Entropy Region and Network Information Theory

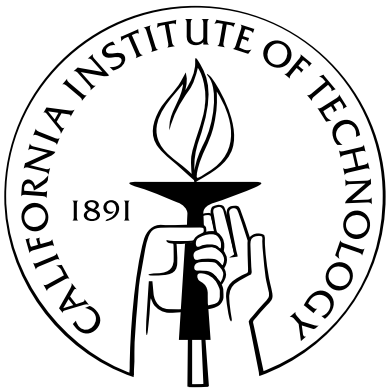
Thesis by

Sormeh Shadbakht

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



California Institute of Technology

Pasadena, California

2011

(Defended May 16, 2011)

To my family

Acknowledgements

I am deeply grateful to my advisor Professor Babak Hassibi, without whom this dissertation would have not been possible. He has never hesitated to offer his support and his guidance and mindful comments, during my Ph.D. years at Caltech, have indeed been crucial. I have always been inspired by his profound knowledge, unparalleled insight, and passion for science, and have enjoyed countless hours of intriguing discussions with him. I feel privileged to have had the opportunity to work with him.

My sincere gratitude goes to Professors Jehoshua Bruck, Michelle Effros, Tracey Ho, Robert J. McEliece, P. P. Vaidyanathan, and Adam Wierman for being on my candidacy and Ph.D. examination committees, and providing invaluable feedback to me. I would like to thank Professor Alex Grant with whom I had the opportunity of discussion during his visit to Caltech in Winter 2009 and Professor John Walsh for interesting discussions and kind advice. I am also grateful to Professor Salman Avestimehr for the worthwhile collaborations—some instances of which are reflected in this thesis—and Professor Alex Dimakis for several occasions of brainstorming. I am further thankful to David Fong, Amin Jafarian, and Matthew Thill, with whom I have had pleasant collaborations; parts of this dissertation are joint works with them.

My warmest thanks go to Shirley Slattery for her amiable assistance in administrative matters and beyond. I am also indebted to all my friends and colleagues who made my life more pleasant during my graduate studies. In particular, I would like to express my gratitude to my former and current group-mates in the Systems Lab, Moore 155, who created a nice environment that made long hours of working more

joyful. I would like to thank my dear friend, Ali Vakili, for his unlimited friendship and support. To him, my gratitude is beyond words.

Last but not least, my deepest appreciations go to my family for their endless love, care and encouragement. I am sincerely thankful to my parents, Farideh Ghavidel and Bahram Shadbakht, who have unconditionally put forth anything that they could for my success and progress. Their love, care and support have been boundlessly heartwarming and their appreciation for education and science inspiring. My mother's presence here has been a blessing for me, and my father's encouragements have always helped me pursue my goals decisively. I wholeheartedly thank my dear sister, Bahareh Shadbakht, for her unreserved kindness and being my wonderful friend and support. She and her husband, Sohrab Zamani, have been of great help to me especially during my first year of stay in the U.S. I am greatly thankful to them and I have always enjoyed spending time with them and my lovely nieces, Yasmin and Hannah. To my family, for all that they have done for me, I shall always be grateful.

Sormeh Shadbakht

Caltech

May 2011

Abstract

The ever-growing need for data transmission over networks calls for optimal design of systems and efficient use of coding schemes over networks. However, despite the attention given to the analysis of networks and their optimum performance, many issues remain unsolved. An important subject of study in any network is its capacity region, which is defined through the limits of the set of data rates at which the sources can *reliably* communicate with their corresponding destinations. Although the capacity of a single user communication channel is completely known, the capacity region of many multiuser information theory problems are still open. A main hurdle in obtaining the capacity of multiuser networks is that the problem is usually nonconvex and it involves limiting expressions in terms of the number of channel uses (data transmissions). This thesis takes a step toward a *general* framework for solving network information theory problems by studying the capacity region of networks through the *entropy region*. An entropy vector of n random variables with a fixed probability distribution is the vector of all their joint entropies. The entropy region of n random variables accordingly is the space of all entropy vectors that can arise from different probability distributions of those random variables.

In the first part, it is shown that the capacity of a large class of acyclic memoryless multiuser information theory problems can be formulated as *convex* optimization over the region of entropy vectors of the network random variables. The advantage of this characterization over the previous approaches, beside its universality, is that the capacity optimization need not be performed over the limit of an infinite number of

channel uses. This formulation on the other hand reveals the fundamental role of the entropy region in determining the capacity of network information theory problems.

With this viewpoint, the rest of the thesis is dedicated to the study of the entropy region and its consequences for networks. A full characterization of the entropy region has proven to be a very challenging problem and so we have mostly examined the space of entropy vectors through inner bound constructions. For discrete random variables our approaches include the characterization of entropy vectors with a lattice-derived probability distribution, the entropy region of binary random variables, and the linear representable region roughly defined as the entropy region of linearly related random variables over a finite field. Our lattice-based construction can in principle be generalized to any number of random variables and we have explicitly computed its resulting entropy region for up to 5 random variables. The entropy region of binary random variables and the linear representable vectors are mostly considered in the context of the linear coding capacity of networks. In particular, we formulate the binary scalar linear coding capacity of networks and give the necessary and sufficient conditions for its set of solutions. Moreover, we also obtain similar necessary and sufficient conditions in the case of linearly coded arbitrary alphabet-size scalar random variables of a network with 2 sources and determine the optimality of linear codes among all scalar codes for such a network.

For continuous random variables we have studied the entropy region of jointly Gaussian random variables and have determined that the convex cone of the corresponding region of 3 Gaussian random variables obtains the entropy region of 3 continuous random variables in general. For more than 3 random variables we point out the set of minimal necessary and sufficient conditions for a vector to be an entropy vector of jointly Gaussian random variables.

Finally in the absence of a full analytical characterization of the entropy region, it is desirable to be able to perform numerical optimization over this space. In this

regard, a certain Monte Carlo method is proposed that enables one to numerically optimize the achievable rates in an arbitrary wired network under linear or nonlinear network coding schemes. This method can be adjusted for decentralized operation of networks and can also be used for optimization of any function of entropies of discrete random variables. The promise of this technique is shown through various simulations of several interesting network problems.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The growing interest in information transmission over networks in recent years has encouraged optimal design of communication networks. Although since its birth in the late 1940s, information theory has had a significant role in the development and improvement of point-to-point communication systems, it is fair to say that it has had far less impact on the design of most of the networks currently in use, and especially on the Internet. The importance of an interaction between networking and information theory has become more apparent as the current networking tools have been recognized to be inadequate for addressing the challenges of, e.g., the mobile ad hoc wireless networks, and as network coding has proven to be advantageous over traditional routing [ACLY00].

One of the main difficulties of incorporating information theory in the design of networks is that when it comes to multi-user information theory, the capacity region (the rate region for a reliable information transfer) of even some of the simplest networks, such as the relay channel, remain unsolved. In fact characterizing the capacity region of most network information theory problems requires one to solve an infinite-letter nonconvex optimization problem which is an almost impossible task to do [vdM77, Sha61]. Hence most multi-user problems have been tackled individually through some network-specific subtle techniques. This is in contrast to traditional networking, where the multi-commodity flow viewpoint allows one to obtain the op-

timal rates in an arbitrary setting by solving a linear program over the network. The lack of such an optimization-based approach within the information theory framework that can compute and realize the achievable rates can be considered a major reason for the minimal interaction between networking and information theory. Developing such methods, has been the main motivation for the current thesis.

An "*entropy vector*" of n random variables with a specific joint distribution is defined as the vector of all their 2^n-1 joint entropies. Accordingly, the "*entropy region*" is identified as the space of all such entropy vectors and is denoted by Γ_n^* [YLCZ06]. As a step toward solving network information theory problems via a general framework, we have developed a new optimization formulation for obtaining the achievable rates in an arbitrary network. We show that by using the notion of entropy region the optimal rates can be computed through a *convex* optimization problem. This formulation of the capacity region, not only does away with the infinite-letter and nonconvexity of previous characterizations, but also reveals the fundamental role of the entropy region in determining the capacity region of network information theory problems. Notably, for wired networks due to the separation of channel and network coding [KEM09], to determine the rate region one only needs to characterize the (unconstrained) entropy region. For wireless networks on the other hand, due to the broadcast and interference nature of these channels, study of the network-constrained entropy region is required. Study of the information inequalities which involve sums of random variables such as the entropy-power inequality is particularly important.

