Exact Repair Problems with Multiple Sources

Jayant Apte, Congduan Li, John MacLaren Walsh, Steven Weber
Drexel University, Dept. of ECE, Philadelphia, PA 19104, USA
jsa46@drexel.edu,congduan.li@drexel.edu, jwalsh@coe.drexel.edu, sweber@coe.drexel.edu

***Abstract*—We consider a new variant of the exact repair distributed storage problem, the multi-source exact repair problem, wherein the reconstruction decoders are each only required to provide a subset of the source variables. To best illustrate the idea, we generalize the $(n, k, d) = (3, 2, 2)$ exact repair distributed storage problem to the multisource case. When every decoder demands all source variables, the rate region of the $(3, 2, 2)$ exact repair problem is known to be same as that of the $(3, 2, 2)$ functional repair problem, while the rate region for $(3, 2, 2)$ case with multiple sources is unknown. We find achievable rate regions for vector binary and scalar binary codes via an automated approach.**

***Index Terms*—Distributed storage, exact repair, multi-source network coding**

I. Introduction

The repair problem for distributed storage systems has received considerable attention recently. Distributed storage systems use erasure codes for protection against disk failure. To recover the system after failure, the surviving nodes send information to a new node. The disks or arrays of disks are connected by a network and might be separated by large distances. The amount of information sent to the new node during the repair process has emerged as parameter of interest along with its relationship with amount of information stored per node.

In their pioneering work Dimakis et al.[3] considered a version of repair problem called *functional repair* problem. In this version, while the system would repair the failed node, the overall erasure code was not guaranteed to remain the same over time. This case was reduced to single source multicast network coding problem and a wide array of polynomial code construction techniques including deterministic [8] and randomized [6] algorithms could be directly used for the functional repair problem. However, functional repair has disadvantages including high overhead and the lack of systematic codes. To avoid these difficulties, a new version of the problem was introduced called *exact repair* problem [16], [15], [14], [1] where the regenerated data in the new node is constrained to be same as that on the failed node. This way, the system continues to maintain the same erasure code in spite of the failure of nodes and the regeneration of data. Tian[17] proved that, in general, rate region of the exact repair problem can be smaller than that of the functional repair case. This paper seeks to extend the exact repair paradigm to include the possibility of multiple sources with decoders requiring only subsets of sources.

The paper is organized as follows: in §II we introduce the multi-source exact repair problem. We generalize the framework in [15], [14] and more recently [17] for exact repair regenerating codes to include the possibility of multiple source variables and that different decoders are required to produce different subsets of the source variables, and provide some motivation for why such a generalization may be interesting in applications. In §III we review the relationships between matroids and scalar codes and between subspace arrangements and vector codes, upon which our computer assisted proof technique for achievability is based. §IV details the computer assisted achievability proof technique and presents the main results of this paper: the achievable rate regions for multi-source (3,2,2) exact repair problems with 1) 1-bit variables and various decoder demands and 2) 1 2-bit variable and rest 1-bit variables with various decoder demands. We discuss possible algorithms for computing rate regions and introduce a dual projection algorithm for computing inner bounds on rate regions from matroids and subspace arrangements that is used for computing the aforementioned achievable rate regions.

II. Problem Definition

In a $(n, k, d) = (3, 2, 2)$ repair problem we have $n = 3$ storage nodes and $\binom{n}{k} = 3$ decoders with each decoder having access to only a 2-subset ($k = 2$) of storage nodes. When any of the three storage nodes fails, both ($d = 2$) of the remaining nodes participate in the repair process. By a multisource exact repair problem, we will additionally mean that instead of requiring every one of the $\binom{n}{k} = 3$ decoders to reconstruct the source, different subsets of decoders will be required to decode different subsets of a collection of sources. This is inspired by the fact that different users of a distributed storage pattern will have different localities in which they will typically access their data, so it may be desirable to have asymmetry regarding which subsets of disks can reconstruct which sources. To provide the simplest case of such a model, we will consider the problem depicted in Fig. 1, where there are two sources $\mathcal{S} = (S_1, S_2)$, and one pair of storage nodes is capable of decoding the first source, another pair is capable of reconstructing the second source, and the third pair is capable of reconstructing both of them.

