 1, M = k(k-1)]$. In the original description, the data file $\mathscr{F}$ is represented by two symmetric matrices S_1, S_2 of order $k-1$, accounting for a total of M independent symbols, and the encoding is defined as multiplication of $(S_1|S_2)^\intercal$ by some well-chosen Vandermonde-like matrices. To phrase this differently, let $s_1(y, z)$ and $s_2(y, z)$ be two symmetric polynomials over F of degree at most $k-2$ in each of the two variables. Because of symmetry, the total number of independent coefficients is M, so s_1, s_2 can be used to represent $\mathscr{F}$. Letting $a_1, \dots, a_n$ be distinct points of F, we let node i store the l coefficients of the polynomial $g^{(i)}(z) = s_1(a_i, z) + a_i^{k-1} s_2(a_i, z)$ for all $i \in [n]$.

Using this definition of the codes, the IP repair process of [8] can be phrased as follows. Let $f \in [n]$ be the failed node, let D be the set of d helpers, and let A be a set of helper nodes of size at least $d-k+1 = k-1$. For $h \in D$ define the polynomial

$$l^{(h)}(z) = \sum_{j=0}^{d-1} l_j^h z^j := \prod_{\substack{i \in D \\ i \neq h}} \frac{z - a_i}{a_h - a_i} \tag{5}$$

of degree at most $d-1$. Then the set A transmits the l-vector

$$\xi(f, A) = \sum_{h \in A} g^{(h)}(a_f) \begin{bmatrix} l_0^h + a_f^{k-1} l_{k-1}^h \\ l_1^h + a_f^{k-1} l_k^h \\ \vdots \\ l_{k-2}^h + a_f^{k-1} l_{2k-3}^h \end{bmatrix}. \tag{6}$$

We show that (i) the failed node can recover its value based on the vector $\xi(f, D)$, and (ii) the intermediate nodes can save on the repair bandwidth by processing the received information. To show (i) we prove

Lemma III.1. *The content of the failed node f coincides with the vector $\xi(f, D)$, i.e., $g^{(f)}(z) = \sum_{i=0}^{l-1} (\xi(f, D))_i \, z^i$.*

Proof. Consider the polynomial $H(z) = s_1(a_f, z) + z^{k-1} s_2(a_f, z)$ and note that $\deg(H) \le 2k-3 = d-1$. Thus if we write $H(z) = \sum_{j=0}^{d-1} g_j z^j$, then the polynomial $g^{(f)}$ defined above can be written as

$$g^{(f)}(z) = \textstyle\sum_{j=0}^{k-2} (g_j + a_f^{k-1} g_{k-1+j}) z^j.$$

Rephrasing, the contents of the node f is

$$(g_0 + a_f^{k-1} g_{k-1}, g_1 + a_f^{k-1} g_k, \dots, g_{k-2} + a_f^{k-1} g_{2k-3})^{\intercal}.$$

At the same time, using (5) we can write $H(z)$ in the Lagrange form $H(z) = \sum_{h \in D} g^{(h)}(a_f) l^{(h)}(z)$. The coefficient vector of this polynomial is nothing but $\xi(f, D)$. $\square$

To show part (ii) we note that the polynomials $\{l_h(z)\}_{h \in D}$ do not depend on $\mathscr{F}$ and can be computed at any node in the network. So what we care to receive from the helper nodes are the multipliers $\{g^{(h)}(a_f)\}_{h \in D}$. Hence, for any set of helper nodes with $|A| < d-k+1$, it is gainful to send $\{g^{(h)}(a_f)\}_{h \in A}$ rather than the vector $\xi(f, A)$ since the former requires fewer than l transmissions. To conclude, we see that when $|A| \ge d-k+1$, we can transmit the vector $\xi(f, A)$ of dimension l, meeting the bound of Lemma II.2 and reproducing the result from [8].