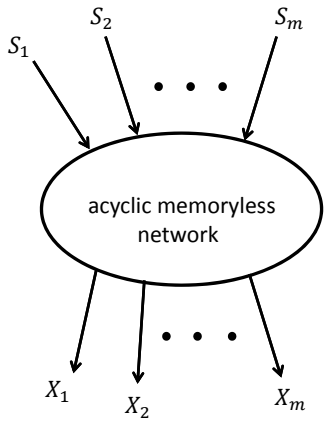
The full characterization of the entropy region has proven to be a formidable task. Therefore with an eye toward solving the obtained network optimization problem, this thesis studies new inner bound constructions of the entropy region and the relevant consequences for network coding. Numerical optimization over the entropy region as an alternative is investigated as well.

1.1 Capacity Region in Multiuser Information Theory

Multi-user information theory involves the study of the limits of information transfer among many users. However in contrast to the point-to-point communication (single-user) case, where the problem is well understood, many multiuser problems are still open. This is mainly due to the fact that a general approach to evaluate the capacity region of multi-user problems requires one to optimize a linear combination of mutual information terms over all possible joint distributions of source variables of the network and all feasible network operations while letting the number of channel uses go to infinity. In this formulation the objective function is usually nonconvex in the joint probability distribution of sources and also the network operations. Moreover considering infinite number of channel uses means that one should consider a sequence of random variables or equivalently vector-valued random variables whose lengths are growing unboundedly. This is referred to as "infinite-letter". Altogether, one has to solve an infinite-letter nonconvex optimization problem.

This method is extremely difficult and although it can theoretically express the capacity region it has rarely been used for the computation of the achievable rates [CV93]. A few networks whose capacity regions are known are the cases for which an equivalent single-letter characterization has been found. In particular the capacity of the memoryless point-to-point communication channel can be expressed via a single-letter convex optimization and is therefore completely solved.

In traditional networking, a.k.a, the multi-commodity viewpoint, on the other hand, considering the information as flows allows one to obtain optimal rates via solving a linear program, subject to the conservation of flows at each node and that the total flow on each edge of the network not exceed the capacity of that edge. Therefore one might wonder if there exists a similar framework for solving network



$$\lim_{T\to\infty} \sup_{p_{S_i{}^T} \text{ and network operations}} \sum_{i=1}^{m} \alpha_i \frac{1}{T}\left(H\left(X_i{}^T\right) - H\left(X_i{}^T|S_i{}^T\right)\right)$$

Figure 1.1: Determining the rate region of a memoryless acyclic network involves an infinite letter characterization.

information theory problems in general.

This issue is addressed in Chapter 2 where we have shown that the a large class of acyclic memoryless networks can be formulated as convex optimization over the region of entropy vectors. This formulation avoids the infinite-letter characterization and reveals the fundamental region that needs to be characterized: The entropy region. Moreover it suggests that similar to the traditional networking where distributively solving the network problem made algorithms such as TCP-IP possible, distributive solutions to our proposed convex optimization may also lead to effective protocols. These issues are dealt with in subsequent chapters.

1.2 Entropy Region Characterization

Characterizing the entropy region of any number of random variables has long been an interesting open problem. In fact in addition to its central role in determining the capacity of networks, it is closely related to many other issues in information theory and statistics.

A linear combination of the joint entropies of n random variables which is positive for all the entropy vectors in the Γ_n^* is referred to as a *linear information inequality* for the entropies. Linear information inequalities which follow from the positivity of conditional mutual information are known as *Shannon-type* inequalities [ZY98]. Although for up to 3 random variables the entropy region is completely characterized by a finite set of Shannon-type inequalities, the full characterization of the region for 4 or more number of random variables involves *non-Shannon information inequalities* [ZY97, MMRV02, Zha03, DFZ06a] and remains a challenging problem. In fact it is proven that no finite set of linear inequalities can completely characterize Γ_n^* for $n \geq 4$ [Mat07a]. In other words the region is not a polytope for $n \geq 4$, in spite of the fact that the closure of the entropy region is known to be a convex cone for all n. In summary for $n \geq 4$, only partial characterization of the entropy region through inner or outer bounds, exist. From the network problem perspective, inner bounds of the entropy region are interesting in that they yield achievable rates. Yet, an approach that can be easily extended to any number of random variables for obtaining an inner bound is missing. This thesis takes a step in this direction by constructing an achievable entropy region through different methods.

Discrete Random Variables

While the discovery of new families of non-Shannon-type inequalities has lead to new outer bounds, the most well-known inner bound for the entropy region is the

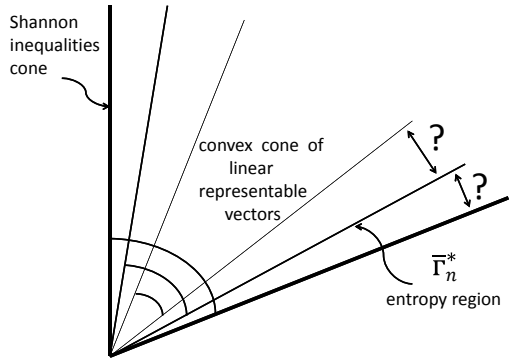


Figure 1.2: Entropy region, cone of Shannon inequalities, and the linear representable region

so-called *linear representable region* [YLCZ06] which is the entropy region of a set of random variables defined over a finite field that are obtained via a linear matrix transformation from another set of uniformly distributed random variables over that finite field. Although proven to be a strict subset of the entropy region, the general characterization of this inner bound also remains an open problem.

In summary, there exists no general method for creating inner bounds for the entropy region. This problem is addressed in Chapters 3 and 4 where in Chapter 3 a systematic inner bound construction of the entropy region is proposed and in Chapter 4 characterizing the linear representable region under simplifying constraints is investigated.

Continuous Random Variables

Although most of the research regarding the characterization of the entropy region has been focused toward discrete random variables, it turns out that there is a close connection between the entropy region of discrete and that of continuous random variables such that one can be computed from the other [Cha03]. Nevertheless the entropy region of continuous random variables or in particular the entropy region of

a *family* of continuous probability distributions has not been studied independently. This issue is dealt with in Chapter 5 where the entropy region of jointly Gaussian random variables is studied. Gaussian random variables demonstrate some interesting properties which make them a potentially good candidate for consideration. As an example they easily violate the best known inner bound for the entropy region of 4 random variables.

1.3 Linear Coding and Linear Representability

Linear network codes, although known to be suboptimal for achieving capacity in an arbitrary network, are appealing due to their simple structure. They turn out to be intimately related to the *linear representable* entropy region which is the most acclaimed inner bound of Γ_n^*. Therefore from the linear network coding perspective characterizing the region of linear representable entropy vectors is an important topic of study. Moreover this problem has connections with the matroid representability subject which makes it an interesting problem even on its own.

Despite some recent advances in this area [DFZ10, CGK10], the general characterization of the linear representable region is very complex and remains an open problem. As a result, full analysis of network problems even under the linear coding assumption seems to be far from reach. However it turns out that focusing on networks with a fixed number of sources or assuming linear network codes over a certain finite field makes the problem much more tractable [SJH09]. This subject is addressed in Chapter 4 where linear codes for networks with two sources are studied. In particular, we determine that among all scalar codes for networks with two sources, linear codes are optimal. Furthermore, binary linear codes are also investigated for general networks, and the method to obtain the capacity of networks under such coding schemes is presented.

1.4 Numerical Optimization Over the Entropy Region

The problem of characterizing the entropy region is very challenging and its analysis will be a topic of research for years to come. In the absence of such an explicit characterization, the next best thing is to present a method to numerically perform optimization over this region. Nonetheless any deterministic numerical optimization over this region would require some knowledge about the boundaries of the entropy space. Therefore stochastic optimization techniques such as Markov Chain Monte Carlo seem to be more suitable. For all that a Markov Chain Monte Carlo (MCMC) method calls for sampling from Γ_n^*. This issue is dealt with in Chapter 6 where an MCMC approach is presented that allows numerical optimization of any desired accuracy over the entropy region by doing a random walk on quasi-uniform distributions. These distributions, as defined by Chan in [Cha01], are a set of joint distributions of n random variables which take on either zero or a constant value for all marginals and they turn out to be sufficient for characterizing the entropy region.