Definition 1. An $(\mathcal{S}, N, K_d, K, \{\beta(j) \subseteq \mathcal{S}\}_{j \in I_3})$ exact repair regenerating code for a (3,2,2) case consists of 3 encoding functions $f_i^E(., \ldots, .)$, 3 decoding functions $f_A^D(., ., .)$ 6 repair encoding functions and 3 repair decoding functions $F_i^D(., ., .)$, where the encoders have access to all of the sources

$$f_i^E : I_N{}^s \mapsto I_{K_d}, \quad i \in I_3, s = |\mathcal{S}| \tag{1}$$

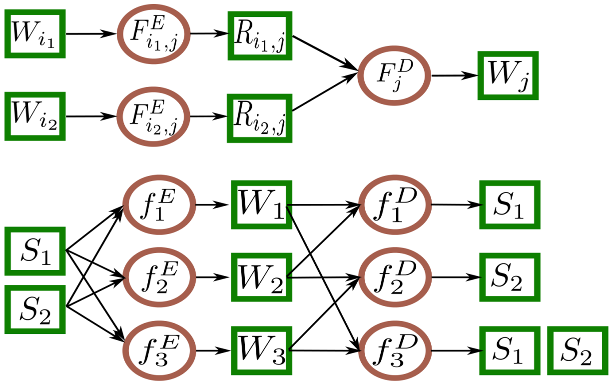

Figure 1. The 2-source (3,2,2) exact repair problem with source messages $\mathcal{S} = \{m_1, m_2\}$ decoder demands $\beta(1) = \{m_1\}, \beta(2) = \{m_2\}$ and$\beta(3) = \{m_1, m_3\}$. S_1, S_2 are the source variables, $W_i, i \in I_3$ are storage variables and $R_{i,j}, i \in I_3, j \in I_3$ are the repair variables. Also, $i_1, i_2 \in I_3$. Decoders 1,2,3 have access to storage variable sets $A_1 = \{1, 2\}, A_1 = \{2, 3\}$ and $A_3 = \{1, 3\}$ respectively

and the decoders

$$f^D_{A_j} : {I_{K_d}}^2 \mapsto {I_N}^{|\beta(j)|} \quad A_j \subset I_3, |A| = 2, j \in I_3 \tag{2}$$

each map 2 pieces of coded information stored on a set $A_j, |A_j| = 2$ of nodes to the subset $\beta(j)$ of original messages, and repair message encoders

$$F^E_{i,j} : I_{K_d} \mapsto I_K, \quad j \in I_3 \text{ and } i \in I_3 \setminus j \tag{3}$$

each of which maps a piece of coded information at node i to a message that will be used to reconstruct data at node j, and repair decoders

$$F^D_j : {I_K}^2 \mapsto I_{K_d} \quad j \in I_3 \tag{4}$$

each of which uses the information sent by the helper nodes to reconstruct information stored at the failed node.

The above functions must satisfy the following data reconstruction condition

$$f^D_{A_i}\left(\prod_{j \in A_i} f^E_j(\mathcal{S})\right) = \beta(i), \quad \beta(i) \in {I_N}^{|\beta(i)|}, A_i \subset I_3, j \in I_3 \tag{5}$$

and the following repair conditions

$$F^D_j\left(\prod_{i \in I_3 \setminus \{j\}} F^E_{i,j}(f^E_i(\mathcal{S}))\right) = f^E_j(S), j \in I_3. \tag{6}$$

As noted in [17], the exact repair problem reduces to multisource multicast network coding (MSNC) problem [19]. An implicit characterization of the rate region of MSNC problem in terms of the closure of the region of entropic vectors $\overline{\Gamma^*_N}$ is provided in [19]. While characterization of $\overline{\Gamma^*_N}$ remains an open problem, polyhedral inner and outer bounds respectively on $\overline{\Gamma^*_N}$ can be utilized to obtain polyhedral inner and outer bounds on the rate region.

The $(3, 2, 2)$ multisource exact repair problem under consideration in this paper is depicted in Fig. 1, where we have labeled the random variables involved with the labels in the green boxes. In particular, we have source variables $\mathcal{S} = \{S_i | i \in I_2\}$, encoded storage variables $\mathcal{W} = \{W_i |, i \in I_3\}$, and repair message variables $\mathcal{P} = \{R_{i,j} | i, j \in I_3, i \neq j\}$ as shown in Fig. 1. Since there are 11 random variables, the rate region of this network can be implicitly written in terms of $\overline{\Gamma^*_{11}}$. The problem structure defines following constrained regions in $\mathcal{H}_N := \mathbb{R}^{2^{11}+1}$, where the first $2^{11} - 1$ coordinates are identified with different non-empty subsets of the 11 random variables in the usual way encountered when working with $\bar{\Gamma}^*_{11}$, and the last 2 coordinates are variables α and β related to the storage and repair bandwidths, respectively