Using multilinear algebra notation, explained in the next section, we can rephrase the code description as follows. The encoding is defined as a linear functional $\phi \in (F^2 \otimes S^2 F^{k-1})^*$, where $S^2 F^{k-1}$ is the second symmetric power (this is another way of saying that the encoding relies on evaluations of symmetric polynomials). Node i stores a restriction of ϕ to $x_i \otimes y_i \otimes F^{k-1}$, where $x_i = [1, a_i^{k-1}], y_i = [1, a_i, \dots, a_i^{k-2}]$. The contents of the failed node is a vector in the l-dimensional subspace $(x_f \otimes y_f \otimes F^{k-1})^*$, and the IP procedure recovers the coordinates of this vector in stages that correspond to moving along the repair graph toward the failed node. A general version of this idea underlies the repair procedure in the following sections.

B. Linear-algebraic notation

In this section we introduce elements of notation used below to define code families for which we design IP procedures of node repair.

For a linear space U over F we denote by U^* its dual space; its elements are linear functionals of the form $\phi : U \to F$. The spaces U and U^* have the same dimension and $(U^*)^* \cong U$. A *restriction* of ϕ to a subspace $V \subset U$ is denoted as $\phi \restriction V$.

Let U, V be linear spaces of dimensions m and n, respectively, and let us fix bases $\{\overline{u}_i\}_{i=1}^m$ and $\{\overline{v}_j\}_{j=1}^n$. The tensor product of U and V is a linear space $U \otimes V = \{\sum_{ij} a_{ij} \overline{u}_i \otimes \overline{v}_j, a_{ij} \in F\}$ where $a_{ij} \in F$ and the tensors $\overline{u}_i \otimes \overline{v}_j$ form a basis in $U \otimes V$ (thus $\dim(U \otimes V) = mn$). By definition, $u \otimes V = \{\sum_j a_j u \otimes \overline{v}_j, a_j \in F\}$ and $u \otimes V \subseteq U \otimes V$ The dual of a tensor product is the tensor product of duals, i.e., $(U \otimes V)^* = U^* \otimes V^*$. We denote by $T^p V := V \otimes V \otimes \dots \otimes V$ the pth tensor power of V. The dimension of $T^p V$ is n^p.

The *symmetric power* $S^p V$ is the linear space of symmetric tensors, i.e., elements of $T^p V$ invariant under transformations of the form $\overline{v}_1 \otimes \dots \otimes \overline{v}_p \mapsto \overline{v}_{\sigma(1)} \otimes \dots \otimes \overline{v}_{\sigma(p)}$ for any permutation σ. We write symmetric tensors as

$$\sum_{\substack{i_1, i_2, \cdots, i_p \\ 1 \le i_1 \le i_2 \le \cdots \le i_p \le n}} a_{i_1 i_2 \cdots i_p} \overline{v}_{i_1} \odot \overline{v}_{i_2} \odot \dots \odot \overline{v}_{i_p},$$

where $\odot$ denotes the symmetric product. By definition, $\dim(S^p V) = \binom{n+p-1}{p}$.

Finally, $x \wedge y$ denotes the exterior (alternating) product of vectors, characterized by $x \wedge y = -y \wedge x$; hence $\overline{v}_{\sigma(1)} \wedge \overline{v}_{\sigma(2)} \wedge \dots \wedge \overline{v}_{\sigma(n)} = \operatorname{sgn}(\sigma) \overline{v}_1 \wedge \overline{v}_2 \wedge \dots \wedge \overline{v}_n$, where $\operatorname{sgn}(\sigma)$ is the signature of the permutation σ. The *exterior power* $\Lambda^p V$ is a vector subspace of dimension $\binom{n}{p}$ spanned by elements of the form $\overline{v}_{i_1} \wedge \overline{v}_{i_2} \wedge \dots \wedge \overline{v}_{i_p}, 1 \le i_1 < i_2 < \dots < i_p \le n$.

C. Generalized PM Codes

An extension of the PM construction was recently proposed in [3]. The construction of [3, Sec.4] yields a family of MSR codes with parameters $n, k, d = \frac{(k-1)t}{t-1}, l = \binom{k-1}{t-1}, M = t\binom{k}{t}$ for any $t \ge 2$. In this section we follow the paradigm of evaluation codes to introduce an IP node repair procedure for this code family.

