Lexicographic Products and the Power of Non-Linear Network Coding

Anna Blasiak
Dept. of Computer Science
Cornell University.
ablasiak@cs.cornell.edu

Robert Kleinberg
Dept. of Computer Science
Cornell University.
rdk@cs.cornell.edu

Eyal Lubetzky
Microsoft Research
Redmond, WA, USA.
eyal@microsoft.com

Abstract— We introduce a technique for establishing and amplifying gaps between parameters of network coding and index coding problems. The technique uses linear programs to establish separations between combinatorial and coding-theoretic parameters and applies hypergraph lexicographic products to amplify these separations. This entails combining the dual solutions of the lexicographic multiplicands and proving that this is a valid dual solution of the product. Our result is general enough to apply to a large family of linear programs. This blend of linear programs and lexicographic products gives a recipe for constructing hard instances in which the gap between combinatorial or coding-theoretic parameters is *polynomially large*. We find polynomial gaps in cases in which the largest previously known gaps were only small constant factors or entirely unknown. Most notably, we show a polynomial separation between linear and non-linear network coding rates. This involves exploiting a connection between matroids and index coding to establish a previously unknown separation between linear and non-linear index coding rates. We also construct index coding problems with a polynomial gap between the broadcast rate and the trivial lower bound for which no gap was previously known.

1. Introduction

The problem of *Network Coding*, introduced by Ahlswede *et al* [2] in 2000, asks for the maximum rate at which information can be passed from a set of sources to a set of targets in a capacitated network. In practice, there are many examples where network coding provides faster transmission rates compared to traditional routing, e.g. [9] details a recent one in wireless networks. However, despite tremendous initial success in using network coding to solve some *broadcast* problems (those in which every receiver demands the same message), very little is known about how to compute or approximate the network coding rate in general. (See [13] for a survey of the topic.)

In the absence of general algorithms for solving network coding, attention has naturally turned to restricted models of coding (e.g. linear functions between vector spaces over finite fields) and to approximating network coding rates using graph-theoretic parameters (e.g. minimum cut [1] and the independence number [3]). Several of these variants provide bounds on the network coding rate, but the worst-case approximation factor of these bounds remains unknown. For example, it is known that there exists a network in which non-linear network coding can achieve a rate which exceeds the best linear network code by a factor of $\frac{11}{10}$ [7], but it is not known whether this gap[1] can be improved to $n^{1-\varepsilon}$, or even possibly to $\Theta(n)$.

In this paper we introduce a general technique for amplifying many of these gaps by combining linear programming with hypergraph product operations. For instance, this enables us to construct a family of network coding instances with n messages, in which the rate of the best non-linear network code exceeds the rate of the best (vector-)linear network code by a factor of at least n^{ε}. A crucial ingredient in our technique is *index coding* [5], [4], a class of communication problems in which a server holds a set of messages that it wishes to broadcast over a noiseless channel to a set of receivers. Each receiver is interested in one of the messages and has side-information comprising some subset of the other messages. The objective is to devise an optimal encoding scheme (one minimizing the broadcast length) that allows all the receivers to retrieve their required information. Following [3], we use β to denote the limiting value of the information rate (i.e., ratio of broadcast length to message length) of this optimal scheme, as the message length tends to infinity.

In our framework, index coding is most useful for isolating a sub-class of network coding problems that can be combined using lexicographic products. However, it is also an important and well-studied problem in its own right. Index coding is intimately related to network coding in general. It is essentially equivalent to the special case of network coding in which only one edge has finite capacity.[2] Additionally, [11] shows that linear network coding can be

The first author was supported by an NDSEG Graduate Fellowship, an AT&T Labs Graduate Fellowship, and an NSF Graduate Fellowship. The second author was supported in part by NSF grant CCF-0729102, AFOSR grant FA9550-09-1-0100, a Microsoft Research New Faculty Fellowship, and an Alfred P. Sloan Foundation Fellowship.

[1]The literature on network coding distinguishes between *linear* network codes, in which the messages are required to be elements of a finite field, and *vector-linear* network codes, in which the messages are elements of a finite-dimensional vector space over a finite field. Linear coding is weaker, and a gap of size $n^{1-\varepsilon}$ is known [10]. Vector-linear coding is much more powerful, and no gap larger than 11/10 was known prior to our work.

[2]The unique finite-capacity edge represents the broadcast channel. Each sender is connected to the tail of this edge, each receiver is connected to its head, and each receiver has incoming edges directly from a subset of the senders, representing the side-information.

reduced to linear index coding, thus implying that index coding captures much of the difficulty of network coding.

Index coding is also intricately related to other well-studied areas of mathematics. Connections between matroids and index coding were established in [12]; for example, that paper shows that realizability of a matroid over a field $\mathbb{F}$ is equivalent to linear solvability of a corresponding index coding problem. Index coding is also closely connected to graph theory: a special case of index coding can be described by an undirected graph G, representing a communication problem where a broadcast channel communicates messages to a set of vertices, each of whom has side-information consisting of the neighbors' messages. Letting $\alpha(G), \overline{\chi}(G)$ denote the independence and clique-cover numbers of G, respectively, one has

$$\alpha(G) \leq \beta(G) \leq \overline{\chi}(G)\,. \tag{1}$$

The first inequality above is due to an independent set being identified with a set of receivers with no mutual information, whereas the last one due to [5], [4] is obtained by broadcasting the bitwise XOR of the vertices per clique in the optimal clique-cover of G. As one consequence of the general technique we develop here, we settle an open question of [3] by proving that $\alpha(G)$ can differ from $\beta(G)$; indeed, we show that their ratio can be as large as $n^{0.139}$.

1.1. Contributions

We present a general technique that amplifies lower bounds for index coding problems using lexicographic hypergraph products in conjunction with linear programs that express information-theoretic inequalities. The use of such linear programs to prove lower bounds in network coding theory is not new, but, perhaps surprisingly, they have not gained widespread use in the analysis of index coding problems. We give an information-theoretic linear program, whose solution, b, gives the best known lower bound on β. However, our main innovation is the insight that this linear programming technique can be combined with the combinatorial technique of graph products to yield lower bounds for sequences of index coding and network coding problems. Specifically, we provide a lexicographic product operation on index coding problems along with an operation that combines dual solutions of the corresponding two linear programs. We show that the combined dual yields a dual solution of the linear program corresponding to the lexicographic product. Using this operation, we demonstrate that index coding lower bounds proven using linear programming behave supermultiplicatively under lexicographic products. This technical tool enables us to prove some new separation results answering open questions in the field.

Our technique not only applies to the standard linear programs used in network information theory (those that express entropy inequalities such as submodularity) but to *any* family of linear programs constructed using what we call a *tight homomorphic constraint schema*. In particular, if one can develop a tight homomorphic constraint schema that applies to a restricted class of codes (e.g. linear) then it becomes possible to prove lower bounds for this class of codes and amplify them using lexicographic products. We pursue this approach in establishing a large multiplicative gap between linear and non-linear network coding.

Theorem 1.1. *Lower bounds for index coding problems can be proven by solving a linear program whose constraints are valid for the class of coding functions being considered. If the linear program is constructed using a tight homomorphic constraint schema (see Section 3), then its optimum is supermultiplicative under the lexicographic product of two index coding problems.*

To separate linear from non-linear coding, we first produce a pair of linear inequalities that are valid information inequalities for tuples of random variables defined by linear functions over fields of odd (resp., even) characteristic, but not vice-versa. We obtain these inequalities by considering the Fano and non-Fano matroids; the former is a matroid that is only realizable in characteristic 2, while the latter is only realizable in odd characteristic and in characteristic 0. For each of the two matroids, we are able to transform a proof of its non-realizability into a much stronger quantitative statement about dimensions of vector spaces over a finite field. This, in turn, we transform into a tight homomorphic constraint schema of valid information inequalities for linear random variables.

We then use the connection between matroids and index coding [7], [8], [12] and these inequalities to give a pair of index coding instances where the best non-linear coding rate is strictly better than the best linear rate over a field of odd (resp.,even) characteristic. We do this by establishing a general theorem that says that for a matroid M, and an inequality that is violated for the rank function of M, there is an index coding problem for which the bound obtained by adding this inequality to the LP is strictly greater than b.

We can now plug the constraint schema into our lexicographic product technique and apply it to these two index coding problems to yield the aforementioned separation between (vector-)linear and non-linear network coding.