$$\begin{aligned}
&\mathcal{L}_1 = \{\mathbf{h} \in \mathcal{H}_N : h_{\mathcal{S}} = h_{S_1} + h_{S_2}\} \\
&\mathcal{L}_2 = \left\{\mathbf{h} \in \mathcal{H}_N : h_{W_i | S_1, S_2} = 0, \;\; i \in I_3\right\} \\
&\mathcal{L}_3 = \left\{\mathbf{h} \in \mathcal{H}_N : h_{W_j | R_{k,j}, R_{\ell,j}} = 0, \;\; \{k, j, \ell\} = I_3\right\} \\
&\mathcal{L}_4 = \{\mathbf{h} \in \mathcal{H}_N : h_{W_i} \leq \alpha, \;\; i \in I_3\} \\
&\mathcal{L}_5 = \left\{\mathbf{h} \in \mathcal{H}_N : h_{R_{i,j}} \leq \beta, \;\; i, j \in I_3\right\} \\
&\mathcal{L}_6 = \left\{\mathbf{h} \in \mathcal{H}_N : h_{\beta(i) | W_{A_i}} = 0\right\} \\
&\mathcal{L}_* = \bigcap_{i \in \{1,2,3,4,5,6\}} \mathcal{L}_i
\end{aligned} \tag{7}$$

A minor extension of [19] shows that given a polyhedral inner(outer) bound Γ^{poly}_{11} on $\overline{\Gamma^*_{11}}$ one can compute a polyhedral inner(outer) bound on the rate region as:

$$\mathcal{R}^{in} = Ex(proj_{S_1, S_2, \alpha, \beta}(\Gamma^{poly}_{11} \cap \mathcal{L}_*)) \tag{8}$$

where $Ex(\mathcal{B}) = \{\mathbf{x} \in \mathbb{R}^4 : \mathbf{h} = \mathbf{x} \geq \mathbf{x}' \text{ for some } \mathbf{x}' \in \mathcal{B}\}$ and $proj_{S_1, S_2, \alpha, \beta}(\mathcal{G}) = \{(h'_{S_1}, h'_{S_2}, \alpha', \beta')^T \in \mathbb{R}^4 : \exists \mathbf{h}' \in \mathcal{G} \text{ with } h_{S_1} = h'_{S_1}, h_{S_2} = h'_{S_2}, \alpha = \alpha' \text{ and } \beta = \beta'\}$.

III. Codes and their abstract counterparts

In this section we review the relationships between matroids and scalar codes and between subspace arrangements and vector codes. We use these relationships to get inner bounds on $\overline{\Gamma^*_{11}}$ that will be used in the next section to compute achievable inner bounds on the rate region of the network.

A. Inner bounds on $\overline{\Gamma^*_N}$ from matroids

Definition 2. A *matroid* [13] on a ground set $\mathcal{S}$ of size $|\mathcal{S}| = N$ can be defined via its *rank function*, which is a function $r : 2^{\mathcal{S}} \to \{0, \ldots, N\}$ obeying for all $A, B \subseteq \mathcal{S}$:

1) Cardinality: $r(A) \leq |A|$;
2) Submodularity: $r(A \cup B) + r(A \cap B) \leq r(A) + r(B)$.
3) Monotonicity: if $A \subseteq B \subseteq \mathcal{S}$ then $r(A) \leq r(B)$.

Given a matroid rank function, we can stack ranks of different nonempty subsets into a vector $\mathbf{r} \in \mathbb{R}^{2^{|S|}-1}$. We denote the conic hull of all vectors obtained from ground set size N matroids as Γ^{mat}_N. Representable matroids are the class of matroids whose rank functions are in fact ranks of subsets of columns of a matrix. In particular, a matroid M with ground set $\mathcal{S}$ of size $|\mathcal{S}| = N$ and rank $r(\mathcal{S}) = k$ is representable over the finite field $\mathbb{F}_q$ of size q if there exists a matrix $\mathbf{A} \in \mathbb{F}_q^{k \times N}$ such that $\forall B \subseteq \mathcal{S}$ $r(B) = \text{rank}(\mathbf{A}_{:,B})$, the matrix rank of the columns of $\mathbf{A}$ indexed by B. Denote by Γ_N^q the conic hull of rank vectors obtained from matroids representable over $\mathbb{F}_q$. As reviewed in detail in [9], $\Gamma_N^q \subset \overline{\Gamma_N *}$. For this work, we will focus on binary matroids, i.e. matroids representable over $\mathbb{F}_2$, which have the following forbidden minor characterization due to Tutte.