We start with a brief description of the code construction. Let $X = F^t$ and $Y = F^{k-t+1}$. Let $L := X \otimes S^t Y$ and note that $\dim(L) = M$. The encoding $\phi : L \to F^{nl}$ is an F-linear map. To define a concrete encoding procedure, we fix a basis in L^* and let the coordinates of ϕ be the contents of the stored data.

To support the data reconstruction and node repair tasks, we further choose, for each $i \in [n]$, a pair of vectors $x_i \in X$ and $y_i \in Y$ such that

(i) Any t-subset of x_i's spans X.

(ii) Any $(k-t+1)$-subset of y_i's spans Y.
(iii) Any d subspaces $x_i \otimes y_i \odot S^{t-2}Y$ span $X \otimes S^{t-1}Y$.

The first two properties enable data reconstruction, while the node repair property depends on the third condition [3].

With these assumptions, the contents of node i corresponds to the restriction $\phi \restriction x_i \otimes y_i \odot S^{t-1}Y \in (x_i \otimes y_i \odot S^{t-1}Y)^*$. This is consistent with the code parameters: indeed, an element in $(x_i \otimes y_i \odot S^{t-1}Y)^*$ is completely described by its evaluations on a basis of the space $x_i \otimes y_i \odot S^{t-1}Y$, which requires storing exactly $l = \binom{k-1}{t-1}$ evaluations.

As before, let $f \in [n]$ be (the index of) the failed node and let $D \subseteq [n] \setminus \{f\}$ be the helper set. Note that we wish to recover the restriction $\phi \restriction x_f \otimes y_f \odot S^{t-1}Y$. Choose a basis for $x_f \otimes y_f \odot S^{t-1}Y$ and let $x_f \otimes y_f \odot (\overline{y}_{i_1} \odot \cdots \odot \overline{y}_{i_{t-1}})$ be one of the basis vectors. Let $\{\underline{y}_{j_1} \odot \cdots \odot \underline{y}_{j_{t-2}}, 1 \le j_1 \le j_2 \cdots \le j_{t-2} \le n\}$ be a basis of $S^{t-2}Y$. Similarly to the PM codes, the helper node $i \in D$ transmits to the failed node the restriction of ϕ to the set of vectors $\{x_i \otimes y_i \odot (\underline{y}_{j_1} \odot \cdots \odot \underline{y}_{j_{t-2}}) \odot y_f\}$.

By (iii) above we can write

$$x_f \otimes y_f \odot (\overline{y}_{i_1} \odot \cdots \odot \overline{y}_{i_{t-1}}) = x_f \otimes (\overline{y}_{i_1} \odot \cdots \odot \overline{y}_{i_{t-1}}) \odot y_f$$
$$= \sum_{i \in D} \sum_{j_1, \cdots, j_{t-2}} a_{i, j_1 \cdots j_{t-2}} x_i \otimes y_i \odot \mathscr{Y}_{j_1, \ldots, j_{t-2}} \odot y_f.$$

where we denoted $\mathscr{Y}_{j_1,\ldots,j_{t-2}} = \underline{y}_{j_1} \odot \cdots \odot \underline{y}_{j_{t-2}}$. Again similarly to the PM codes, any set $A \subseteq D$ with $|A| \ge d-k+1$ can transmit the following single evaluation of ϕ along the path to f:

$$\phi\Big(\sum_{i \in A} \sum_{j_1, \cdots, j_{t-2}} a_{i, j_1 \cdots j_{t-2}} x_i \otimes y_i \odot \mathscr{Y}_{j_1, \ldots, j_{t-2}} \odot y_f\Big)$$
$$= \sum_{i \in A} \sum_{j_1, \cdots, j_{t-2}} a_{i, j_1 \cdots j_{t-2}} \phi(x_i \otimes y_i \odot \mathscr{Y}_{j_1, \ldots, j_{t-2}} \odot y_f).$$

This can be done for all basis vectors of the chosen basis of $x_f \otimes y_f \odot S^{t-1}Y$, and that requires $l = \binom{k-1}{t-1}$ transmissions which matches the lower bound of Lemma II.2. Note that the AF repair would require any set A of helpers to transmit $\beta|A|$ symbols of F, which is greater than l for $|A| > d-k+1$.