Theorem 1.2. *There exists an explicit family of network coding instances (based on index coding instances) with n messages and some fixed $\varepsilon > 0$ such that the non-linear rate is $\Omega(n^{\varepsilon})$ times larger than the linear rate.*

The largest previously known gap between the non-linear and linear rates for network coding was a factor of $\frac{11}{10}$ ([8]). No separation was known between these parameters for index coding (see [10], [3] for related separation results focusing on the weaker setting of scalar linear codes).

As explained above, given any index coding problem G we can write down an LP whose constraints are based

on information inequalities that gives a lower bound on β. It is the best known lower bound, and in many cases, strictly better than any previously known bound. Notably, we can show that the broadcast rate of the 5-cycle is at least $\frac{5}{2}$, giving the first known gap between the independence number α (which equals 2 for the 5-cycle) and the broadcast rate β. Amplifying this gap using lexicographic products, we can boost the ratio β/α to grow polynomially with n in a family of n-vertex graphs.

Theorem 1.3. *There exists an explicit family of index coding instances with n messages such that $\beta(G)$ is at least $\Omega(n^\delta)$ times larger than $\alpha(G)$, where $\delta = 1 - 2\log_5(2) \approx 0.139$.*

The remainder of the paper is organized as follows. In Section 2 we give a formal definition of index coding and the lexicographic product of two index coding problems. In Section 3 we describe a general class of LPs and prove they behave supermultiplicatively under lexicographic products. Section 4 is devoted to the proof of Theorem 1.3. In Section 5 we give a construction from matroids to index coding and prove a number of connections between properties of the matroid and the parameters of the corresponding index coding problem. Finally, in Section 6 we establish inequalities that are valid for linear codes over fields of odd (resp., even) characteristic and then use these to prove Theorem 1.2.

2. Definitions

An index coding problem is specified by a *directed hypergraph* $G = (V, E)$, where elements of V are thought of as messages, and $E \subseteq V \times 2^V$ is a set of directed hyperedges (v, S), each of which is interpreted as a receiver who already knows the messages in set S and wants to receive message v. Messages are drawn from a finite alphabet Σ, and a solution of the problem specifies a finite alphabet Σ_P to be used by the public channel, together with an encoding scheme $\mathcal{E} : \Sigma^{|V|} \to \Sigma_P$ such that, for any possible values of $(x_v)_{v\in V}$, every receiver (v, S) is able to decode the message x_v from the value of $\mathcal{E}(\vec{x})$ together with that receiver's side information. The minimum encoding length $\ell = \lceil \log_2 |\Sigma_P| \rceil$ for messages that are t bits long (i.e. $\Sigma = \{0,1\}^t$) is denoted by $\beta_t(G)$. As noted in [10], due to the overhead associated with relaying the side-information map to the server the main focus is on the case $t \gg 1$ and namely on the following *broadcast rate*.

$$\beta(G) \triangleq \lim_{t\to\infty} \frac{\beta_t(G)}{t} = \inf_t \frac{\beta_t(G)}{t} \tag{2}$$

(The limit exists by subadditivity.) This is interpreted as the average asymptotic number of broadcast bits needed per bit of input, that is, the asymptotic broadcast rate for long messages. We are also interested in the optimal rate when we require that Σ is a finite-dimensional vector space over a finite field $\mathbb{F}$, and the encoding function is linear. We denote this by $\lambda^{\mathbb{F}}$, and we denote the optimal linear rate over any field as λ.

A useful notion in index coding is the following *closure* operation with respect to G, a given instance of the problem: for a set of messages $S \subseteq V$, define

$$\mathrm{cl}(S) = \mathrm{cl}_G(S) = S \cup \{x \mid \exists (x, T) \in E \text{ s.t. } T \subseteq S\}\,. \tag{3}$$

The interpretation is that every message $x \in \mathrm{cl}(S)$ can be decoded by someone who knows all of the messages in S in addition to the broadcast message. In Section 5 when we discuss a transformation that associates an index coding problem to every matroid, the closure operation defined in this paragraph — when specialized to the index coding problems resulting from that transformation — will coincide with the usual matroid-theoretic closure operation.

We next define the lexicographic product operation for directed hypergraphs, then proceed to present Theorem 2.2 which demonstrates its merit in the context of index coding by showing that β is submultiplicative for this operation. The proof gives further intuition for the product operation.

Definition 2.1. The lexicographic product of two directed hypergraphs G, F, denoted by $G \bullet F$, is a directed hypergraph whose vertex set is the cartesian product $V(G) \times V(F)$. The edge set of $G \bullet F$ contains a directed hyperedge e for every pair of hyperedges $(e_G, e_F) \in E(G) \times E(F)$. If $e_G = (w_G, S_G)$ and $e_F = (w_F, S_F)$, then the head of $e = (e_G, e_F)$ is the ordered pair (w_G, w_F) and its tail is the set $(S_G \times V(F)) \cup (\{w_G\} \times S_F)$. Denote by $G^{\bullet n}$ the n-fold lexicographic power of G.

Remark. In the special case where the index coding problem is defined by a graph[3] the above definition coincides with the usual lexicographic graph product (where $G \bullet F$ has the vertex set $V(G) \times V(F)$ and an edge from (u, v) to (u', v') iff either $(u, u') \in E(G)$ or $u = u'$ and $(v, v') \in E(F)$).

Theorem 2.2. *The broadcast rate is submultiplicative under the lexicographic product of index coding problems. That is, $\beta(G \bullet F) \le \beta(G)\,\beta(F)$ for any two directed hypergraphs G and F.*

Proof: Let $\varepsilon > 0$ and, recalling the definition of β in (2) as the limit of β_t/t, let K be a sufficiently large integer such that for all $t \ge K$ we have $\beta_t(G)/t \le \beta(G) + \varepsilon$ as well as $\beta_t(F)/t \le \beta(F) + \varepsilon$. Let $\Sigma = \{0,1\}^K$ and consider the following scheme for the index coding problem on $G \bullet F$ with input alphabet Σ, which will consist of an inner and an outer code.

Let $\mathcal{E}_F$ denote an encoding function for F with input alphabet Σ achieving an optimal rate, i.e. minimizing

[3]When there are n messages and exactly n receivers, w.l.o.g. receiver i wants the message x_i and one can encode the side-information by a graph on n vertices which contains the edge (i, j) iff receiver i knows the message x_j.

$\log(|\Sigma_P|)/\log(|\Sigma|)$. For each $v \in V(G)$, the inner code applies $\mathcal{E}_F$ to the $|V(F)|$-tuple of messages indexed by the set $\{v\} \times V(F)$, obtaining a message m_v. Note that our assumption on $|\Sigma|$ implies that the length of m_v is equal to K' for some integer K' such that $K \le K' \le (\beta(F)+\varepsilon)K$. Next, let $\mathcal{E}_G$ denote an optimal encoding function for G with input $\{0,1\}^{K'}$. The outer code applies $\mathcal{E}_G$ to $\{m_v\}_{v\in V(G)}$ and the assumption on K ensures its output is at most $(\beta(G)+\varepsilon)K'$ bits long.

To verify that the scheme is a valid index code, consider a receiver in $G \bullet F$ represented by $e = ((w_G, w_F), (S_G \times V(F)) \cup (\{w_G\} \times S_F))$. To decode (w_G, w_F), the receiver first computes m_v for all $v \in S_G$. Since $\mathcal{E}_G$ is valid for G, receiver e can compute m_{w_G}, and since $\mathcal{E}_F$ is valid for F, this receiver can use the messages indexed by $\{w_G\} \times S_F$ along with m_{w_G} to compute (w_G, w_F).

Altogether, we have an encoding of K bits using at most $(\beta(F)+\varepsilon)(\beta(G)+\varepsilon)K$ bits of the public channel, and the required result follows from letting $\varepsilon \to 0$. ∎

3. Linear programming

In this section we derive a linear program whose value constitutes a lower bound on the broadcast rate, and we prove that the value of the LP behaves supermultiplicatively under lexicographic products. In fact, rather than working with a specific linear program, we work with a general class of LP's having two types of constraints: those dictated by the network structure (which are the same for all LP's in the general class), and additional constraints depending only on the vertex set, generated by a *constraint schema*, i.e. a procedure for enumerating a finite set of constraints given an arbitrary finite index set. We identify some axioms on the constraint schema that constitute a sufficient condition for the LP value to be supermultiplicative. An example of a constraint schema which is important in network information theory is *submodularity*. For a given index set I, the submodularity schema enumerates all of the constraints of the form $z_S + z_T \ge z_{S\cap T} + z_{S\cup T}$ where S, T range over subsets of I.