Theorem 3. *[13](Tutte) A matroid is representable over $\mathbb{F}_2$ if and only if it has no $U_{2,4}$ minor.*

Denote by Γ_N^{bin} the conic hull of rank vectors obtained from matroids representable over $\mathbb{F}_2$. Partially owing to the large amount of symmetry, and partially to simple combinatorial explosion, the number of matroids grows rapidly with size of ground set. One can remove a part of the explosion due to the symmetries by working with lists of *non-isomorphic* matroid rank functions, which have long been available for $N \leq 8$ and have recently become available for $N = 9$ [11] and partially (i.e. only the matroids of specific ranks) for $N = 10$ [10]. It is possible to adapt the work in [11] and [10] to directly list only nonisomorphic binary matroids. Using this adaptation, the authors have listed all non-isomophic binary matroids upto the ground set size $N = 12$ and some of the ranks for ground set size $N = 13$. The problem of counting the number of binary, ternary and in general $\mathbb{F}_q$-representable matroids was studied by Wild [18] who counted the number of orbits of the set of all $r \times n$ matrix representations over $\mathbb{F}_q$ under action of group that produces isomorphic representations using Burnside lemma, but did not provide a method for listing the non-isomorphic matroids that were counted. Sizes of lists of binary matroids obtained via authors' adaptation of [11] and [10] to the exclusively binary case agree with Wild's counting results. Unfortunately, the representable matroidal inner bounds Γ_N^q are poor inner bounds to $\bar{\Gamma}_N^*$. This is because for $N \geq 4$ the majority of extreme rays of Γ_N are integral polymatroids (meaning that they obey all of the requirements of matroids except cardinality) but not matroids. Hence, far superior inner bounds are obtained by projecting representable matroids to get linear polymatroids, also known as subspace dimensions, as we describe presently.

B. Inner bounds on $\overline{\Gamma_N^*}$ from subspaces

Consider a collection of N vector subspaces $\mathcal{V} = (V_1, \ldots, V_N)$ of a finite dimensional vector space, and define the set function $\mathrm{d} : 2^{\mathcal{V}} \to \mathbb{N}_+$, where $\mathrm{d}(A) = \dim\left(\sum_{i \in A} V_i\right)$ for each $A \subseteq [N]$ is the dimension of the vector space generated by the sum of subspaces indexed by A. For any collection of subspaces $\mathcal{V}$, the function d is integer valued, and obeys monotonicity and submodularity. Additionally, for every subspace dimension function d, we can stack the dimension d of different non-empty subsets of subspaces into a vector $\mathbf{d} \in \mathbb{R}^{2^N - 1}$. Denote the conic hull of all such vectors $\mathbf{d}$ obtained from arrangements of N finite dimensional subspaces by $\Gamma_N^{\mathrm{space}}$. It is known that Γ_N^{space} forms an inner bound on $\overline{\Gamma_N^*}$[5].

Integrality, monotonicity, and submodularity are necessary but insufficient for for a given set function $\mathrm{d} : 2^{\mathcal{V}} \to \mathbb{N}_+$ to be dimension function of subspace arrangements. That is, there exist additional inequalities that are necessary to describe the conic hull of all possible subspace dimension set functions. As discussed in [5], Ingleton's inequality [7] together with the Shannon outer bound Γ_4, completely characterizes $\Gamma_4^{\mathrm{space}}$.

Although, for $N > 5$, Γ_N^{space} is not known, we can form strict inner bounds to $\Gamma_N^{\mathrm{space}}$ by projecting representable matroids on a larger groundset. We shall denote inner bounds created this way as $\Gamma_{N,k}^{\subset space}$. Consider a representable matroid M with groundset $E(M)$ with cardinality $|E(M)| = k$ and rank function r. We can obtain dimension function d of subspace arrangements of $t \leq k$ subspaces from r as follows:

- For a given $t \leq k$ define a t-partition of $E(M)$ and denote it as $\mathcal{P}$. Let A be the set indexing the different sets in the partition $\mathcal{P}$
- With each $P \in \mathcal{P}$ associate a subspace V_P spanned by the associated vectors in the representable matroid index by the ground set elements P.
- Now define the dimension function of subspace arrangement to be $d(B) = r(\cup_{i \in B} P_i)$ for each $B \subseteq A$.

It is easy to see that dimension vector $\mathbf{d}$ of subspace arrangement created in such a way is a linear projection of rank vector r of the original representable matroid. Similarly, it is easy to see that given any collection of subspaces, one can create a matrix by stacking together their bases, and this is by construction a representation for a representable matroid which when projected gives the proscribed subspace dimensions. From this logic it is evident that $\Gamma_N^{space} = \cup_{k=1}^{\infty} \Gamma_{N,k}^{\subset space}$.

From previous subsection, we know that lists of non-isomorphic representable matroids of size k are available for field size $q = 2$. We denote the conic hull of the set of dimension vectors of subspace arrangements obtained by projecting binary representable matroids by $\Gamma_{N,k,2}^{\subset space}$.