D. IP for Interior Point Codes

The intuition accumulated in the previous sections motivates the IP protocol for linear regenerating codes of the evaluation type. We exemplify this approach by using a family of non-MSR exact repair codes introduced recently in [2] (see also [9, Sec. 7.2]). Let s be an integer such that $n-1 \ge d \ge k \ge s-1 \ge 1$. The family of *moulin codes* that we discuss (the name was given by the authors of [2]) has parameters $[n,k,d,l,\beta,M]$ where

$$l = \textstyle\sum_{p+q=s-1} (d-k)^p \binom{k}{q}$$
$$\beta = \textstyle\sum_{p+q=s-2} (d-k)^p \binom{k-1}{q}$$
$$M = \textstyle\sum_{p+q=s-1} d(d-k)^p \binom{k}{q} - \sum_{p+q=s} (d-k)^p \binom{k}{q}. \quad (7)$$

Let $V = F^{d-k}$, $W = F^k$, and $U = V \oplus W \cong F^d$. We shall be dealing with spaces of the form $T^pV \otimes U \otimes \Lambda^q W$ where $p+q = s-1$ and $p,q \ge 0$. While the general idea is the same as before (node contents are given by restrictions of linear maps to subspaces), the detailed description relies on two operations on tensor products called *co-wedge multiplication* ∇ and *coboundary operators* ∂, defined formally in the appendix.

For a fixed s satisfying the constraints above, the file $\mathscr{F}$ is chosen to be an element ϕ of the dual space

$$\bigoplus_{p+q=s-1} (T^pV \otimes U \otimes \Lambda^q W)^* \quad (8)$$

which satisfies the parity checks that place constraints on ϕ that guarantee that the diagram

$$\begin{tikzcd}
T^pV \otimes \Lambda^{q+1}W \arrow[rr, "\phi"] \arrow[dr, "\nabla"'] & & F \\
& T^pV \otimes W \otimes \Lambda^q W \arrow[ur, "\phi"'] &
\end{tikzcd}$$

commutes for all $p \ge 1, q \ge 0$ with $p+q = s-1$ (plus the boundary cases of $p=0, q=-1$, not discussed for space constraints). The file size is the dimension of the direct sum of the vector spaces (8) minus the dimension of the parity check space, which is exactly (7). To each node $i \in [n]$ we associate a vector $u_i \in U$ such that any d of these vectors span U and any k vectors span U/V under the quotient map $U \to U/V$. The i-th node stores the following restriction of the mapping ϕ:

$$\phi \restriction \bigoplus_{p+q=s-1} T^pV \otimes u_i \otimes \Lambda^q W.$$

The size l of the node equals $\dim(T^pV \otimes u_i \otimes \Lambda^q W)$.

Now suppose that node $f \in [n]$ fails and we are provided with a set $D \subseteq [n] \setminus \{f\}$ of d helpers. Each helper $h \in D$ provides the restrictions of its contents to coboundaries:

$$\phi \restriction \partial^U_{u_f}(T^pV \otimes u_h \otimes \Lambda^q W) \quad (9)$$

for each pair p,q with $p+q = s-2$. We shall need the following result.

Lemma III.2 ([2], Thm. 4.1)**.** *For all possible $p,q \ge 0$, such that $p+q = s-1$, for all $\nu \in T^pV, \omega \in \Lambda^q W$, we have*

$$\phi(\partial^U_{u_f}(\nabla(\nu \otimes \omega))) - \phi(\partial^U_{u_f}(\nu \otimes \omega)) = (-1)^p \phi(\nu \otimes u_f \otimes \omega).$$

The right-hand side of the above equation is one coordinate of the failed node, and the left-hand side can be computed from (9). The statement of next lemma appears in [2] without a proof. We include it here to set up the notation.