1.5 Scope and Contributions of the Thesis

In this section we review the contents of each chapter and mention the main results. Chapters may be read more or less independently.

In Chapter 2, we propose a convex formulation of the capacity region of a general class of acyclic memoryless network information theory problems. This scheme is advantageous over the previous characterizations in the sense that the infinite-letter and nonconvexity dilemmas are no longer present. Our method is based on a slightly different notion of entropy vectors, i.e., *normalized entropy vectors*—which we define as the entropy vector normalized by the log of the alphabet size of the underlying dis-

tribution. We denote the corresponding *normalized entropy region* by Ω_n^*, as opposed to the region of non-normalized entropy vectors, which is denoted by Γ_n^*. We show that this definition is more natural as the cost function in capacity characterization of a multiuser information theory problem is a linear combination of joint entropies divided by the number of channel uses and therefore proportional to the linear combination of normalized entropies. Hence we show that the network capacity problem reduces to an optimization of a linear function of normalized entropies over the *network-constrained entropy region*, which is essentially Ω_n^* constrained by the network conditions which either follow from the topology of the network or are imposed via the channel constraints. By proving the convexity of the network-constrained entropy region we prove the convexity of this optimization formulation. For wired networks the formulation simplifies to a convex optimization over the unconstrained entropy region, and as a result determining the capacity of wired networks only requires the characterization Ω_n^*. Moreover we show how using the optimization machinery in this framework can bring forth interesting results. We obtain cutset outerbounds via a duality argument as an example.

In Chapter 3, we study the entropy region of discrete random variables and present an inner bound construction that is in principle generalizable to any number of random variables. By using lattices as generator of points in the Euclidean space $\mathbb{R}^{2^n-1}$ we construct an inner region for Ω_n^*. Our method hinges on defining a uniform probability on lattice points inside a hypercube of $\mathbb{R}^{2^n-1}$ and offers a systematic method for constructing inner bounds for any number of random variables. Moreover the obtained innerbound is a polytope which is specially desired from the network information theory viewpoint, as such innerbounds render the network problem as a simple linear program. We have explicitly calculated this region for up to $n = 5$ random variables, have shown its tightness for $n \leq 3$ and have proved its equivalence with the linear representable region for $n = 4$. In general we have established that due

to the connection of our construction with *Abelian groups*, the lattice-derived inner region will always satisfy the known *Ingleton* inequality; an inequality that is valid for a strict subset of the entropy region points. We also study the entropy region of binary quasi-uniform random variables and make comparisons with the lattice-derived entropy region. Studying the entropy region of binary random variables is an interesting problem which has been a subject of research in the community as well [WW09].

In Chapter 4, we focus on the linear representable entropy region. In particular we study the scalar linear representable region (i.e., the linear representable vectors whose underlying random variables are scalar valued) in a systematic fashion and explicitly compute the region for 4 random variables. We then turn our attention to networks with 2 sources and show the optimality of linear codes among all scalar codes for the network. We explicitly compute the entropy region of linearly encoded random variables of a network with 2 sources and maximum number of variables of 6. We also study linear network codes over binary operations, which essentially is the case where network random variables are binary and the nodes of the network either route the packets they receive, or combine them using XORs, or timeshare between these two operations. We then give the necessary and sufficient conditions for an entropy vector to correspond to a solution of such network and accordingly formulate the capacity region of networks under a binary linear coding assumption.

In Chapter 5 we determine the entropy region of 3 jointly Gaussian random variables by finding the structure of the covariance matrices of the boundary points. We show that the closure of the convex cone of this region generates the whole entropy region of 3 continuous random variables. This result is very encouraging and to our knowledge is the first result about the entropy region of any class of continuous distributions. For 4 or more number of Gaussian random variables, the problem is closely related to *Cayley's hyperdeterminant* relation which is a generalization of the

determinant to higher dimensions [HS07b]. We determine $2^n - 1 - \frac{n(n+1)}{2}$ nonlinear constraints as the set of necessary and sufficient conditions for the entries of a given $2^n - 1$ dimensional vector to correspond to the entropy vector of n jointly Gaussian random variables. These necessary and sufficient conditions lay the foundation for determining the whole entropy region of Gaussian random variables for $n \geq 4$ via obtaining the convex cone of the realizable entropy vectors. Finally the entropy region of continuous random variables in the context of the capacity of wireless networks is considered in this chapter and the role of information inequalities such as the entropy power are discussed.

In Chapter 6 we present a method for doing a random walk on quasi-uniform distributions that allows one to numerically stake out the entropy region to any desired accuracy. When coupled with Markov Chain Monte Carlo (MCMC) methods, one may bias the random walk so as to maximize certain functions of the entropy vector in a fashion similar to the Metropolis algorithm. Moreover this method can be used in networks for optimizing a particular function of rates, where the random walk will be over the input-output mappings at the network nodes. In cases where the network is solvable, the obtained mappings yield a network code which could be linear or nonlinear. Finally this approach can be performed in a decentralized fashion in networks. We show the promise of this technique in finding solutions for arbitrary networks and optimization of entropy functions by applying it to different examples. As an instance we have employed this method to find entropy vectors that violate the so-called *Ingleton* bound. This bound is identified by an inequality that does not hold for all entropy vectors, yet only a handful of explicit examples of entropy vectors are known to violate it. By using the MCMC method, we have interestingly discovered entropy vectors that violate this bound meaningfully more than the previously known examples. We have applied this technique to some networks as well. As an instance we have considered a repair problem in a distributed storage system where there are

source and storage nodes and the goal is to find network codes that in the case of failure of a storage node will enable the rest of the network nodes to reliably recover the lost data [CDH]. We have easily found a linear solution for storage problems with 2 and 3 sources and 4 and 5 storage nodes, correspondingly. Moreover we have found nonlinear codes for some of the matroidal networks studied in the literature [DFZ07].

Finally in **Chapter 7** we discuss some open problems in this field and possible future directions.

Chapter 2

Network Information Theory and Convex Optimization

2.1 Introduction

Determining the capacity region of network information theory problems has long been an interesting problem. Nonetheless, as opposed to the point-to-point communication systems where the problem is well understood, the capacity region of many multiuser problems remain open. A simple example is the 3-node relay channel which one may consider to be the immediate extension of the point-to-point channel in which the receiver not only receives information directly from the transmitter but also through a relay. As simple as the setup may sound, the capacity region remains unsolved. Thus far the approaches that have been taken toward solving multiuser problems have been either in the regime of large number of users [GK00] or through development of network-specific techniques. The handful of cases for which the capacity region has been completely determined are the cases where the obtained inner or outer bounds for the capacity region have matched the cutset bounds. All in all, a general theory of multiuser information theory is still lacking.

While, "in principle", it is possible to write down a characterization for the capacity region of most network information theory problems, the difficulty is that this characterization is *infinite-letter* and *nonconvex*. In other words, evaluating the capac-

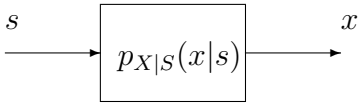


Figure 2.1: A point-to-point communication problem

ity region requires solving an infinite succession of nonconvex optimization problems over certain distributions whose number of variables goes to infinity. This is in stark contrast with point-to-point (single-user) memoryless channels where the characterization is both *single-letter* and *convex*.

To make this more explicit, consider the point-to-point memoryless channel of Fig. 2.1. The capacity is clearly

$$C = \max_{p_S(\cdot)} I(S;X) = \max_{p_S(\cdot)} \{H(X) - H(X|S)\}, \tag{2.1}$$

where $p_S(\cdot)$ is the input distribution and $H(X)$ and $H(X|S) = H(X,S) - H(S)$ are the usual entropy and conditional entropies. Problem (2.1) is referred to as single letter, since all entropies are over only a single channel use. The problem is one of "convex optimization" since $I(X;Y)$ is a concave function of the input distribution and so we are maximizing a concave function.