Now we explain the general class of LPs which behave submultiplicatively under the lexicographic product and give bounds on β. Given an index code, if we sample each message independently and uniformly at random, we obtain a finite probability space on which the messages and the public channel are random variables. If S is a subset of these random variables, we will denote the Shannon entropy of the joint distribution of the variables in S by $H(S)$. If $S \subseteq T \subseteq \mathrm{cl}(S)$ then every message in $T \setminus S$ can be decoded given the messages in S and the public channel p, and consequently $H(S\cup\{p\}) = H(T\cup\{p\})$. More generally, if we normalize entropy (i.e. choose the base of the logarithm) so that $H(x) = 1$ for each message x, then for every $S \subseteq T$ we have

$$H(T \cup \{p\}) - H(S \cup \{p\}) \le |T \setminus \mathrm{cl}(S)| \triangleq c_{ST}, \qquad (4)$$

$$\begin{array}{lll}
\min & z_\emptyset & \\
\text{s.t.} & z_I = |I| & (w) \\
\forall S \subset T & z_T - z_S \le c_{ST} & (x) \\
 & Az \ge 0 & (y) \\
 & & \\
\max & |I| \cdot w - \sum_{S\subset T} c_{ST} x_{ST} & \\
\text{s.t.} & \sum_q a_{qS} y_q + \sum_{T\supset S} x_{ST} - \sum_{T\subset S} x_{TS} = 0 & \\
 & \qquad\qquad\qquad\qquad\qquad\qquad \forall S \ne \emptyset, I & \\
 & \sum_q a_{q\emptyset} y_q + \sum_{T\ne\emptyset} x_{\emptyset T} = 1 & \\
 & \sum_q a_{qI} y_q - \sum_{T\ne I} x_{TI} + w = 0 & \\
 & x, y \ge 0 &
\end{array}$$

Figure 1. The LP and its dual.

where the above is taken as the definition of c_{ST}. This implies that for any index code we obtain a feasible solution of the primal LP in Figure 1 by setting $z_S = H(S \cup \{p\})$ for every S. Indeed, the first constraint expresses the fact that the value of p is determined by the values of the n messages, which are mutually independent. The second constraint was discussed above. The final line of the LP represents a set of constraints, corresponding to the rows of the matrix $A = (a_{qS})$, that are universally valid for any tuple of random variables indexed by the message set I. For instance, it is well known that the entropy of random variables has the submodularity property: $H(S) + H(T) \ge H(S \cup T) + H(S \cap T)$ if S, T are any two sets of random variables on the same sample space. So, for example, the rows of the constraint matrix A could be indexed by pairs of sets S, T, with entries in the (S, T) row chosen so that it represents the submodularity constraint (namely $a_{qS} = a_{qT} = 1$, $a_{q\,S\cap T} = a_{q\,S\cup T} = -1$ and all other entries of row a of A are zero). Noting that $H(\{p\}) \le \beta(G)$ we can altogether conclude the following theorem.

Theorem 3.1. *For an index coding problem G, let $\mathfrak{B}(G)$ be the LP in Figure 1 when A represents the submodularity constraints and let $b(G)$ be its optimal solution. Then $b(G) \le \beta(G)$.*

It is known that entropies of sets of random variables satisfy additional linear inequalities besides submodularity; if desired, the procedure for constructing the matrix A could be modified to incorporate some of these inequalities. Alternatively, in the context of restricted classes of encoding and decoding functions (e.g. linear functions) there may be additional inequalities that are specific to that class of functions, in which case the constraint matrix A may incorporate these inequalities and we obtain a linear program that is valid for this restricted model of index coding but not

valid in general. We will utilize such constraints in Section 6 when proving a separation between linear and non-linear network coding.

Definition 3.2. A *constraint schema* associates to each finite index set I a finite set $\mathcal{Q}(I)$ (indexing constraints) and a matrix $A(I)$ with rows indexed by $\mathcal{Q}(I)$ and columns indexed by $\mathcal{P}(I)$, the power set of I. In addition, to each Boolean lattice homomorphism[4] $h : \mathcal{P}(I) \to \mathcal{P}(J)$ it associates a function $h_* : \mathcal{Q}(I) \to \mathcal{Q}(J)$.

Let $\mathbf{1}$ be the $\mathcal{P}(I)$-indexed vector such that $\mathbf{1}_S = 1$ for all S, and let $\mathbf{1}_i$ be the vector where $(\mathbf{1}_i)_S = 1$ for all S containing i and otherwise $(\mathbf{1}_i)_S = 0$. We say that a constraint schema is *tight* if $A(I)\mathbf{1} = A(I)\mathbf{1}_i = 0$ for every index set I and element $i \in I$.

Given h and h_* let P_h and Q_h be matrices representing the linear transformations they induce on $\mathbb{R}^{\mathcal{P}(I)} \to \mathbb{R}^{\mathcal{P}(J)}$ and $\mathbb{R}^{\mathcal{Q}(I)} \to \mathbb{R}^{\mathcal{Q}(J)}$, respectively. That is, P_h and Q_h have zeros everywhere except $(P_h)_{h(S)S} = 1$ and $(Q_h)_{h_*(q)q} = 1$. We say that a constraint schema is *homomorphic* if it satisfies $A(J)^\mathsf{T} Q_h = P_h A(I)^\mathsf{T}$ for every Boolean lattice homomorphism $h : \mathcal{P}(I) \to \mathcal{P}(J)$.

Example 3.3. Earlier we alluded to the *submodularity constraint schema*. This is the constraint schema that associates to each index set I the constraint-index set $\mathcal{Q}(I) = \mathcal{P}(I) \times \mathcal{P}(I)$, along with the constraint matrix $A(I)$ whose entries are as follows. In row (S,T) and column U, we have an entry of 1 if $U = S$ or $U = T$, an entry of -1 if $U = S \cap T$ or $U = S \cup T$, and otherwise 0. (If any two of S, T, $S \cap T$, $S \cup T$ are equal, then that row of $A(I)$ is set to zero.) It is easy to verify that $A(I)\mathbf{1} = A(I)\mathbf{1}_i = 0$ for all $i \in I$, thus the schema is tight. For a homomorphism h, the corresponding mapping of constraint sets is $h_*(S,T) = (h(S), h(T))$. We claim that, equipped with this mapping of $h \to h_*$, the constraint schema is homomorphic. Indeed, to verify that $A(J)^\mathsf{T} Q_h = P_h A(I)^\mathsf{T}$ take any two sets $S, T \subset I$ and argue as follows to show that $u = P_h\, A(I)^\mathsf{T}\, e_{S,T}$ and $v = A(J)^\mathsf{T}\, Q_h\, e_{S,T}$ are identical (here and henceforth $e_{X,Y}$ denotes the standard basis vector of $\mathbb{R}^{\mathcal{P}(I)}$ having 1 in coordinate (X,Y) for $X, Y \subset I$). First observe that $A(I)^\mathsf{T}\, e_{S,T}$ is the vector $\tilde{u} \in \mathbb{R}^{\mathcal{P}(()I)}$ which has 0 entries everywhere except $\tilde{u}_S = \tilde{u}_T = 1$ and $\tilde{u}_{S\cup T} = \tilde{u}_{S\cap T} = -1$ provided that $S \not\subseteq T \not\subseteq S$, otherwise $\tilde{u} = 0$. As such, $u = P_h \tilde{u}$ has 0 entries everywhere except

$$u_{h(S)} = u_{h(T)} = 1\,, \quad u_{h(S\cup T)} = u_{h(S\cap T)} = -1$$

provided that $S \not\subseteq T \not\subseteq S$ and furthermore $h(S) \not\subseteq h(T) \not\subseteq h(S)$, otherwise $u = 0$ (for instance, if $S \subseteq T$ then $\tilde{u} = 0$ and so $u = 0$, whereas if $h(S) \subseteq h(T)$ then $\tilde{u}$ belongs to the kernel of P_h). Similarly, $Q_h\, e_{S,T} = e_{h(S),h(T)}$ and therefore $v = A(J)^\mathsf{T}\, e_{h(S),h(T)}$ has 0 entries everywhere except

$$v_{h(S)} = v_{h(T)} = 1\,, \quad v_{h(S)\cup h(T)} = v_{h(S)\cap h(T)} = -1$$

provided that $h(S) \not\subseteq h(T) \not\subseteq h(S)$, otherwise $v = 0$. To see that $u = v$ note that if $h(S) \subseteq h(T)$ then $u = v = 0$, and if $S \subseteq T$ then again we get $h(S) \subseteq h(T)$ due to monotonicity (recall that h is a lattice homomorphism) and so $u = v = 0$. Adding the analogous statements obtained from reversing the roles of S, T, it remains only to verify that $u = v$ in case $h(S) \not\subseteq h(T) \not\subseteq h(S)$, which reduces by the above definitions of u and v to requiring that $h(S \cup T) = h(S) \cup h(T)$ and $h(S \cap T) = h(S) \cap h(T)$. Both requirements are satisfied by definition of a Boolean lattice homomorphism, and altogether we conclude that the submodularity constraint schema is homomorphic.