Lemma III.3. *1) For all possible $p \ge 1, q \ge 0$, such that $p+q = s-1$, for all $\nu \in T^pV, \omega \in \Lambda^q W$, the tensor $\nu \otimes \omega$ is contained in the linear span of the union of the spaces $\{T^pV \otimes u_h \otimes \Lambda^q W\}_{h \in D, p+q=s-2}$.*
2) For all possible $p,q \ge 0$, such that $p+q = s-1$, for all $\nu \in T^pV, \omega \in \Lambda^q W$, $\nabla(\nu \otimes \omega)$ is contained in the linear span of the union of the spaces $\{T^pV \otimes u_h \otimes \Lambda^q W\}_{h \in D, p+q=s-2}$.

Proof: 1) Fix $p_1 \ge 1, q_1 > 0$ such that $p_1 + q_1 = s-1$. Let $\nu \in T^{p_1}V$ and $\omega \in \Lambda^{q_1}W$. Fix a basis $\{\overline{\nu}_i \otimes u_h \otimes \overline{\omega}_i\}_{i=1}^{(d-k)^{p_1-1}\binom{k}{q_1}}$ of $T^{p_1-1}V \otimes u_h \otimes \Lambda^q W$. Since the set $\{u_h\}_{h \in D}$ spans U, we can write

$$\nu \otimes \omega = \overbrace{(\nu_1 \otimes \cdots \otimes \nu_{p_1-1})}^{T^{p_1-1}V} \otimes \underbrace{\nu_{p_1}}_{U} \otimes \overbrace{\omega}^{\Lambda^q W}$$

and hence $\nu \otimes \omega$ is an element of $T^{p_1-1}V \otimes U \otimes \Lambda^{q_1}W$. So we can write $\nu \otimes \omega = \sum_{i,h} a_{i,h}(\overline{\nu}_i \otimes u_h \otimes \overline{\omega}_i)$.

2) Similarly, for a basis $\{\underline{\nu}_j \otimes u_h \otimes \underline{\omega}_j\}_{j=1}^{(d-k)^{p_1}\binom{k}{q_1-1}}$ of $T^{p_1}V \otimes u_h \otimes \Lambda^{q_1-1}W$, we can write $\nabla(\nu \otimes \omega) = \sum_{j,h} b_{j,h}(\underline{\nu}_j \otimes u_h \otimes \underline{\omega}_j)$.

Lemma III.4. *Let $A \subseteq D$. The nodes in A need to transmit only l symbols for the repair of f.*

Proof. Fix $p, q \geq 0$ such that $p+q = s-1$. Let $\nu \in T^pV, \omega \in \Lambda^qW$. If $p \geq 1$,

$$\phi(\partial^U_{u_f}(\nabla(\nu\otimes\omega)))-\phi(\partial^U_{u_f}(\nu\otimes\omega)) = \sum_{j,h} b_{j,h}\phi(\partial^U_{u_f}(\underline{\nu}_j\otimes u_h\otimes\underline{\omega}_j)) - \sum_{i,h} a_{i,h}\phi(\partial^U_{u_f}((\overline{\nu}_i \otimes u_h \otimes \overline{\omega}_i)).$$

If $p = 0$ then the second term on the LHS is already 0. If $p > 0$, then by Lemma III.2, the LHS is $(-1)^p\phi(\nu \otimes u_f \otimes \omega)$, and we have recovered one symbol of the failed node. For this, the set A need to transmit the element

$$\sum_{h\in A}\Big[\sum_j b_{j,h}\phi(\partial^U_{u_f}(\underline{\nu}_j \otimes u_h \otimes \underline{\omega}_j)) - \sum_i a_{i,h}\phi(\partial^U_{u_f}((\overline{\nu}_i \otimes u_h \otimes \overline{\omega}_i))\Big].$$

Doing this for any fixed $\{\nu, \omega\}$ basis of $T^pV \otimes u_f \otimes \Lambda^qW$, for all values of p, q, requires the set A to transmit a total of l symbols. $\square$

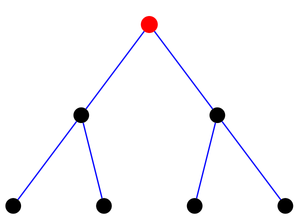
Observe that whenever $|A| \geq \lceil \frac{l}{\beta} \rceil$, the IP protocol given by this lemma results in communication savings compared to the AF repair.