Consider now the network problem of Figure 2.2. Assume that the network is acyclic and memoryless (in the sense that all channels internal to the network are memoryless) and that there is no feedback from the destinations to the sources. Assume that each source S_i needs to transmit to its corresponding destination X_i

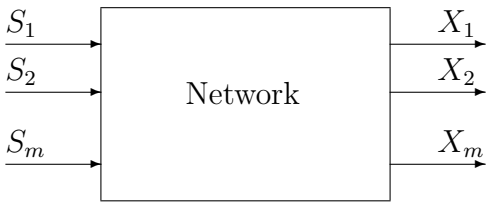


Figure 2.2: A communication problem over an acyclic memoryless network

at some rate $R_i, i = 1, \ldots, m$. Note that this assumption can allow for general setups; if a source is desired by many destinations, then we may repeat that source as many times as desired, and if a destination requires many sources, then we may repeat that destination as many times as necessary.

Now in this case it is not too hard to show that the *rate region*, defined as the set of rates $\{R_i\}_{i=1}^m$ which can be reliably exchanged between the sources and destinations, is given by (see, e.g., [Sha61, vdM77, Kra03]):

$$\mathcal{R} = \mathrm{cl}\left\{R_i, i = 1, \ldots, m \mid R_i < \frac{1}{T}\left(H(X_i^T) - H(X_i^T|S_i^T)\right)\right\} \quad \text{as } T \to \infty \tag{2.2}$$

where $\mathrm{cl}\{\cdot\}$ refers to the closure of the set. Here S_i^T and X_i^T are random variables obtained from concatenating the corresponding source and destination random variables over T channel uses. Of course, the characterization of the rate region $\mathcal{R}$ as in (2.2) is not surprising—in some sense it can be considered almost as the definition of the rate region. Computing it, however, is another matter.

An equivalent way of representing the rate region is through its tangent hyperplanes. These can be obtained via solving the following optimization problem

$$\lim_{T\to\infty} \sup_{\substack{p_{S_i^T}(\cdot) \text{ and} \\ \text{network operations}}} \sum_{i=1}^m \alpha_i \frac{1}{T}\left(H(X_i^T) - H(X_i^T|S_i^T)\right) \tag{2.3}$$

where $\{\alpha_i\}_{i=1}^m$ represents the normal vector to the tangent hyperplane, and where "network operations" represents all permissible internal operations of the network. The problem (2.3) is notoriously difficult since it is *infinite-letter* (i.e., it involves optimization over joint distributions whose number of variables goes to infinity) and *nonconvex* (the cost function $\sum_{i=1}^m \alpha_i \frac{1}{T}\left(H(X_i^T) - H(X_i^T|S_i^T)\right)$ is highly nonconvex in the $p_{S_i^T}(\cdot)$ and "network operations"). For this reason, the characterization of (2.3)

has very rarely been explicitly used.[1]

Our goal in this chapter is to suggest an alternative representation and study of the aforementioned network information theory problem. The main idea is to define the space of (suitably normalized) entropic vectors and to show that a very wide range of network information theory problems reduce to the optimization of a linear cost over the convex set of (constrained) entropic vectors. This viewpoint has several advantages: first, it does away with the complications of infinite-letter characterizations (in fact, the infinite limit simplifies the representation considerably), second, it renders the problem convex and, third, it shows how through duality one may obtain classical results such as cutset bounds. While by no means solving the network information theory problem in itself, it does point to what the heart of the problem is: characterizing the space of (channel constrained) entropic vectors.

The next section defines the notion of entropic vectors and shows that the resulting space is convex. This is then used to formulate network information theory problems as convex optimizations.

In Section 2.3.3 cutset bounds, as well as a separation between network coding and channel coding, are studied as some special instances of this formulation.

2.2 Entropy Vectors and Network Information Theory

The notion of entropy has been around for a long time. It is a measure of information and hence it is what the theory of information is based on [Sha48]. Hence there has been a lot of research about its properties and extensions [Fuj78, Han81, CT91]. In particular, for a given set of random variables $X_1, \ldots, X_n$, it has been interesting

[1] The only work that we are aware of that uses the infinite-letter characterization (2.2) is [CV93], which shows that it can be reduced to a single-letter characterization for memoryless multiple-access channels.

to find out the relations between different joint entropies of those random variables [Han75, Yeu91]. In 1997 Yeung [Yeu97] formalized the following definition:

Definition 2.2.1 (Entropy Vector) *[Yeu97] Let $X_1, \ldots, X_n$ be a collection of n jointly distributed discrete random variables with alphabet size N each.*[1] *For any set $\alpha \subseteq \{1, \ldots, n\}$, let $h_\alpha = h(X_i, i \in \alpha)$ denote the joint entropy of the random variables indexed by the subset α. There are $2^n - 1$ such subsets and thus the collection of all h_α forms a $2^n - 1$ dimensional vector which is called an "entropy vector". Conversely any $2^n - 1$ dimensional vector that can be regarded as the entropy vector of n discrete random variables $X_1, \ldots, X_n$ is called "entropic". The region of all entropic vectors of n random variables is denoted by Γ_n^*.*

Now recall that the objective (2.3) is just a linear function of entropies:

$$\sum_{i=1}^{m} \alpha_i \frac{1}{T} \left(H(X_i^T) + H(S_i^T) - H(X_i^T, S_i^T) \right). \tag{2.4}$$

This motivates the following definition:

Definition 2.2.2 (Normalized Entropy) *Let h be the entropy vector of n discrete random variables $X_1, \ldots, X_n$ with alphabet size N. We define $\underline{h} = \frac{1}{\log N} h$ as the "normalized entropy vector". Conversely any $2^n - 1$ dimensional vector that can be considered as the entropy vector of n random variables for some value of alphabet size N is called normalized entropic. We denote the space of all normalized entropic vectors of n random variables by Ω_n^*.*

We remark that the motivation for the definition of the normalized entropy is the fact that what appears in (2.4), i.e., $\frac{1}{T} H(X_i^T)$, $\frac{1}{T} H(S_i^T)$, and $\frac{1}{T} H(X_i^T, S_i^T)$ are essentially normalized entropies, since the alphabet-sizes of S_i^T and X_i^T are exponential

[1] There is no loss of generality in this assumption. If the random variables have different alphabet-sizes, we can always take N to be the largest alphabet-size and to make the probability mass functions zero wherever appropriate.

in T. Therefore our definition [HS07a] of entropic vectors is slightly different from the non-normalized version that is conventionally used in the literature (see, e.g., [Yeu02]). We believe our definition to be more natural. One indication is the more direct connection to (2.3) and (2.4). The other is the fact that the set Γ_n^* is known to be quite complicated: it has an irregular boundary [Mat07a] and many "holes". Its closure, $\bar{\Gamma}_n^*$, is therefore more often studied, which can be shown to be a convex cone [Yeu02, ZY97]. The set Ω_n^* is, however, much simpler. It is clearly bounded, since,

$$\underline{h}_\alpha \leq |\alpha|\,, \quad \forall \alpha \subseteq \{1, \ldots, n\} \tag{2.5}$$

where $|\alpha|$ is the cardinality of the set α. Furthermore, it is straightforward to show that the closure of Ω_n^* is a convex set.

Theorem 2.2.3 (Convexity of $\bar{\Omega}_n^*$) *The closure of the set of normalized entropic vectors, $\bar{\Omega}_n^*$ is convex.*

We will present two proofs, since both are instructive.

Proof 1: (*Time sharing*) Suppose $\underline{h}_x \in \bar{\Omega}_n^*$, corresponding to random variables $X_1, \ldots, X_n$ with alphabet-size N_x and $\underline{h}_y \in \bar{\Omega}_n^*$ corresponding to random variables $Y_1, \ldots, Y_n$ with alphabet-size N_y. Make n_x independent copies of the first set and n_y independent copies of the second so that together the new concatenated random variables have alphabet-size $N_x^{n_x} N_y^{n_y}$. The resulting entropy vector is clearly

$$\frac{n_x \log N_x}{n_x \log N_x + n_y \log N_y} h_x + \frac{n_y \log N_y}{n_x \log N_x + n_y \log N_y} h_y,$$

which, since n_x and n_y are arbitrary, implies that one can get *arbitrarily* close to *any* point on the convex hull of $\underline{h}_x$ and $\underline{h}_y$. This implies the convexity of the closure of Ω_n^*. □

Now let us form the convex combination of the distribution of two sets of random variables X_i and Y_i, $i = 1, \ldots, n$, with fixed alphabet size N. For this matter define the random variables,

$$Z_i = \begin{cases} X_i & \text{when} \quad \theta = 0 \\ Y_i & \text{when} \quad \theta = 1 \end{cases} \tag{2.6}$$

where θ is a random variable that is 0 with probability p_θ and 1 with probability $1 - p_\theta$. Then the probability distribution of Z_i is,

$$p_{Z_1,\ldots,Z_n}(z_1, \ldots, z_n) = p_\theta p_{X_1,\ldots,X_n}(z_1, \ldots, z_n) + (1 - p_\theta) p_{Y_1,\ldots,Y_n}(z_1, \ldots, z_n). \tag{2.7}$$

If we denote the entropy vectors of the set of random variables X_i, Y_i, and Z_i by h_X, h_Y, and h_Z, respectively, then clearly it is not *true* that,

$$h_Z = p_\theta h_X + (1 - p_\theta) h_Y. \tag{2.8}$$

However, the next proof shows that this is true in the limit!