Theorem 3.4. *Let A be a tight homomorphic constraint schema. For every index coding problem let $\rho(G)$ denote the optimum of the LP in Figure 1 when $I = V(G)$ and the constants c_{ST} are defined as in* (4)*. Then for every two index coding problems G and F, we have $\rho(G \bullet H) \geq \rho(G)\,\rho(F)$.*

Remark 3.5. The condition that A needs to be tight and homomorphic is surprisingly unrestrictive. In Section 6, Lemmas 6.3 and 6.4 show that a large class of inequalities can be expressed as tight homomorphic constraint schema.

Proof: It will be useful to rewrite the constraint set of the dual LP in a more succinct form. First, if x is any vector indexed by pairs S, T such that $S \subset T \subseteq I$, let $\nabla x \in \mathbb{R}^{\mathcal{P}(I)}$ denote the vector such that for all S, $(\nabla x)_S = \sum_{T \supset S} x_{ST} - \sum_{T \subset S} x_{TS}$. Next, for a set $S \subseteq I$, let e_S denote the standard basis vector vector in $\mathbb{R}^{\mathcal{P}(I)}$ whose S component is 1. Then the entire constraint set of the dual LP can be abbreviated to the following:

$$A^\mathsf{T} y + \nabla x + w e_I = e_\emptyset\,, \quad x, y \geq 0\,. \tag{5}$$

Some further simplifications of the dual can be obtained using the fact that the constraint schema is tight. For example, multiplying the left and right sides of (5) by the row vector $\mathbf{1}^\mathsf{T}$ gives

$$\mathbf{1}^\mathsf{T} A^\mathsf{T} y + \mathbf{1}^\mathsf{T} \nabla x + w = 1\,.$$

By the tightness of the constraint schema $\mathbf{1}^\mathsf{T} A^\mathsf{T} = 0$. It is straightforward to verify that $\mathbf{1}^\mathsf{T} \nabla x = 0$ and after eliminating these two terms from the equation above, we find simply that $w = 1$. Similarly, if we multiply the left and right sides of (5) by the row vector $\mathbf{1}_i^\mathsf{T}$ and substitute $w = 1$, we obtain $\mathbf{1}_i^\mathsf{T} A^\mathsf{T} y + \mathbf{1}_i^\mathsf{T} \nabla x + 1 = 0$ and consequently (again by the tightness) we arrive at $1 = -\mathbf{1}_i^\mathsf{T} \nabla x$. At the same time, $-\mathbf{1}_i^\mathsf{T} \nabla x = \sum_{\substack{S \subset T \\ i \in T \setminus S}} x_{ST}$ by definition of ∇x, hence summing over all $i \in I$ yields that $|I| = \sum_{S \subset T} |T \setminus S|\, x_{ST}$. Plugging in this expression for $|I|$ and $w = 1$, the LP

[4] A Boolean lattice homomorphism preserves unions and intersections, but does not necessarily map the empty set to the empty set nor the universal set to the universal set, and does not necessarily preserve complements.

objective of the dual can be rewritten as

$$|I| - \sum_{S\subset T} c_{ST} x_{ST} = \sum_{S\subset T} \left(|T\setminus S| - c_{ST}\right) x_{ST} = \sum_{S\subset T} |T\cap(\mathrm{cl}(S)\setminus S)|\, x_{ST},$$

where the last equation used the fact that $c_{ST} = |T\setminus \mathrm{cl}(S)|$. We now define

$$d(S,T) = |T\cap(\mathrm{cl}(S)\setminus S)|$$

and altogether we arrive at the following reformulation of the dual LP.

$$\begin{array}{ll} \max & \sum_{S\subset T} d(S,T)\, x_{ST} \\ \text{s.t.} & A^{\mathsf{T}} y + \nabla x = e_\emptyset - e_I \\ & x, y \geq 0\,. \end{array} \tag{6}$$

Now suppose that $(\xi^G,\eta^G),(\xi^F,\eta^F)$ are optimal solutions of the dual LP for G,F, achieving objective values $\rho(G)$ and $\rho(F)$, respectively. (Here ξ,η play the role of x,y from (6), resp.) We will show how to construct a pair of vectors $(\xi^{G\bullet F},\eta^{G\bullet F})$ that is feasible for the dual LP of $G\bullet F$ and achieves an objective value of at least $\rho(G)\,\rho(F)$. The construction is as follows. Let $g : \mathcal{P}(V(G)) \to \mathcal{P}(V(G\bullet F))$ be the mapping $g(X) = X\times V(F)$. For sets $S\subset T\subseteq V(G)$, let $h^{ST} : \mathcal{P}(V(F)) \to \mathcal{P}(V(G\bullet F))$ be the mapping $h^{ST}(X) = (T\times X)\cup(S\times V(F))$. Observe that both mappings are Boolean lattice homomorphisms.

To gain intuition about the mappings g, h^{ST} it is useful to think of obtaining the vertex set of $G\bullet F$ by replacing every vertex of G with a copy of F. Here $g(\{v\})$ maps the vertex v in G to the copy of F that replaces v. The mapping $h^{ST}(\{u\})$ maps a vertex u in F to the vertex u in the copies of F that replace vertices in T, and then adds the set $\{u\}\times V(F)$.

Recall that Definition 3.2 associates two matrices P_h, Q_h to every Boolean lattice homomorphism $h : \mathcal{P}(I) \to \mathcal{P}(J)$. It is also useful to define a matrix R_h as follows: the columns and rows of R_h are indexed by pairs $S\subset T\subseteq I$ and $X\subset Y\subseteq J$, respectively, with the entry in row XY and column ST being equal to 1 if $X = h(S)$ and $Y = h(T)$, otherwise 0. Under this definition,

$$\nabla(R_h x) = P_h \nabla x \quad \text{for any } x\in\mathbb{R}^{\mathcal{P}(I)}\,. \tag{7}$$

Indeed, if $x = e_{S,T}$ for some $S\subset T\subseteq I$ then $\nabla e_{S,T} = e_S - e_T$ and so $P_h\, e_{S,T} = e_{h(S)} - e_{h(T)}$, whereas $\nabla(R_h e_{S,T}) = \nabla(e_{h(S),h(T)}) = e_{h(S)} - e_{h(T)}$.

We may now define

$$\xi^{G\bullet F} = \sum_{S\subset T} (\xi^G)_{ST}\,(R_{h^{ST}}\,\xi^F)\,, \tag{8}$$

$$\eta^{G\bullet F} = Q_g\,\eta^G + \sum_{S\subset T} (\xi^G)_{ST}\,(Q_{h^{ST}}\,\eta^F)\,. \tag{9}$$

In words, the dual solution for $G\bullet F$ contains a copy of the dual solution for F lifted according to h^{ST} for every pair $S\subset T$ and one copy of the dual solution of G lifted according to g. The feasibility of $(\xi^{G\bullet F},\eta^{G\bullet F})$ will follow from multiple applications of the homomorphic property of the constraint schema and the feasibility of (ξ^F,η^F) and (ξ^G,η^G), achieved by the following claim.