C. Matroids, subspace arrangements and network codes for multisource exact repair problem

An important feature of inner bounds on $\overline{\Gamma_N^*}$ obtained from representable matroids and subspace arrangements is that they not only provide us with an inner bound on the rate region but also a linear code for every extreme point in the inner bound. Relationship between network codes and corresponding achievable points in $\mathcal{R}_{in}$ can be specified via a network to $\mathbb{F}_q$ matroid mapping. Let V be the set of all random variables in Fig. 1 and $E(M)$ be the ground set of a matroid M such that $|\mathcal{S}| \geq 11$. Consider a mapping $f : V \mapsto 2^{E(M)} \setminus \phi$ such that $f(S_1), f(S_2), f(W_i), i \in I_3$ and $f(R_{i,j}, i, j \in I_3)$ form a partition of the set $E(M)$ and it satisfies:

- f is one-to-one
- $r(f(S_1) \cup f(S_2)) = r(f(S_1)) + r(f(S_2)) (\mathcal{L}_1)$
- $r(f(S_1) \cup f(S_2) \cup f(W_i)) = r(f(S_1) \cup f(S_2)$ $(\mathcal{L}_2)$
- $r(\cup_{j \in A_i} r(f(W_j)) = r(\{f(W_j, j \in A_i\} \cup \{\cup_{j \in \beta(i)} W_j, \}), \ \forall i \in I_3$ $(\mathcal{L}_3)$

The matrix representation of the matroid M is a valid *basic solution* for the problem. If $E(M) = 11$, it is called the *basic scalar network code* for the problem. Otherwise, if

$|E(M)| > 11$, it is called *basic vector network code* which also corresponds to a subspace-arrangement containing 11 subspaces. An arbitrary rational point in the rate region can be obtained via time sharing together with associated basic scalar or vector network codes according to a construction detailed in [9]. We summarize this fact and the results in the previous sections as the following lemma:

Lemma 4. *$\mathcal{R}_{in}$ obtained from Γ_N^q and $\Gamma_{N,k,q}^{\subset space}$ is achievable*

IV. Computer Assisted Achievability Proof

Having established the achievability of rate regions obtained from Γ_N^q and $\Gamma_{N,k,q}^{\subset space}$, we turn our attention to actually computing $\mathcal{R}_{in}$ via polyhedral computation techniques. Note that there are two ways of representing every polyhedron: either as intersection of finite number of half-spaces(H-representation) or as conic hull of finite number of extreme rays(V-representation). The following table gives algorithms used to obtain representations of polyhedra that are then used to compute $\mathcal{R}_{in}$: e.g. from non-isomorphic rank vectors available from matroid enumeration (more precisely, matroid *listing*) mentioned in subsection III-A one can obtain the extreme ray representtion of Γ_N^q by first forming the matroid isomorphs under all permutations of ground set and then performing redundancy removal(i.e. remove all rays that can be represented as conic combination of others). When dealing with the matroidal inner bounds, the matroids that are connected are extremal, and hence the redundancy removal can be replaced with a connectedness check [2]. The extreme ray representation of $\Gamma_{N,k,q}^{\subset space}$ can be obtained via linear projection of extreme rays of Γ_k^q followed by redundancy removal to form the conic hull. In general, given the extreme ray representation of Γ_N^{poly}(an arbitrary polyhedral inner bound on $\overline{\Gamma_N^*}$), computation of extreme ray representation of $\Gamma_N^{poly} \cap \mathcal{L}_i$ for any $i \in I_6$ can be interpreted as single iteration Double Description method of representation conversion[9], [4], [12] of polyhedra. Further, computing $\mathcal{V}_R = \Gamma_N^{poly} \cap \mathcal{L}_i, i \in \{1, 2, 3, 6\}$ can be done in $\mathcal{O}(n)$ where n is number of extreme rays of Γ_N^{poly} with the resultant intersection having $\mathcal{O}(n)$ extreme rays[9]. On the other hand, computing extreme rays of $\Gamma_N^{poly} \cap \mathcal{L}_i, i \in \{4, 5\}$ amounts to general double description iteration and is computationally burdensome. Hence, we first compute the extreme rays $\mathcal{V}_R$ as shown above and then use a dual projection approach as described in next subsection to avoid costly iterations of double description method corresponding to $\mathcal{L}_4$ and $\mathcal{L}_5$.