Proof 2: (*Convex combination of distributions*) Make T independent copies of each of the sets of random variables X_i and Y_i and consider the distribution

$$p_\theta \prod_{t=1}^{T} p_{X_1,\ldots,X_n}(z_1^t, \ldots, z_n^t) + (1 - p_\theta) \prod_{t=1}^{T} p_{Y_1,\ldots,Y_n}(z_1^t, \ldots, z_n^t). \tag{2.9}$$

Now for any $\alpha \subseteq \{1, \ldots, n\}$, we have

$$\underbrace{H(Z_\alpha^T|\theta)}_{p_\theta H(X_\alpha^T) + (1-p_\theta) H(Y_\alpha^T)} \leq H(Z_\alpha^T) \leq \underbrace{H(Z_\alpha^T, \theta)}_{= H(Z_\alpha^T|\theta) + H(p_\theta)} . \tag{2.10}$$

Denote Z_i^T by $\tilde{Z}_i$ whose alphabet size is N^T and its corresponding normalized entropy vector by $h_{\tilde{z}}$. Then normalizing (2.10) by $\log N^T$ yields,

$$p_\theta \underline{h}_X + (1-p_\theta)\underline{h}_Y \leq \underline{h}_{\tilde{Z}} \leq p_\theta \underline{h}_X + (1-p_\theta)\underline{h}_Y + \frac{-p_\theta \log p_\theta - (1-p_\theta)\log(1-p_\theta)}{T \log N} \quad (2.11)$$

which shows the convexity of the closure as $T \to \infty$. □

We remark that, for any fixed N, the set of entropic vectors is highly nonconvex. It is the fact that N is arbitrary (and can grow unbounded) that yields convexity.

We end this section by emphasizing that our choice of normalized entropy vectors, and letting N be arbitrary, retains all the information needed to solve network information theory problems, yet "smooths out" all the irregularities in Γ_n^*. In fact, the relationship between the two sets is as follows:

Theorem 2.2.4 ($\boldsymbol{\Omega_n^*}$ and $\boldsymbol{\Gamma_n^*}$) *Define the ray of a set $\mathcal{S}$ as*

$$ray(\mathcal{S}) = \{\alpha X | \alpha > 0, X \in \mathcal{S}\}. \quad (2.12)$$

Then we have

$$ray(\bar{\Omega}_n^*) = \bar{\Gamma}_n^*, \quad (2.13)$$

i.e., the ray of $\bar{\Omega}_n^*$ is the closure of Γ_n^*.

Proof: Let $V \in \bar{\Gamma}_n^*$. This means that for any $\epsilon > 0$ and $\alpha \subset \{1, \ldots, n\}$ there exists random variables $X_1, \ldots, X_n$ of some alphabet size N such that $|H(X_\alpha) - V_\alpha| \leq \epsilon$. Therefore $\frac{1}{\log N} V \in \bar{\Omega}_n^*$ and so $V \in \text{ray}(\bar{\Omega}_n^*)$.

Conversely suppose $V \in \text{ray}(\bar{\Omega}_n^*)$; then by definition there exists a β such that $\frac{1}{\beta} V \in \bar{\Omega}_n^*$, from which it follows that for any $\epsilon \geq 0$ there exist random variables $X_1, \ldots, X_n$ with alphabet size N for which $|\underline{h}(X_\alpha) - \frac{1}{\beta} V_\alpha| \leq \epsilon$. Thus, $\frac{\log N}{\beta} V$ is a non-normalized entropic vector. Since $\bar{\Gamma}_n^*$ is a convex cone [Yeu02], this implies that $V \in \bar{\Gamma}_n^*$. □

2.3 Network Capacity as a Convex Optimization

2.3.1 Objective and Constraints

Let us now return to the network problem (2.3) and study the consequences of what we have considered so far. Consider *all* the random variables in the network and designate them by X_i, $i = 1, \ldots, n$ (the X_i will thus include both the sources, the destinations, as well as any random variables internal to the network). Now due to the normalization $\frac{1}{T}$ in (2.3), we can simply write the objective as a linear combination of entropic vectors constructed from the X_i. Furthermore, since we consider the closure of the set of entropic vectors, the $\lim_{T\to\infty}$ does not pose any problems. Finally, since the set of entropic vectors is dense in its closure, replacing optimization over Ω_n^* by optimization over $\bar{\Omega}_n^*$ does not cause a problem.

The upshot of all these arguments is that (2.3) can be rewritten as

$$\sup \alpha^t h, \tag{2.14}$$

where α is the vector of coefficients and $\cdot^t$ refers to transpose. The optimization (2.14) should be performed subject to $h \in \bar{\Omega}_n^*$ and subject to the constraints imposed by the network. These are of two kinds.

2.3.1.1 Topological Constraints

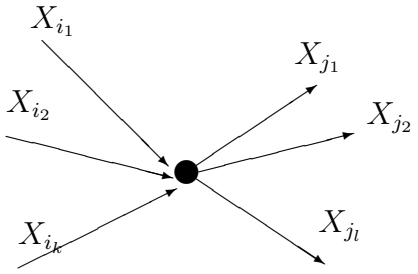


Figure 2.3: Topological constraints at any non-source node

Topological constraints have to do with the information flow in the network. Thus, consider a "non-source" node in the network with incoming messages $\{X_{i_p}\}_{p=1}^{k}$ and outgoing messages $\{X_{j_q}\}_{q=1}^{l}$ (see Fig. 2.3). Then clearly, we have the following linear constraints on the entropy

$$h\left(X_{j_q}|X_{i_1},\ldots,X_{i_k}\right)=0 \tag{2.15}$$

or, equivalently:

$$h\left(X_{j_q},X_{i_1},\ldots,X_{i_k}\right)-h\left(X_{i_1},\ldots,X_{i_k}\right)=0 \tag{2.16}$$

for all $q=1,\ldots l$. Alternatively, at source nodes we have $h(S_i,S_j)-h(S_i)-h(S_j)=0$, if source nodes i and j are independent or $h(S_i,S_j)=h(S_i)=h(S_j)$, if source nodes i and j are identical.

The conclusion is that topological constraints simply introduce linear constraints on the entries of the entropy vector.

2.3.1.2 Channel Constraints

Channel constraints do not translate directly to entropies. What they do is constrain the joint distribution of all random variables in the network (which then determines the admissible entropy vectors). Thus, referring to Fig. 2.4, let a certain discrete memoryless channel relate the messages X_i and X_j. Therefore,

$$p(X_i,X_j)=p(X_j|X_i)p(X_i), \tag{2.17}$$

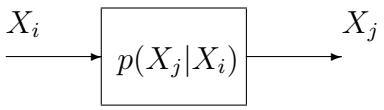


Figure 2.4: A channel internal to the network

or, equivalently,

$$\int \prod_{k\neq i,j} dX_k \; p(X_1,\ldots,X_n) = p(X_j|X_i) \int \prod_{k\neq i} dX_k \; p(X_1,\ldots,X_n) \tag{2.18}$$

which is simply a linear constraint on the joint distribution. Now the fact that the underlying distribution satisfies linear constraints has no effect on the validity of the two proofs we gave for Theorem 2.2.3: time-sharing two sets of random variables will satisfy the linear channel constraints if the original ones do and similarly convex combination of any two distributions also satisfies the same linear channel constraints that the initial ones do.