Claim 3.6. *The pair $(\xi^{G\bullet F},\eta^{G\bullet F})$ as defined in* (8),(9) *is a feasible dual solution.*

Proof: The matrices Q_g, $R_{h^{ST}}$, $Q_{h^{ST}}$ all have $\{0,1\}$-valued entries thus clearly $\xi^{G\bullet F},\eta^{G\bullet F}\geq 0$. Letting $A = A(G\bullet F)$, we must prove that $A^{\mathsf{T}}\eta^{G\bullet F} + \nabla\xi^{G\bullet F} = e_\emptyset - e_{V(G\bullet F)}$. Plugging in the values of $(\xi^{G\bullet F},\eta^{G\bullet F})$ we have

$$\begin{aligned} A^{\mathsf{T}}\eta^{G\bullet F} + \nabla\xi^{G\bullet F} &= A^{\mathsf{T}} Q_g \eta^G + \sum_{S\subset T} (\xi^G)_{ST}(A^{\mathsf{T}} Q_{h^{ST}}\eta^F) \\ &\quad + \sum_{S\subset T} (\xi^G)_{ST}\,\nabla(R_{h^{ST}}\,\xi^F) \\ &= P_g A(G)^{\mathsf{T}}\eta^G + \sum_{S\subset T} (\xi^G)_{ST}\Big(P_{h^{ST}} A(F)^{\mathsf{T}}\eta^F + \nabla(R_{h^{ST}}\,\xi^F)\Big) \end{aligned} \tag{10}$$

where the second equality applied the homomorphic property of the constraint schema. To treat the summation in the last expression above, recall (7) which implies that

$$\begin{aligned} P_{h^{ST}} A(F)^{\mathsf{T}}\eta^F + \nabla(R_{h^{ST}}\,\xi^F) &= P_{h^{ST}} A(F)^{\mathsf{T}}\eta^F + P_{h^{ST}}\nabla\xi^F \\ &= P_{h^{ST}}(e_\emptyset - e_{V(F)})\,, \end{aligned} \tag{11}$$

with the last equality due to the fact that (ξ^F,η^F) achieves the optimum of the dual LP for F. Recalling that $P_h e_S = e_{h(S)}$ for any h and combining it with the facts $h^{ST}(\emptyset) = S\times V(F)$ and $g(S) = S\times V(F)$ gives $P_{h^{ST}} e_\emptyset = e_{S\times V(F)} = P_g e_S$. Similarly, since $h^{ST}(V(F)) = T\times V(F)$ we have $P_{h^{ST}} e_{V(F)} = e_{T\times V(F)} = P_g e_T$, and plugging these identities in (11) combined with (10) gives:

$$A^{\mathsf{T}}\eta^{G\bullet F} + \nabla\xi^{G\bullet F} = P_g\left[A(G)^{\mathsf{T}}\eta^G + \sum_{S\subset T}(\xi^G)_{ST}\,(e_S - e_T)\right].$$

Collecting together all the terms involving e_S for a given $S\in\mathcal{P}(I)$, we find that the coefficient of e_S is $\sum_{T\supset S}(\xi^G)_{ST} - \sum_{T\subset S}(\xi^G)_{ST} = (\nabla\xi^G)_S$. Hence,

$$\begin{aligned} A^{\mathsf{T}}\eta^{G\bullet F} + \nabla\xi^{G\bullet F} &= P_g\left[A(G)^{\mathsf{T}}\eta^G + \nabla\xi^G\right] \\ &= P_g\left[e_\emptyset - e_{V(G)}\right] = e_\emptyset - e_{V(G\bullet F)}\,, \end{aligned}$$

where the second equality was due to (ξ^G,η^G) achieving the optimum of the dual LP for G. ∎

To finish the proof, we must evaluate the dual LP objective and show that it is at least $\rho(G)\,\rho(F)$, as the next claim establishes:

Claim 3.7. *The LP objective for the dual solution given in Claim 3.6 has value at least $\rho(G)\,\rho(F)$.*

Proof: To simplify the notation, throughout this proof we will use K, L to denote subsets of $V(G\bullet F)$ while

referring to subsets of $V(G)$ as S,T and to subsets of $V(F)$ as X,Y. We have

$$\begin{aligned}
&\sum_{K\subset L} d(K,L)(\xi^{G\bullet F})_{KL} \qquad (12)\\
&= \sum_{K\subset L} d(K,L) \sum_{S\subset T} (\xi^G)_{ST}\,(R_{h^{ST}}\,\xi^F)_{KL}\\
&= \sum_{S\subset T} (\xi^G)_{ST}\Big(\sum_{K\subset L} d(K,L)\,(R_{h^{ST}}\,\xi^F)_{KL}\Big)\\
&= \sum_{S\subset T} (\xi^G)_{ST}\Big(\sum_{X\subset Y} d\big(h^{ST}(X),h^{ST}(Y)\big)\,(\xi^F)_{XY}\Big),
\end{aligned}$$

where the last identity is by definition of R_h.

At this point we are interested in deriving a lower bound on $d\big(h^{ST}(X),h^{ST}(Y)\big)$, to which end we first need to analyze $\mathrm{cl}_{G\bullet F}(h^{ST}(X))$. Recall that $E(G\bullet F)$ consists of all hyperedges of the form (w,K) with $w=(w_G,w_F)$ and $K=(W_G\times V(F))\cup(\{w_G\}\times W_F)$ for some pair of edges $(w_G,W_G)\in E(G)$ and $(w_F,W_F)\in E(F)$. We first claim that for any $S\subset T$ and $X\subset V(F)$, if $\hat{X}$ denotes $h^{ST}(X)$, then

$$\mathrm{cl}_{G\bullet F}(\hat{X})\setminus\hat{X} \;\supseteq\; \big((\mathrm{cl}_G(S)\setminus S)\cap T\big)\times\big(\mathrm{cl}_F(X)\setminus X\big). \qquad (13)$$

To show this, let L denote the set on the right side of (13). Note that L contains no ordered pairs whose first component is in S or whose second component is in X, and therefore L is disjoint from $\hat{X}=(T\times X)\cup(S\times V(F))$. Consequently, it suffices to show that $\mathrm{cl}_{G\bullet F}\left(\hat{X}\right)\supseteq L$. Consider any $w=(w_G,w_F)$ belonging to L. As $w_G\in\mathrm{cl}_G(S)\setminus S$, there must exist an edge $(w_G,W_G)\in E(G)$ such that $W_G\subseteq S$. Similarly, there must exist an edge $(w_F,W_F)\in E(F)$ such that $W_F\subseteq X$. Letting $K=(W_G\times V(F))\cup(\{w_G\}\times W_F)$, we find that $K\subseteq(S\times V(F))\cup(T\times X)=\hat{X}$ and that $(w,K)\in E(G\bullet F)$, implying $w\in\mathrm{cl}_{G\bullet F}\left(\hat{X}\right)$ as desired.

Let $\hat{X}=h^{ST}(X)$ and $\hat{Y}=h^{ST}(Y)$, and recall that $d(\hat{X},\hat{Y})$ is defined as $\left|\left(\mathrm{cl}_{G\bullet F}(\hat{X})\setminus\hat{X}\right)\cap\hat{Y}\right|$. Using (13) and noting that $\hat{Y}\supseteq(T\times Y)$ we find that

$$\begin{aligned}
&(\mathrm{cl}_{G\bullet F}(\hat{X})\setminus\hat{X})\cap\hat{Y} \;\supseteq\\
&\quad((\mathrm{cl}_G(S)\setminus S)\cap T)\times((\mathrm{cl}_F(X)\setminus X)\cap Y)
\end{aligned}$$

and hence

$$\begin{aligned}
d(\hat{X},\hat{Y}) &\geq |(\mathrm{cl}_G(S)\setminus S)\cap T|\cdot|(\mathrm{cl}_F(X)\setminus X)\cap Y|\\
&= d(S,T)\,d(X,Y)\,.
\end{aligned}$$

Plugging this bound into (12) we find that

$$\begin{aligned}
&\sum_{K\subset L} d(K,L)(\xi^{G\bullet F})_{KL}\\
&\geq \sum_{S\subset T}(\xi^G)_{ST}\sum_{X\subset Y} d(S,T)d(X,Y)(\xi^F)_{XY}\,,
\end{aligned}$$

and since the last expression above is precisely equal to

$$\Big(\sum_{S\subset T} d(S,T)(\xi^G)_{ST}\Big)\Big(\sum_{X\subset Y} d(X,Y)(\xi^F)_{XY}\Big)=\rho(G)\,\rho(F)$$

this concludes the proof. ∎

Combining Claims 3.6 and 3.7 concludes the proof of the Theorem 3.4. ∎

4. Separation between α and β

To prove Theorem 1.3, we start by using Theorem 3.1 to show that $\beta(C_5)>\alpha(C_5)$ where C_5 is the 5-cycle. Then we apply the power of Theorem 3.4 to transform this constant gap on C_5 to a polynomial gap on C_5^k.