A. Use of duality for projection

Let's consider the pointed polyhedral cone given by a subspace inner bound to the region of entropic vectors obeying all network constraints except those involving α and β. For instance, consider

$$\mathcal{V}_R = \Gamma_{11,12,2}^{\subset space} \cap \mathcal{L}_{1236} \tag{9}$$

whose extreme ray representation can be calculated with the methods described in the previous section. The rate region we wish to calculate can be expressed as

$$\mathcal{R}_{in} = Ex(proj_{S_1,S_2,\alpha,\beta}(proj_{\mathcal{S},\mathcal{W},\mathcal{P},\alpha,\beta}(\Gamma_{11}^{poly} \cap \mathcal{L}_{123456})) \tag{10}$$

Since

$$\Gamma_{11,12,2}^{\subset space} \cap \mathcal{L}_{123456} = \mathcal{V}_R \cap \mathcal{L}_{45} \tag{11}$$

we can write

$$\mathcal{R}_{in} = Ex(proj_{S_1,S_2,\alpha,\beta}(\mathcal{V}_R \cap \mathcal{L}_{45})) \tag{12}$$

the intersection $\mathcal{V}_R \cap \mathcal{L}_{45}$ is an intersection of a cone for which the extreme ray representation is easily calculated, and a series of linear inequalities. One way to compute this intersection is to perform representation conversion on $\mathcal{V}_R$ to obtain $\mathcal{H}_R$, the halfspace representation of same polyhedral set, then add the inequalities to the list. However this representation conversion of $\mathcal{V}_R$ can be time consuming, even for polyhedra in $\mathbb{R}^{11}$, and hence it is preferable to provide a direct method calculating the tradeoff between the entropies and α and β.

In order to create this method, let $R = \{\mathbf{r}_1, \ldots, \mathbf{r}_m\}$ be the extreme ray representation of $\mathcal{V}_R$, and define the polyhedron

$$\mathcal{C} = \left\{ (h_{S_1}, h_{S_2}, \alpha, \beta, \boldsymbol{\lambda}) \left| \begin{array}{c} \sum_{i=1}^m \lambda_i r_{S_j} = h_{S_j}, \ , \forall j \in I_2 \\ \sum_{I=1}^m \lambda_i r_{W_j} \leq \alpha, \forall j \in I_3 \\ \sum_{i=1}^m \lambda_i r_{R_{j,k}} \leq \beta \forall j,k \in I_3 \\ \boldsymbol{\lambda} \geq \mathbf{0} \end{array} \right. \right\} \tag{13}$$

When written this way, $\mathcal{C}$ is defined in terms of an inequality representation, and the rate region can be calculated through the projection

$$\mathcal{R}_{in} = proj_{h_{S_1},h_{S_2},\alpha,\beta}(\mathcal{C}) \tag{14}$$

When expressed this way, one can use any one of a number of methods for polyhedral projection for polyhedra expressed in their inequality representation to calculate the rate region.

B. Results

Having developed the necessary inner bounds and algorithmic tools, we now pass to calculating inner bounds for the $(3, 2, 2)$ multiple source exact repair problem discussed in §II. First we consider the inner bound formed from combinations of basic scalar network codes.

Theorem 5. *The inner bound on the rate region obtained from Γ_{11}^2 for 2-source (3,2,2) exact repair problem is:*

$$\mathcal{R}_{in} = \left\{ (\alpha, \beta, h_{S_1}, h_{S_2}) \left| \begin{array}{r} h_{S_1} \geq 0 \\ h_{S_2} \geq 0 \\ 3\alpha - h_{S_1} - 2h_{S_2} \geq 0 \\ \alpha + \beta - h_{S_1} - h_{S_2} \geq 0 \\ 3\beta - h_{S_1} - h_{S_2} \geq 0 \\ 3\alpha - 2h_{S_1} - h_{S_2} \geq 0 \end{array} \right. \right\} \tag{15}$$

Polyhedron (Representation)	Input	Algorithm
Γ^q_N(V-rep)	List of non-isomorphic $\mathbb{F}_q$-representable matroids	Permute rank vectors to create isomorphisms and then conic hull
$\Gamma^{\subset space}_{N,k,q}$($V$-rep)	Γ^q_N(V-rep)	Linear projection (by simply deleting unwanted co-ordinates and redundancy removal)
$\mathcal{V}_R = \Gamma^q_N \cap \mathcal{L}_{1236}$ or $\Gamma^{\subset space}_{N,k,q} \cap \mathcal{L}_{1236}$ (V-rep)	Γ^q_N(V-rep) or $\Gamma^{\subset space}_{N,k,q}$($V$-rep)	Iterations of double description method of representation conversion (As noted in [9])
$\mathcal{R}_{in}$(H-rep)	$\mathcal{V}_R$ (V-rep)	Dual projection approach: Subsection IV-A

Figure 2. (column 1)Various polyhedra one has to compute during computer assisted achievability proof and their representations that are computed (column2) input polyhedra with specified representation and (column 3) algorithm used to compute polyhedra in column 1 using the input

Next we consider the inner bound created from combinations of basic vector network codes which allow one of the random variables to be represented as two bits from the matroid.