Therefore the presence of channels inside the network does not affect the convexity of the space of admissible entropy vectors. We formalize this result in the following theorem.

Theorem 2.3.1 (Channel-constrained entropic vectors) *Let $\Omega^*_{n,c}$ denote the space of entropic vectors that are constrained by the discrete memoryless channels in the network. Then the closure of this set, i.e., $\bar{\Omega}^*_{n,c}$, is convex.*

2.3.2 Capacity Formulation as Convex Optimization

From the above discussions we can conclude that the problem (2.3) is equivalent to,

$$\max_{h\in\bar{\Omega}^*_{n,c}, Ah=0} \alpha^t h, \tag{2.19}$$

where $\bar{\Omega}^*_{n,c}$ denotes the convex space of channel-constrained entropic vectors, and A is a matrix multiplying the entropy vector h such that the $Ah = 0$ represents the topological constraints (which, as was stated earlier, enforce linear equalities on the joint entropies). Note that, since the constraint set is closed, we can use max, rather than sup.

Remarks: The formulation (2.19) is significant for at least two reasons:

1. By going to the space of normalized entropy vectors, we have circumvented the problem of "infinite-letter characterization".

2. We have also circumvented the "nonconvexity". (2.19) is a convex optimization problem. In fact, the infinite-letter characterization is what yields convexity (the space of entropic vectors is not convex for any finite T).

2.3.3 Some Applications

2.3.3.1 Duality and Cutset Bounds

As a first attempt, a simple use of some basic machinery from convex optimization yields some interesting results. In network flow problems, the duality between *max-flow* and *min-cut* is well known [EFS56, FF56]. In information theory cutset outerbounds are also well known (see, e.g., [CT91]); however, to the best of our knowledge, these have been obtained by relaxing the network problem to a point-to-point problem (assuming certain nodes can fully cooperate with the sources and others can fully cooperate with the destinations), rather than through any duality argument.

Note that in (2.19) we can enforce the linear constraints through a Lagrange multiplier λ to obtain,

$$\max_{h \in \bar{\Omega}^*_{n,c}, Ah=0} \alpha^t h = \max_{h \in \bar{\Omega}^*_{n,c}} \min_{\lambda} \left(\alpha^t h + \lambda^t A h \right). \tag{2.20}$$

Using the duality of convex optimization we can interchange the max and min to obtain

$$\max_{h \in \bar{\Omega}^*_{n,c}, Ah=0} \alpha^t h = \min_{\lambda} \max_{h \in \bar{\Omega}^*_{n,c}} \left(\alpha^t h + \lambda^t A h \right). \tag{2.21}$$

In particular, for any λ, we have the upper bound

$$\max_{h \in \bar{\Omega}^*_{n,c}, Ah=0} \alpha^t h \leq \max_{h \in \bar{\Omega}^*_{n,c}} \left(\alpha^t h + \lambda^t A h \right). \tag{2.22}$$

Consider now an arbitrary cut through the network, such that all the source nodes reside on one side of the cut and all the destination nodes on the other side of the cut. Set to zero all components of the Lagrange multiplier λ that correspond to edges that do not cross the cut. Now all nodes on either side of the cut can fully cooperate and so the problem becomes a point-to-point problem whose value is simply the cut capacity. Therefore the upper bound in (2.22), after minimizing over the remaining components of λ, is simply the cutset upper bound corresponding to this cut. We have therefore obtained an interpretation of cutset bounds through duality and Lagrange multipliers. More clever choices of the Lagrange multiplier can lead to improved upper bounds over the cutset bound.

2.3.3.2 Wired Networks

In the current framework, solving network information theory problems requires characterizing the set $\bar{\Omega}^*_{n,c}$. This seems formidable (to say the least). However, as we shall presently see, for wired networks things simplify considerably. Wired networks are defined through three main characteristics:

1. Each link represents a (discrete memoryless) channel.

2. The signals transmitted on outgoing edges of a node (X_i, X_j in Fig. 2.5) can be distinct.

3. The signals impinging on a node (X_k, X_l in Fig. 2.5) are received without interference.

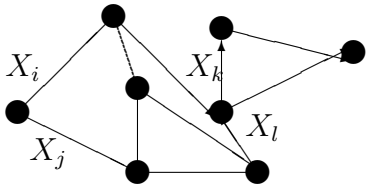


Figure 2.5: A wired network

Due to these properties the network capacity problem can be greatly simplified. In fact for wired networks the separation of channel and network coding [SYC06, KEM09, JE10], implies that any noisy link with information theoretic capacity C can be replaced with a lossless bit pipe of the same capacity. In such scenario it is shown that a set of rates are achievable in a noisy network if they are achievable in an equivalent network whose links are replaced with noiseless links of the same capacity. Therefore instead of studying the noisy network where for each link there is a transmitted signal at the input and a received signal at its output, we may obtain the rate region of the equivalent noiseless network by assigning a single random variable X_e to every noiseless bit pipe e with the constraint that $h_{X_e} \leq C_e$ where C_e is the capacity of the noiseless link e. Hence if we assume that the total number of random variables (sources and all X_e) in the noiseless network is k, then the channel constrained entropy region simplifies to

$$\bar{\Omega}_k^* \cap \left\{ \underline{h} \,|\, \underline{h}_{X_e} \leq C_e \right\}. \tag{2.23}$$

From the above argument we can easily conclude the following theorem.

Theorem 2.3.2 (Wired Networks Capacity) *Consider a wired network that is acyclic and memoryless and has a total of k edges and sources. Moreover assume that there is no feedback from the destinations to sources. Then any network information*

theory problem reduces to a problem of the form,

$$\max_{h\in\bar{\Omega}_k^*,\ \forall e, h_{X_e}\leq C_e,\ Ah=0} \alpha^T h, \tag{2.24}$$

where $Ah = 0$ *represents the topological constraints of the network and* C_e *is the capacity of edge* e*.*

Therefore for wired networks determining the rate region solely requires the characterization of the unconstrained entropy region $\bar{\Omega}_k^*$ of k variables.

2.4 Conclusions

In this chapter, we introduced the notion of normalized entropic vectors—slightly different from the standard definition in the literature in that we normalize entropy by the logarithm of the alphabet size. We argued that this definition is more natural for determining the capacity region of networks and, in particular, that it smooths out the irregularities of the space of non-normalized entropy vectors and renders the closure of the resulting space convex (and compact). Furthermore, the closure of the space remains convex even under constraints imposed by memoryless channels internal to the network. It therefore followed that, for a large class of acyclic memoryless networks, the capacity region for an arbitrary set of sources and destinations can be found by maximization of a linear function over the *convex* set of channel-constrained normalized entropic vectors and some linear constraints. This formulation circumvents the "infinite-letter characterization" issue, as well as the nonconvexity of earlier formulations, and exposes the core of the problem: characterization of the entropy region. We showed that the approach allows one to obtain the classical cutset bounds via a duality argument. Furthermore, for wired networks where the separation of channel and network coding holds the channel constrained entropy region simplifies considerably where one only needs to characterize the unconstrained entropy region

to determine the rate region of such networks.

Chapter 3

Lattice-Based Entropy Construction

3.1 Introduction

Let $X_1, \ldots, X_n$ be a collection of n jointly distributed finite-alphabet random variables and consider the $2^n - 1$ dimensional vector whose entries are the joint entropies of each non-empty subset of these n random variables. Any $2^n - 1$ dimensional vector that can be constructed from the entropies of n such random variables is referred to as *entropic*. The region of all entropic vectors for n random variables is referred to as Γ_n^*.

Characterizing the region of entropic vectors has long been an interesting open problem. Many issues in information theory and probabilistic reasoning, such as optimizing information quantities or characterizing the compatibility of conditional independence relations, involve or are closely related to this problem. Moreover, as it has been proved that the closure of Γ_n^* is a convex cone for any n, characterizing this region is fundamental in the sense that many network information theory problems can be formulated as convex optimization problems over this region. Thus, determining this region can lead to the solution of a whole host of information-theoretic problems. On the other hand many proofs of the converse of coding theorems involve information inequalities, the complete set of which can be found as a result of

characterizing this region.