First we show that $\beta(C_5)\geq b(C_5)=\frac{5}{2}$. We can show that $b(C_5)\geq\frac{5}{2}$ by providing a feasible dual solution for the LP $\mathfrak{B}$ with value $\frac{5}{2}$. This can easily be achieved by listing a set of primal constraints whose variables sum and cancel to show that $z_\emptyset\geq\frac{5}{2}$. Labeling the vertices of C_5 by $1,2,3,4,5$ sequentially, such a set of constraints is given below. It is helpful to note that in an index coding problem defined by an undirected graph, $x\in\mathrm{cl}(S)$ if $x\in S$ or all the neighbors of x are in S.

$$\begin{aligned}
2 &\geq z_{\{1,3\}} - z_\emptyset\\
2 &\geq z_{\{2,4\}} - z_\emptyset\\
1 &\geq z_{\{5\}} - z_\emptyset\\
0 &\geq z_{\{1,2,3\}} - z_{\{1,3\}}\\
0 &\geq z_{\{2,3,4\}} - z_{\{2,4\}}\\
z_{\{2,3,4\}} + z_{\{1,2,3\}} &\geq z_{\{2,3\}} + z_{\{1,2,3,4\}}\\
z_{\{2,3\}} + z_{\{5\}} &\geq z_\emptyset + z_{\{2,3,5\}}\\
0 &\geq z_{\{1,2,3,4,5\}} - z_{\{1,2,3,4\}}\\
0 &\geq z_{\{1,2,3,4,5\}} - z_{\{2,3,5\}}\\
z_{\{1,2,3,4,5\}} &= 5\\
z_{\{1,2,3,4,5\}} &= 5
\end{aligned}$$

Applying Theorem 3.4 we deduce that for any integer $k\geq 1$ the k-th lexicographic power of C_5 satisfies $\beta(C_5^k)\geq b(C_5^k)\geq\left(\frac{5}{2}\right)^k$. Furthermore, $\alpha(C_5)=2$ and it is well known that the independence number is multiplicative on lexicographic products and so $\alpha(C_5^k)=2^k$. Altogether, C_5^k is a graph on $n=5^k$ vertices with $\alpha=n^{\log_5(2)}$ and $\beta\geq n^{1-\log_5(2)}$, implying our result.

5. Matroids and index coding

Recall that a matroid is a pair $M=(E,r)$ where E is a ground set and $r:2^E\to\mathbb{N}$ is a rank function satisfying

(i) $r(A)\leq|A|$ for all $A\subseteq E$;
(ii) $r(A)\leq r(B)$ for all $A\subseteq B\subseteq E$ (monotonicity);
(iii) $r(A)+r(B)\geq r(A\cup B)+r(A\cap B)$ for all $A,B\subseteq E$ (submodularity).

The rank vector of a matroid, $\vec{\mathbf{r}}(M)$, is a $2^{|E|}$-dimensional vector indexed by subsets of E, such that its S-th coordinate

is $r(S)$. A subset $S \subseteq E$ is called *independent* if $r(S) = |S|$ and it is called a *basis* of M if $r(S) = |S| = r(E)$.

In this section we give a construction mapping a matroid to an instance of index coding that exactly captures the dependencies in the matroid. We proceed to show some useful connections between matroid properties and the broadcast rate of the corresponding index coding problem.

Definition 5.1. Let $M = (E, r)$ be a matroid. The hypergraph index coding problem *associated* to M, denoted by G_M, has a message set E and all receivers of the form

$$\left\{(x, S) \mid x \in E\,,\, S \subseteq E\,,\, r(S) = r(S \cup \{x\})\right\}.$$

Remark. A similar yet slightly more complicated construction was given in [12]. Our construction is (essentially) a subset of the one appearing there. A construction that maps a matroid to a network coding problem is given in [7], [8]. They prove an analog of Proposition 5.2.

Proposition 5.2. *For a matroid $M = (E, r)$, $b(G_M) = |E| - r(E)$.*

Proof: In what follows we will let $n = |E|$ and $r = r(E)$. To show that $b(G_M) \leq n - r$ it suffices to show $z_S = r(S) + n - r$ is a feasible primal solution to the LP $\mathfrak{B}(G_M)$. The feasibility of constraints (w) and (x) follows trivially from the definition of G_M and properties of a matroid. The feasibility of (y) : $z_T - z_S \leq c_{ST}$ $\forall S \subset T$ follows from repeated application of submodularity:

$$\begin{aligned} z_T - z_S = r(T) - r(S) &\leq \sum_{x \in T \setminus S} r(S \cup \{x\}) - r(S) \\ &\leq \sum_{x \in \mathrm{cl}(S)} (r(S \cup \{x\}) - r(S)) \\ &\quad + \sum_{x \in T \setminus \mathrm{cl}(S)} r(\{x\}) \leq |T \setminus \mathrm{cl}(S)| = c_{ST}. \end{aligned}$$

For the reverse inequality, let S be any basis of M and note that $z_\emptyset = z_E - (z_E - z_S) - (z_S - z_\emptyset) \geq n - r$. ∎

The following definition relaxes the notion of a representation for a matroid.

Definition 5.3. A matroid $M = (E, r)$ with $|E| = n$ is *under-representable* in d dimensions over a finite field $\mathbb{F}$ if there exists a $d \times n$ matrix with entries in $\mathbb{F}$ and columns indexed by elements of E such that (i) the rows are independent and (ii) if $r(x \cup S) = r(S)$ then the columns indexed by $x \cup S$ are dependent.

Observe that if a matrix represents M then it also under-represents M. We next show a relation between under-representations for M over $\mathbb{F}$ and the *scalar* linear rate $\lambda_1^{\mathbb{F}}$, where the alphabet vector space, over which the encoding functions are required to be linear, is single-dimensional. Note that $\lambda^{\mathbb{F}} \leq \lambda_1^{\mathbb{F}}$. The following is the analogue of Theorem 8 in [12] for our version of the matroid to index coding mapping.

Theorem 5.4. *A matroid $M = (E, r)$ with $|E| = n$ is under-representable in d dimensions over a finite field $\mathbb{F}$ if and only if $\lambda_1^{\mathbb{F}}(G_M) \leq n - d$. In particular, if M is representable over $\mathbb{F}$ then $\lambda^{\mathbb{F}}(G_M) = \beta(G_M) = n - r(E)$.*

Proof: Let R be a $d \times n$ matrix which under-represents M in d dimensions over $\mathbb{F}$. Let Q be an $(n-d) \times n$ matrix whose rows span the kernel of R. We will show that Q is a valid encoding matrix for G_M. Let $y \in \mathbb{F}^E$ be some input message set and consider a receiver (x, S), who wishes to decode y_x from $\{y_z : z \in S\}$ and the broadcast message Qy. Extend $\ker(Q)$ arbitrarily into a basis B for $\mathbb{F}^E$ and let $y = y' + y''$ be the unique decomposition according to B such that $y' \in \ker(Q)$. Clearly, $Qy'' = Qy$ since $y' \in \ker(Q)$, hence one can recover y'' from the public channel by triangulating Q. It remains for the receiver (x, S) to recover y'_x. To this end, observe that the rows of R span $\ker(Q)$ and recall that by Definitions 5.1 and 5.3, column x of R is a linear combination of the columns of R indexed by S. Since y' is in the row-space of R it follows that y'_x is equal to the exact same linear combination of the components of y' indexed by S, all of which are known to the receiver. Altogether, the receiver can recover both y'_x and y''_x and obtain the message x. As this holds for any receiver, we conclude that Q is a valid encoding matrix and thus $\lambda_1^{\mathbb{F}}(G_M) \leq n - d$. When $d = r(E)$ the inequality is tight because this upper bound coincides with the lower bound given by Proposition 5.2.