IV. Partial Node Repair

In this section we briefly consider the node repair problem for the situation when only a part, say $\gamma l, 0 \leq \gamma \leq 1$, of node's symbols are erased. One of the first works devoted to this question was [6] which derived a bound on the file size of the form (1) that accounts for the parameter γ.

This problem gives rise to a number of open questions, starting with MSR code constructions, that have not been addressed in the literature. Without attempting a comprehensive analysis, we point out that partial repair can be implemented under the IP approach discussed here. The underlying idea is that the helper nodes need to transmit only the linear combinations corresponding to the failed coordinates.

Example IV.1. *Consider the $[n = 7, k = 4, d = 6, l = 3, \beta = 1, M = 12]$ PM MSR code, placed on the following graph.*



Suppose that the root node is erased, and the remaining 6 nodes form the helper set. If the entire contents of the root node is lost, both the AF strategy and the IP strategy of Lemma III-A require a total transmission of 10 symbols. At the same time, if only the first coordinate of the root node needs to be recovered, then the two immediate neighbors of the root node can transmit just the first row of equation (6), and 6 transmissions suffice.

Let T_f be a rooted tree of $G_{f,D}$ rooted at f. Let $D(v)$ be the descendants of node $v \in V_{f,D}$ and let $D^*(v) = D(v) \cup \{v\}$. We have the following lemma that generalizes our earlier result (Theorem III.1 in [8]).

Lemma IV.1. *Given an $[n, k, d, l, \beta, F]$ linear regenerating code. There exists a repair procedure that recovers a $\gamma \in (0, 1)$ fraction of the failed node using the repair bandwidth*

$$\sum_{v\in D} \min\{\gamma l, |D^*(v)|\beta\}.$$

This result entails savings in communication when the erased fraction of the node contents γ is small, namely $\gamma l < \beta$.

Appendix

Here we define operations on tensor product spaces used in Sec. III-D. The notation below was introduced in Sec. III-D and III-B. Below $p, q > 0$.

Co-wedge multiplication:

$$\begin{aligned}
\nabla : T^pV \otimes \Lambda^1W &\to T^pV \otimes W \\
\nu \otimes w_1 &\to \nu \otimes w_1 \\
\nabla : T^pV \otimes \Lambda^2W &\to T^pV \otimes W \otimes \Lambda^1W \\
\nu \otimes w_1 \wedge w_2 &\to \nu \otimes w_1 \otimes w_2 - \nu \otimes w_2 \otimes w_1 \\
\nabla : T^pV \otimes \Lambda^{q+1}W &\to T^pV \otimes W \otimes \Lambda^qW \\
\nu \otimes \omega \wedge w_1 &\to \nabla(\nu \otimes \omega) \wedge w_1 + (-1)^q\nu \otimes w_1 \otimes \omega.
\end{aligned}$$

Coboundary operators (differentials). For any $v \in V$:

$$\begin{aligned}
\partial^V_v : \Lambda^qW &\to U \otimes \Lambda^qW \\
\omega &\to 0 \\
\partial^V_v : U \otimes \Lambda^qW &\to T^1V \otimes U \otimes \Lambda^qW \\
u \otimes \omega &\to v \otimes u \otimes \omega \\
\partial^V_v : T^pV \otimes U \otimes \Lambda^qW &\to T^{p+1}V \otimes U \otimes \Lambda^qW \\
\nu \otimes u \otimes \omega &\to \partial^V_v(\nu) \otimes u \otimes \omega + (-1)^p\nu \otimes v \otimes u \otimes \omega
\end{aligned}$$