The work of Han, Fujishige, Zhang, and Yeung, [Fuj78, Han81, ZY98, ZY97, Yeu97] has resulted in the complete characterization of Γ_n^* for $n = 2, 3$ and their relation to polymatroids and submodular functions. In particular, if we let $\mathcal{N} = \{1, ..., n\}$, $\alpha, \beta \subseteq \mathcal{N}$, $X_\alpha = \{X_i : i \in \alpha\}$ and $X_\beta = \{X_i : i \in \beta\}$, it is clear that the entropy $H(X_\alpha) = H(\alpha)$ (for simplicity) satisfies the properties

1. $H(\emptyset) = 0$
2. For $\alpha \subseteq \beta$: $H(\alpha) \leq H(\beta)$
3. For any α, β: $H(\alpha \cup \beta) + H(\alpha \cap \beta) \leq H(\alpha) + H(\beta)$

the last of which is referred to as the submodularity property. The above inequalities are referred to as the basic inequalities of Shannon information measures (and are derived from the positivity of conditional mutual information). Any inequalities that are obtained as positive linear combinations of these are simply referred to as *Shannon inequalities* [ZY98]. The space of all vectors of $2^n - 1$ dimensions whose components satisfy all such Shannon inequalities is denoted by Γ_n. It is known that all valid information inequalities for up to 3 random variables are "Shannon type inequalities". Therefore $\Gamma_2^* = \Gamma_2$ and $\bar{\Gamma}_3^* = \Gamma_3$, where $\bar{\Gamma}_3^*$ is the closure of Γ_3^* [ZY97].

For 4 or more number of random variables however, it was discovered [ZY97] that there are information inequalities which do not follow from the positivity of conditional mutual information and hence they are "non-Shannon type information inequalities". From the discovery of these inequalities it followed that Γ_4^* is *strictly* smaller than Γ_4. The non-Shannon inequalities have also proven useful in deriving various outer bounds for different network information theory problems [DFZ07, YZ99].

Although various outer bounds have been found for the entropy region of 4 or more number of random variables by discovering new non-Shannon type information inequalities [ZY97, DFZ06a, MMRV02, Zha03, Mat07b], its complete characterization

remains an open problem. Moreover there has been much less focus on determining *inner bounds* on Γ_n^* [ZY98, MS95]. These would be of great interest since they would yield achievable rate regions for many network information theory problems. The most well known inner bound for the entropy region is the so-called "linear representable entropy region" [YLCZ06].

Definition 3.1.1 (Linear representable entropy) *An entropy vector h of n random variables is called linear representable if there are subspaces $v_1, \ldots, v_n$ over GF(q) such that for any $\alpha \subseteq \{1, \ldots, n\}$, we have $h_\alpha = rank\ (\oplus_{i\in\alpha} v_i)$ where $\oplus$ denotes the space spanned by $\{v_i, i \in \alpha\}$. Denote the linear representable entropy region of n random variables by Γ_n^r.*

Clearly, $\Gamma_n^r \subseteq \Gamma_n^* \subseteq \Gamma_n$. In fact it is known that all representable entropy vectors, satisfy an inequality called the "Ingleton bound" which does not hold for all entropies in general.

Definition 3.1.2 (Ingleton inequality) *For a subset of at least 4 random variables i, j, k, l, the Ingleton bound is as follows [Ing71],*

$$h_i + h_j + h_{ijk} + h_{ijl} + h_{kl} \leq h_{ij} + h_{ik} + h_{il} + h_{jk} + h_{jl}. \tag{3.1}$$

Although the linear representable entropy regions of 4 [HRSV00] and very recently 5 [DFZ10] random variables have been determined, the general characterization of Γ_n^r remains an open problem. In essence there exists no generalizable approach for obtaining inner bounds of the entropy region for any number of random variables. Creating such inner bounds is the main goal of this chapter. In fact we present a method that obtains polytope inner bounds for the entropy region and that can be generalized to any number of random variables. Polytope inner bounds are specially useful since they allow one to solve network problems via a linear program.

We should mention that in this chapter, we shall focus on *normalized* entropy

vectors. Recall from Chapter 2, Definition 2.2.2, that the normalized entropy vector of n discrete random variables of alphabet size N is defined as the $2^n - 1$ dimensional vector of all normalized joint entropies $\underline{h}_\alpha$,

$$\underline{h}(\alpha) = \frac{1}{\log N} h(X_\alpha), \quad \forall \alpha \subseteq \{1, \ldots, n\} \tag{3.2}$$

and the region of all such normalized entropy vectors by Ω_n^*. As it was discussed in Chapter 2, there are several reasons for considering this normalized version: it is often the normalized version that comes up in capacity calculations (where the normalization represents the number of channel uses) and it makes the entropy region finite [HS07a]. Moreover it can be shown that $\bar{\Omega}_n^*$ is convex and the notion of normalized entropy makes this proof trivial.

The difficulty in characterizing the entropy region is that one should consider all possible distributions of n random variables over *any* alphabet size N. However it turns out that there is a set of probability distributions that are sufficient for characterizing Γ_n^* and therefore these are the distributions we will focus on in this chapter.

The remainder of the chapter is organized as follows. The next section studies quasi-uniform distributions, which will be the building blocks for our construction. Section 3.3 contains the main results of our method, especially the construction of entropic vectors using lattice-generated probability distributions. Section 3.4 makes the construction explicit for 2,3,4, and 5 random variables and shows the tightness of our construction for $n = 2, 3$. Finally, in Section 3.5, quasi-uniform distributions of alphabet size 2 are studied.

3.2 Quasi-Uniform Distributions

One way of characterizing Γ_n^* is through determining its kissing hyperplanes:

$$a^t H = \sum_{\alpha \subseteq \mathcal{N}} a_\alpha H_\alpha \geq \gamma, \tag{3.3}$$

for $a \in \mathbb{R}^{2^n-1}$ and for all $H \in \Gamma_n^*$. To determine the value of γ, one needs to perform the optimization,

$$\gamma = \min_{H \in \Gamma_n^*} \sum_{\alpha \subseteq \mathcal{N}} a_\alpha H_\alpha. \tag{3.4}$$

One of the difficulties in performing this optimization is that the alphabet size of the underlying distribution is arbitrary. Nonetheless, if we restrict the alphabet size of each X_i to N and attempt to optimize over the unknown joint distribution $p_{X_\mathcal{N}}(x_\mathcal{N})$ then we can use the Lagrange multipliers to write the following unconstrained optimization problem,

$$\min_{p_{X_\mathcal{N}}} \max_{\lambda \geq 0, \mu} \sum a_\alpha H(x_\alpha) + \mu \left(\sum p_\mathcal{N}(x_\mathcal{N}) - 1 \right) - \sum \lambda(x_\mathcal{N}) p_{X_\mathcal{N}}(x_\mathcal{N}). \tag{3.5}$$

Enforcing the KKT conditions by taking the derivative with respect to $p_{X_\mathcal{N}}(x_\mathcal{N})$ gives,

$$\sum_{\alpha \subseteq \mathcal{N}} a_\alpha \log \frac{1}{p_{X_\alpha}(x_\alpha)} = c \quad \text{if} \quad p_{X_\mathcal{N}}(x_\mathcal{N}) \neq 0, \tag{3.6}$$

for some constant c. The KKT conditions imply that, rather than searching over all possible distributions $p_{X_\mathcal{N}}(x_\mathcal{N})$, we need only search over those distributions that satisfy (3.6).

Of course, there can be many solutions to (3.6). However, a rather obvious solution—and one that does not depend on a, the normal vector of the hyperplane—is

the following. For any $\alpha \subseteq \mathcal{N}$:

$$p_{X_\alpha}(x_\alpha) = c_\alpha \text{ or } 0 \tag{3.7}$$

for some constant c_α, independent of the point $x_\alpha \in \{1, \ldots, N\}^{|\alpha|}$. In other words, these are distributions that take on zero or a constant value for all possible marginals, $p_{X_\alpha}(\cdot)$. Such distributions are referred to as *quasi-uniform* [Cha01].