Conversely, suppose that there exists a scalar linear code for G_M over $\mathbb{F}$ with rate $n-d$, and let Q be a corresponding $(n-d) \times n$ encoding matrix of rank $n-d$. Let R be a $d \times n$ matrix whose rows span the kernel of Q. We claim that R under-represents M. Indeed, consider a receiver (x, S). It is easy to verify that this receiver has a linear decoding function of the form $u^{\mathsf{T}} \cdot Qy + v^{\mathsf{T}} \cdot y_S$ for some vectors u, v, where y_S is the vector formed by restricting y to the indices of S. As Q is a valid encoding matrix for G_M, this evaluates to y_x for any $y \in \mathbb{F}^E$. In particular, if y^{T} is a row of R then $Qy = 0$ and so $v^{\mathsf{T}} \cdot y_S = y_x$, and applying this argument to every row of R verifies that column x of R is a linear combination of the columns of R indexed by S (with coefficients from v). Since this holds for any receiver we have that R under-represents M, as required. ∎

We conclude this section with a result that will be useful in establishing lower bounds on the value of the LP for G_M with a given constraint matrix A.

Theorem 5.5. *Suppose that $M = (E, r)$ is a matroid and A is a matrix such that $A\mathbf{1} = 0$ and $A\vec{\mathbf{r}}(M) \not\geq 0$. If the linear program in Figure 1 is instantiated with constraint matrix A, then the value of the LP is strictly greater than $|E| - r(E)$.*

Proof: We will give a dual solution (w, x, y) to the LP with value strictly greater than $|E| - r(E)$.

Recalling the hypothesis $A\vec{\mathbf{r}}(M) \not\geq 0$, let q be a row of A such that $\sum_{S\subseteq E} a_{qS} r(S) < 0$. Let $\mathcal{S}^+ = \{S \subseteq E \mid a_{qS} > 0,\ S \neq E, \emptyset\}$ and $\mathcal{S}^- = \{S \subseteq E \mid a_{qS} < 0,\ S \neq E, \emptyset\}$. Note that the hypothesis that $A\mathbf{1} = 0$ implies that $a_{q\emptyset} + \sum_{S\in\mathcal{S}^+} a_{qS} = -\left(a_{qE} + \sum_{S\in\mathcal{S}^-} a_{qS}\right)$. Assume that A is scaled so $a_{q\emptyset} + \sum_{S\in\mathcal{S}^+} a_{qS} = -\left(a_{qE} + \sum_{S\in\mathcal{S}^-} a_{qS}\right) = 1$. This assumption is without loss of generality since $a_{qE} + \sum_{S\in\mathcal{S}^-} a_{qS}$ is strictly negative, as can be seen from:

$$\begin{aligned}
&r(E)\left(a_{qE} + \sum_{S\in\mathcal{S}^-} a_{qS}\right) \\
&\leq a_{qE} r(E) + \sum_{S\in\mathcal{S}^-} a_{qS} r(S) \\
&\leq a_{qE} r(E) + \sum_{S\in\mathcal{S}^-} a_{qS} r(S) + \sum_{S\in\mathcal{S}^+} a_{qS} r(S) \\
&= \sum_{S} a_{qS} r(S) \ < \ 0\,.
\end{aligned}$$

Define the dual vector y by setting $y_q = 1$ and $y_{q'} = 0$ for rows $q' \neq q$ of A. To define the dual vector x, let us first associate to every set $S \subseteq E$ a matroid basis $b(S)$ such that the set $m(S) = b(S) \cap S$ is a maximal independent subset of S, i.e. $|m(S)| = r(m(S)) = r(S)$. Let $u(S) = S \cup b(S)$. For every $S \in \mathcal{S}^+$, let $x_{\emptyset m(S)} = x_{m(S)S} = a_{qS}$ and for every $S \in \mathcal{S}^-$, let $x_{Su(S)} = x_{u(S)E} = -a_{qS}$. Set all other values of x_{ST} to zero. Finally, set $w = 1$. By construction, (w, x, y) satisfies all of the dual constraints. Using the relations $c_{\emptyset m(S)} = r(S)$, $c_{Su(S)} = r(E) - r(S)$, $c_{m(S)S} = c_{u(S)E} = 0$, we find that the dual LP objective value is

$$\begin{aligned}
&|E|\, w - \sum_{S\subset T} c_{ST} x_{ST} \\
&= |E| - \sum_{S\in\mathcal{S}^+} (c_{\emptyset m(S)} + c_{m(S)S}) a_{qS} \\
&\qquad - \sum_{S\in\mathcal{S}^-} (c_{Su(S)} + c_{u(S)E})(-a_{qS}) \\
&= |E| - \sum_{S\in\mathcal{S}^+} r(S) a_{qS} + \sum_{S\in\mathcal{S}^-} (r(E) - r(S)) a_{qS} \\
&= |E| + \sum_{S\in\mathcal{S}^-} a_{qS} r(E) - \sum_{S} a_{qS} r(S) + a_{q\emptyset} r(\emptyset) + a_{qE} r(E) \\
&= |E| - r(E) - \sum_{S} a_{qS} r(S).
\end{aligned}$$

By hypothesis $\sum_S a_{qS} r(S) < 0$, as required. ∎

6. Separation between linear and non-linear rates

In this section we sketch the proof of Theorem 1.2. To this end we will first show that the linear rate over a field of even characteristic is strictly better than the linear rate over a field of odd characteristic for the index coding problem associated to the Fano matroid, and that the reverse relation holds for the non-Fano matroid. Then we will take the lexicographic product of the two index codes to get a gap between the linear and non-linear coding rates, and then use lexicographic products again to amplify that gap.

The *Fano matroid*, denoted $\mathcal{F}$, and the *non-Fano matroid*, denoted $\mathcal{N}$, are 7 element, rank 3 matroids. The seven columns of the matrix $\left(\begin{smallmatrix} 1&0&0&0&1&1&1 \\ 0&1&0&1&0&1&1 \\ 0&0&1&1&1&0&1 \end{smallmatrix}\right)$ constitute a linear representation of the Fano matroid when $\mathrm{char}(\mathbb{F}) = 2$ and one for the non-Fano matroid when $\mathrm{char}(\mathbb{F}) \neq 2$.

It is well known that the Fano matroid is representable only in a field of characteristic 2, and the non-Fano matroid is representable in any field whose characteristic is different from 2 but not in fields of characteristic 2. We use a generalization of this fact to obtain the following theorem that directly implies Theorem 1.2. All proofs were omitted from this section for space reasons; see [6] for full proofs.

Theorem 6.1 (Separation Theorem)**.** *Let $G = G_{\mathcal{F}} \bullet G_{\mathcal{N}}$. There exists some $\varepsilon > 0$ such that $\beta(G^{\bullet n}) = 16^n$ whereas $\lambda(G^{\bullet n}) \geq (16+\varepsilon)^n$ for all n.*

The fact that $\beta(G^{\bullet n}) = 16^n$ will be a straightforward application of Proposition 5.2 and Theorem 5.4. The lower bound on the linear rate however will require considerably more effort. In order to bound λ from below we will extend the LP $\mathfrak{B}$ to two LPs, one of which will be a lower bound for linear codes over fields with odd characteristic and the other for linear codes over even characteristic. Each one will supplement the matrix A in the LP with a set of constraints, one set derived from dimension inequalities based on the representation of the Fano matroid and the other from the non-Fano matroid. The LP that gives a lower bound for linear codes over a field with even characteristic will be used to show that the linear broadcast rate of $G_{\mathcal{N}}$ over a field of even characteristic is strictly greater than four, and the LP for odd characteristic will imply the corresponding result for $G_{\mathcal{F}}$. Furthermore, the constraints will satisfy the conditions of Theorem 3.4. Putting this all together implies that when we take the lexicographic product of the Fano and non-Fano index coding problems, no linear code is as good as one that combines linear codes over $\mathbb{F}_2$ and $\mathbb{F}_3$.

If $\{V_i\}_{i\in I}$ are subspaces of a vector space V,n let $\vec{\mathbf{d}}(\{V_i\}_{i\in I})$ be a $2^{|I|}$ dimensional vector indexed by the subsets of I such that the coordinate indexed by S is the dimension of the span of $\{V_i\}_{i\in S}$. The following theorem provides the ingredients for proving that $\lambda^{\mathbb{F}}(G_{\mathcal{N}}) > 4$ when $\mathbb{F}$ is a finite field of even characteristic, whereas $\lambda^{\mathbb{F}}(G_{\mathcal{F}}) > 4$ when $\mathbb{F}$ is of odd characteristic.