Theorem 6. *The inner bound on the rate region obtained from $\Gamma^{\subset space}_{11,12,2}$ for 2-source (3,2,2) exact repair problem is:*

$$\mathcal{R}_{in} = \left\{ (\alpha, \beta, h_{S_1}, h_{S_2}) \;\middle|\; \begin{array}{r} h_{S_1} \geq 0 \\ h_{S_2} \geq 0 \\ 2\alpha - h_{S_1} - h_{S_2} \geq 0 \\ 3\beta - h_{S_1} - h_{S_2} \geq 0 \\ \alpha + \beta - h_{S_1} - h_{S_2} \geq 0 \end{array} \right\} \quad (16)$$

These inner bounds were calculated using the techniques detailed in §III and §IV. These rate regions are plotted in Fig. 3 for several entropies for the purposes of comparison.

V. Conclusion

We considered the problem of finding achievable rate regions for $(n, k, d) = (3, 2, 2)$ exact repair problem with 2 sources under vector and scalar binary codes that form inner bounds on the real rate region. We developed an automated approach to the computation of achievable rate region that can be generalized to arbitrary (n, k, d) and arbitrary number of sources. While converse remains the next big step towards characterizing the actual rate region of the problem, achievable rate regions under specific field provide an important insight into the interplay between network structure and field size required to achieve optimum points in the rate region.

Acknowledgments

The authors gratefully acknowledge NSF support of this research through grants CCF-1016588 and CCF-1053702.

References

[1] V.R. Cadambe, S.A. Jafar, H. Maleki, K. Ramchandran, and Changho Suh. Asymptotic interference alignment for optimal repair of mds codes in distributed storage. *Information Theory, IEEE Transactions on*, 59(5):2974–2987, 2013.

[2] Steven Weber Congduan Li, John MacLaren Walsh. Matroid bounds on the region of entropic vectors. In *51th Annual Allerton Conference on Communication, Control and Computing*, October 2013.

[3] A.G. Dimakis, P.B. Godfrey, Y. Wu, M.J. Wainwright, and K. Ramchandran. Network coding for distributed storage systems. *Information Theory, IEEE Transactions on*, 56(9):4539–4551, 2010.

[4] K. Fukuda and A. Prodon. Double description method revisited. *Combinatorics and computer science*, pages 91–111, 1996.

[5] D. Hammer, A. Romashchenko, and N. Vereshchagin A. Shen. Inequalities for shannon entropy and kolmogorov complexity. *Journal of Computer and System Science*, 60(2):442–464, April 2000.

[6] Tracey Ho, Muriel Médard, Ralf Koetter, David R. Karger, Michelle Effros, Jun Shi, and Ben Leong. A random linear network coding approach to multicast. *IEEE Transactions on Information Theory*, 52(10):4413–4430, 2006.

[7] A. W. Ingleton. Representation of matroids. *Combin. Math. Appl.*, pages 149–167, 1971.

[8] Sidharth Jaggi, Peter Sanders, Philip A. Chou, Michelle Effros, Sebastian Egner, Kamal Jain, and Ludo M. G. M. Tolhuizen. Polynomial time algorithms for multicast network code construction. *IEEE Transactions on Information Theory*, 51(6):1973–1982, 2005.

[9] Congduan Li, J. Apte, J.M. Walsh, and S. Weber. A new computational approach for determining rate regions and optimal codes for coded networks. In *Network Coding (NetCod), 2013 International Symposium on*, pages 1–6, 2013.

[10] Yoshitake Matsumoto, Sonoko Moriyama, Hiroshi Imai, and David Bremner. Matroid enumeration for incidence geometry. *Discrete Comput. Geom.*, 47(1):17–43, January 2012.

[11] Dillon Mayhew and Gordon F. Royle. Matroids with nine elements. *Journal of Combinatorial Theory, Series B*, 98(2):415 – 431, 2008.

[12] T.S. Motzkin, H. Raiffa, GL Thompson, and R.M. Thrall. The double description method. *Contributions to the theory of games. Annals of Mathematics Studies. Princeton University Press*, 2(28):51–73, 1953.

[13] J. G. Oxley. *Matroid Theory*. Oxford University, 2011.

[14] 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. *Information Theory, IEEE Transactions on*, 57(8):5227–5239, 2011.