and for every $w \in W$:

$$\begin{aligned}
\partial^W_w : T^pV \otimes U &\to T^pV \otimes U \otimes \Lambda^1W \\
\nu \otimes u &\to (-1)^p\nu \otimes u \otimes w \\
\partial^W_w : T^pV \otimes U \otimes \Lambda^qW &\to T^pV \otimes U \otimes \Lambda^{q+1}W \\
\nu \otimes u \otimes \omega &\to (-1)^{p+q}\nu \otimes u \otimes \omega \wedge w.
\end{aligned}$$

Finally for $u \in U$ such that $u = v + w, v \in V, w \in W$, define

$$\partial^U_u = \partial^V_v + \partial^W_w.$$

Acknowledgment: We are grateful to a reviewer of our paper [8] for suggesting to study the repair problem for non-MSR codes on graphs.

REFERENCES

[1] A. G. Dimakis, P. B. Godfrey, Y. Wu, M. J. Wainwright, and K. Ramchandran, "Network coding for distributed storage systems," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4539–4551, 2010.

[2] I. M. Duursma, X. Li, and H.-P. Wang, "Multilinear algebra for distributed storage," *SIAM J. Appl. Algebra Geom.*, vol. 5, pp. 552–587, 2021.

[3] I. M. Duursma and H.-P. Wang, "Multilinear algebra for minimum storage regenerating codes: a generalization of the product-matrix construction," *Applicable Algebra in Engineering, Communication and Computing*, 2021. [Online]. Available: https://doi.org/10.1007/s00200-021-00526-3

[4] M. Elyasi and S. Mohajer, "Cascade codes for distributed storage systems," *IEEE Trans. Inf. Theory*, vol. 66, no. 12, pp. 7490–7527, 2020.

[5] M. Elyasi and S. Mohajer, "Determinant coding: A novel framework for exact-repair regenerating codes," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 6683–6697, 2016.

[6] M. Gerami, M. Xiao, and M. Skoglund, "Two-layer coding in distributed storage systems with partial node failure/repair," *IEEE Communications Letters*, vol. 21, no. 4, pp. 726–729, 2017.

[7] S. Mohajer and R. Tandon, "New bounds on the (n, k, d) storage systems with exact repair," in *2015 IEEE International Symposium on Information Theory (ISIT)*, 2015, pp. 2056–2060.

[8] A. Patra and A. Barg, "Node repair on connected graphs," *IEEE Trans. Inf. Theory*, 2022, to appear. Preprint: arXiv:2108.00939.

[9] V. Ramkumar, S. Balaji, B. Sasidharan, M. Vajha, M. N. Krishnan, and P. V. Kumar, "Codes for distributed storage," *Foundations and Trends in Communication and Information Theory*, 2022, to appear.

[10] K. V. Rashmi, N. B. Shah, and P. V. Kumar, "Optimal exact-regenerating codes for distributed storage at the MSR and MBR points via a product-matrix construction," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 5227–5239, 2011.

[11] B. Sasidharan, K. Senthoor, and P. V. Kumar, "An improved outer bound on the storage-repair-bandwidth tradeoff of exact-repair regenerating codes," in *2014 IEEE International Symposium on Information Theory (ISIT)*, 2014, pp. 2430–2434.

[12] K. Senthoor, B. Sasidharan, and P. V. Kumar, "Improved layered regenerating codes characterizing the exact-repair storage-repair bandwidth tradeoff for certain parameter sets," in *2015 IEEE Information Theory Workshop (ITW)*, 2015, pp. 1–5.

[13] N. B. Shah, K. V. Rashmi, P. V. Kumar, and K. Ramchandran, "Distributed storage codes with repair-by-transfer and nonachievability of interior points on the storage-bandwidth tradeoff," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1837–1852, 2012.

[14] C. Tian, "Characterizing the rate region of the $(4, 3, 3)$ exact-repair regenerating codes," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 5, pp. 967–975, 2014.