Theorem 6.2. *There exist a pair of 2^7-dimensional vectors Λ_{even}, Λ_{odd} such that for any 7 subspaces $\{V_i\}_{i\in\mathcal{U}}$ of a vector space over a field $\mathbb{F}$, we have*

$$0 \leq \begin{cases} \Lambda_{\mathrm{even}} \cdot \vec{\mathbf{d}}(\{V_i\}_{i\in\mathcal{U}}) & \text{if } \mathrm{char}(\mathbb{F}) = 2 \\ \Lambda_{\mathrm{odd}} \cdot \vec{\mathbf{d}}(\{V_i\}_{i\in\mathcal{U}}) & \text{if } \mathrm{char}(\mathbb{F}) \neq 2. \end{cases}$$

However, $\Lambda_{\mathrm{even}} \cdot \vec{\mathbf{r}}(\mathcal{N})$ and $\Lambda_{\mathrm{odd}} \cdot \vec{\mathbf{r}}(\mathcal{F})$ are strictly negative.

The proof begins with a quantitative strengthening of the non-representability of $\mathcal{N}$ in characteristic 2: we specify a sequence of linear inequalities satisfied by $\vec{\mathbf{r}}(\mathcal{N})$ such that, if $\{V_i\}_{i\in\mathcal{U}}$ are d-dimensional and $\vec{\mathbf{d}}(\{V_i\}_{i\in\mathcal{U}})$ satisfies all but the last of the inequalities, then it violates the last one by d additively. Then we show how to transform a general 7-tuple of subspaces into one satisfying all but the last of the inequalities. The vector Λ_{even} expresses the negation of the final linear inequality in the sequence, combined with correction terms that reflect how the transformation modifies the subspace dimensions. The case $\text{char}(\mathbb{F}) \neq 2$ is handled similarly.

The following pair of lemmas shows how to take a single linear constraint, such as one of those whose existence is asserted by Theorem 6.2, and transform it into a tight homomorphic constraint schema. To state the lemmas, we must first define the set of vectors $D_{\mathbb{F}}(K) \subset \mathbb{R}^{\mathcal{P}(K)}$, for any index set K and field $\mathbb{F}$, to be the set of all vectors $\vec{\mathbf{d}}(\{V_k\}_{k\in K})$, where $\{V_k\}_{k\in K}$ runs through all K-indexed tuples of finite-dimensional vector spaces over $\mathbb{F}$.

Lemma 6.3 (Tightening Modification)**.** *Suppose I is any index set, e is an element not in I, and $J = I \cup \{e\}$. There exists an explicit linear transformation from $\mathbb{R}^{\mathcal{P}(J)}$ to $\mathbb{R}^{\mathcal{P}(I)}$, represented by a matrix B, such that:*

(i) $B \cdot D_{\mathbb{F}}(J) \subseteq D_{\mathbb{F}}(I)$ *for every field* $\mathbb{F}$.

(ii) $B\mathbf{1} = B\mathbf{1}_j = 0$ *for all* $j \in J$.

(iii) If M is a matroid with ground set I and the intersection of all matroid bases of M is the empty set, then $B\vec{\mathbf{r}}(M+e) = \vec{\mathbf{r}}(M)$, where $M+e$ denotes the matroid obtained by adjoining a rank-zero element to M.

Lemma 6.4 (Homomorphic Schema Extension)**.** *Let I be an index set, and let $\vec{\alpha} \in \mathbb{R}^{\mathcal{P}(I)}$ be a vector such that $\vec{\alpha}^{\mathsf{T}}\vec{\mathbf{d}} \geq 0$ for all $\vec{\mathbf{d}} \in D_{\mathbb{F}}(I)$. Then there is a homomorphic constraint schema (Q, A) such that $\vec{\alpha}^{\mathsf{T}}$ is a row of the matrix $A(I)$, and for every index set K and vector $\vec{\mathbf{d}} \in D_{\mathbb{F}}(K)$, $A(K)\vec{\mathbf{d}} \geq 0$. If $\vec{\alpha}^{\mathsf{T}}\mathbf{1} = \vec{\alpha}^{\mathsf{T}}\mathbf{1}_i = 0$ for all $i \in I$, then the constraint schema (Q, A) is tight.*

Finally, it will be useful to describe the following simple operation for combining constraint schemas.

Definition 6.5. The *disjoint union* of two constraint schemas (Q_1, A_1) and (Q_2, A_2) is the constraint schema which associates to every index set I the disjoint union $\mathcal{Q}(I) = \mathcal{Q}_1(I) \sqcup \mathcal{Q}_2(I)$ and the constraint matrix $A(I)$ given by

$$A(I)_{qS} = \begin{cases} A_1(I)_{qS} & \text{if } q \in \mathcal{Q}_1(I) \\ A_2(I)_{qS} & \text{if } q \in \mathcal{Q}_2(I). \end{cases}$$

For a homomorphism $h : \mathcal{P}(I) \to \mathcal{P}(J)$, the function $h_* : \mathcal{Q}(I) \to \mathcal{Q}(J)$ is defined by combining $\mathcal{Q}_1(I) \xrightarrow{h_*} \mathcal{Q}_1(J)$ and $\mathcal{Q}_2(I) \xrightarrow{h_*} \mathcal{Q}_2(J)$ in the obvious way.

Lemma 6.6. *The disjoint union of two tight constraint schemas is tight, and the disjoint union of two homomorphic constraint schemas is homomorphic.*

Theorem 6.1 now follows from combining these results with those of earlier sections. We transform each of $\Lambda_{\text{even}}, \Lambda_{\text{odd}}$ into a tight homomorphic constraint schema using Lemmas 6.3 and 6.4, and we take the disjoint union of each of these with the submodularity schema. Using the resulting pair of linear programs in Theorem 5.4, we obtain $\lambda(G) \geq 16 + \varepsilon$ for some $\varepsilon > 0$. Amplifying this gap using Theorem 3.4 yields the lower bound $\lambda(G^{\bullet n}) > (16+\varepsilon)^n$.

References

[1] M. Adler, N. J. A. Harvey, K. Jain, R. Kleinberg, and A. R. Lehman, "On the capacity of information networks," in *Proc. of the 17th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2006)*, 2006, pp. 241–250.

[2] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inform. Theory*, vol. 46, no. 4, pp. 1204–1216, 2000.

[3] N. Alon, A. Hassidim, E. Lubetzky, U. Stav, and A. Weinstein, "Broadcasting with side information," in *Proc. of the 49th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2008)*, 2008, pp. 823–832.

[4] Z. Bar-Yossef, Y. Birk, T. S. Jayram, and T. Kol, "Index coding with side information," in *Proc. of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006)*, 2006, pp. 197–206.

[5] Y. Birk and T. Kol, "Coding-on-demand by an informed source (ISCOD) for efficient broadcast of different supplemental data to caching clients," *IEEE Trans. Inform. Theory*, vol. 52, pp. 2825–2830, 2006.

[6] A. Blasiak, R. Kleinberg, and E. Lubetzky. Lexicographic products and the power of non-linear network coding. arXiv:1108.2489 [cs.IT].

[7] R. Dougherty, C. Freiling, and K. Zeger, "Insufficiency of linear coding in network information flow," *IEEE Trans. Inform. Theory*, vol. 51, no. 8, pp. 2745–2759, 2005.

[8] ——, "Networks, matroids, and non-Shannon information inequalities," *IEEE Trans. Inform. Theory*, vol. 53, no. 6, pp. 1949–1969, 2007.

[9] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Médard, and J. Crowcroft, "Xors in the air: Practical wireless network coding," *IEEE/ACM Trans. on Networking*, vol. 16, no. 3, pp. 497–510, 2008.

[10] E. Lubetzky and U. Stav, "Non-linear index coding outperforming the linear optimum," *IEEE Trans. Inform. Theory*, vol. 55, pp. 3544–3551, 2009.

[11] S. E. Rouayheb, A. Sprintson, and C. Georghiades, "On the relation between the index coding and the network coding problems," in *IEEE International Symposium on Information Theory (ISIT 2008)*, 2008, pp. 1823–1827.

[12] ——, "A new construction method for networks from matroids," in *IEEE International Symposium on Information Theory (ISIT 2009)*. Piscataway, NJ, USA: IEEE Press, 2009, pp. 2872–2876.

[13] R. W. Yeung, S.-Y. R. Li, and N. Cai, *Network coding theory*. Now Publishers Inc, 2006.