[15] 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. *Information Theory, IEEE Transactions on*, 58(3):1837–1852, 2012.

[16] N.B. Shah, K. V. Rashmi, P.V. Kumar, and K. Ramchandran. Interference alignment in regenerating codes for distributed storage: Necessity and code constructions. *Information Theory, IEEE Transactions on*, 58(4):2134–2158, 2012.

[17] Chao Tian. Rate region of the (4, 3, 3) exact-repair regenerating codes. In *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, pages 1426–1430, 2013.

[18] Marcel Wild. The asymptotic number of binary codes and binary matroids. *SIAM J. Discrete Math.*, 19(3):691–699, 2005.

[19] X. Yan, R.W. Yeung, and Zhen Zhang. An implicit characterization of the achievable rate region for acyclic multisource multisink network coding. *Information Theory, IEEE Transactions on*, 58(9):5625–5639, 2012.

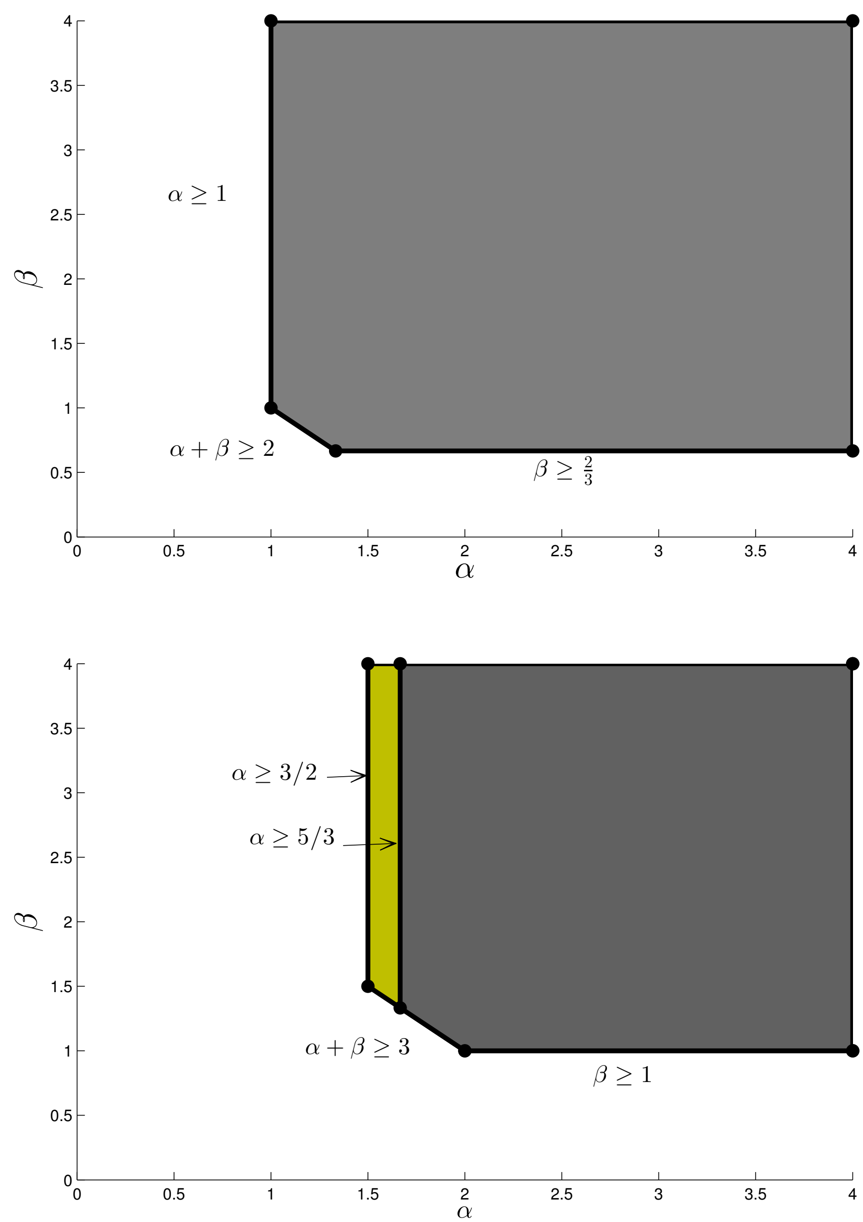

Figure 3. Rate regions obtained in theorem 5 (gray) and theorem 6 (yellow) for different values of source entropies. The region on top is for $h_{S_1} = h_{S_2} = 1$ while bottom region is for $h_{S_1} = 1, h_{S_2} = 2$. The region on top for both scalar and vector cases is the same.