Definition 3.2.1 (Quasi-uniform distribution) *A joint distribution of n discrete random variables $p_{X_\mathcal{N}}(x_\mathcal{N})$ $\mathcal{N} = \{1, \ldots, n\}$ is called "quasi-uniform" [Cha01] if the distribution itself and all its marginals take on a zero or constant value, i.e., $\forall x_\alpha$, $\alpha \subseteq \mathcal{N}$: $p_{X_\alpha}(x_\alpha) = c_\alpha$ or 0 where c_α is a constant depending on α. The space of all quasi-uniform distributions of n random variables is denoted by Λ_n.*

Computing the entropy for quasi-uniform distributions is, of course, straightforward:

$$H(\alpha) = \log \frac{1}{c_\alpha}. \tag{3.8}$$

It also turns out that one can generate quasi-uniform distribution by appealing to the concept of groups as stated in the following.

Theorem 3.2.2 (Quasi-uniforms and groups) *[Cha01, CY02] If G is a finite group whose subgroups are $G_1, \ldots, G_n$, then for any element of $g \in G$ let $X = (X_1, \ldots, X_n)$ be an n dimensional vector whose ith element is the index of coset of G_i to which g belongs. Assigning a constant probability to each X encountered in this fashion yields a quasi-uniform probability for $X_1, \ldots, X_n$ where the joint entropy of a collection of random variables indexed by α is obtained from $h_\alpha = \log \frac{|G|}{|\cap_{i\in\alpha} G_i|}$. The entropy vector such obtained is called "group-derived". The region of all $2^n - 1$ dimensional group-derived entropy vectors is denoted by Υ_n.*

Then the remarkable result of [CY02, Cha01] is that the set of all group-derived

and hence quasi-uniform distributions is sufficient for characterizing the entropy region.

Theorem 3.2.3 (Quasi-Uniform Distribution) $\overline{con}(\Upsilon_n) = \overline{con}(\Lambda_n) = \bar{\Gamma}_n^*$, *i.e., the convex closure of* Λ_n *and* Υ_n *is the closure of* Γ_n^*.

We provide the sketch of the proof, as it is instructive.

Proof: Note that we clearly have $\overline{con}(\Upsilon_n) \subseteq \overline{con}(\Lambda_n) \subseteq \bar{\Gamma}_n^*$. Therefore all we need is to prove that $\bar{\Gamma}_n^* \subseteq \overline{con}(\Upsilon_n)$. The idea is to show that every $2^n - 1$ dimensional entropy vector H is asymptotically characterizable with groups. To provide a sketch of the proof we only show this for H_1. First note that,

$$H_1 = \sum_i p_i \log \frac{1}{p_i} = \log \prod_i \left(\frac{1}{p_i}\right)^{p_i} \tag{3.9}$$

where p is the corresponding marginal distribution. Without loss of generality assume that $p_i = \frac{M_i}{Q}$ where M_i and Q are integers and $\sum_i M_i = Q$. Then using Stirling's approximation we have,

$$H = \frac{1}{Q} \log \prod_i \left(\frac{Q}{M_i}\right)^{M_i} = \frac{1}{Q} \log \left(\frac{Q}{e}\right)^Q \prod_i \frac{1}{\left(\frac{M_i}{e}\right)^{M_i}} \approx \frac{1}{Q} \log \frac{Q!}{\prod_i M_i!}. \tag{3.10}$$

This suggests to define a group G as a permutation group on Q elements. Furthermore partition the set of Q elements into subsets of size M_i and let the subgroup G_1 be the permutation group that permutes within $M_1, \cdots, M_k$. □

Since considering quasi-uniform distributions is sufficient for characterizing $\bar{\Gamma}_n^*$, in this chapter we will focus on the generation of quasi-uniform distributions by considering lattice structures.

3.3 Distributions from Lattices

3.3.1 Principles and Preliminaries of Construction

Determining all quasi-uniform distributions appears to be a hopelessly complicated combinatorial problem. Since we are looking for a construction that can be generalized to any n, it seems reasonable to impose some structure. Some circumspection suggests the use of a lattice structure.

Definition 3.3.1 (Lattice Structure) *In general the points of a lattice in the n dimensional Euclidean space can be represented as follows:*

$$x = Mz, \tag{3.11}$$

where $x \in \mathbb{R}^n$ are points in the lattice, $M \in \mathbb{R}^{n\times n}$ is the so-called lattice-generating matrix, and $z \in \mathbb{Z}^n$ is an integer vector. Since the points we are interested in belong to $\{0,\ldots,N-1\}^n$, we require that x have integer entries. We will therefore henceforth assume that M has non-negative integer entries, so that $M \in (\mathbb{Z}^+)^{n\times n}$. We will refer to the lattice generated by the matrix M as $\mathcal{L}(M)$.

We can assign a probability distribution to a lattice structure.

Definition 3.3.2 (Lattice-Generated Distribution) *A probability distribution over n random variables with alphabet size N each, will be called* lattice-generated*, if for some lattice $\mathcal{L}(M)$, we have,*

$$p_{X_{\mathcal{N}}}(x_{\mathcal{N}}) = \begin{cases} c & x_{\mathcal{N}} \in \{0,\ldots,N-1\}^n \cap \mathcal{L}(M) \\ 0 & \textit{otherwise} \end{cases} . \tag{3.12}$$

Example 3.3.3 (A two-dimensional lattice) *Fig. 3.1 shows a two-dimensional lattice generated by the matrix, $\begin{bmatrix} 2 & 0 \\ 1 & 2 \end{bmatrix}$, when the alphabet-size is $N = 4$. Note that*

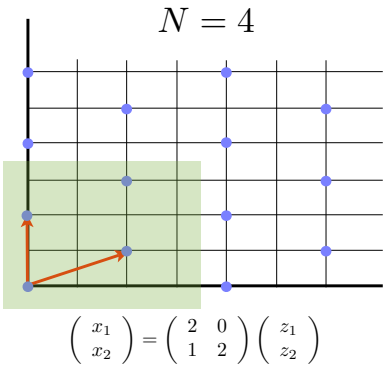


Figure 3.1: An example of a lattice

the probability distributions are as follows,

$$p_{X_1}(0) = p_{X_1}(2) = \frac{1}{2}, \qquad p_{X_1}(1) = p_{X_1}(3) = 0$$
$$p_{X_2}(0) = p_{X_2}(1) = p_{X_2}(2) = p_{X_2}(3) = \frac{1}{4}$$
$$p_{X_1,X_2}(x_1, x_2) = \begin{cases} \frac{1}{4} & (x_1, x_2) \in \{(0,0), (2,1), (0,2), (2,3)\} \\ 0 & \textit{otherwise} \end{cases}.$$

Having this distribution, we can easily compute the corresponding entropies and hence the entropy vector.

We now need a few lemmas.

Lemma 3.3.4 (Bezout Identity) *The following equality holds for 2-by-2 lattices.*

$$\mathcal{L}\left(\begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}\right) = \mathcal{L}\left(\begin{bmatrix} gcd(M_{11}, M_{12}) & 0 \\ M_{21}x + M_{22}y & \frac{M_{11}M_{22} - M_{21}M_{12}}{gcd(M_{11}, M_{22})} \end{bmatrix}\right), \tag{3.13}$$

where x, y are integers found from the Bezout identity $M_{11}x + M_{12}y = gcd(M_{11}, M_{12})$.

Proof: Follows from post-multiplication by $\begin{bmatrix} x & -M_{12}/\text{gcd}(M_{11}, M_{12}) \\ y & M_{11}/\text{gcd}(M_{11}, M_{12}) \end{bmatrix}$, which is a unimodular matrix. □

Lemma 3.3.5 (Lower Triangularization) *Any lattice-generating matrix with non-negative integer entries can be lower triangularized without changing the resulting lattice.*

Proof: Follows from repeated use of Lemma 3.3.4 in a fashion akin to QR factorization. □

Hence, we will assume that the lattice-generating matrix M is lower triangular.

Consider a lattice generated by the matrix,

$$\begin{bmatrix} N & 0 \\ 1 & N \end{bmatrix}. \tag{3.14}$$

The resulting lattice is shown in Figure 3.2. Note that if we want to calculate the entropies of X_1 and X_2 based on the possible values that X_1 and X_2 take in the $(0, 1, \ldots, N-1)$ range (which in turn defines their probability distribution), we end up counting the values for X_2 that fall in this range which are outside of the $(0, 1, \ldots, N-1) \times (0, 1, \ldots, N-1)$ box and we will obtain the normalized entropy $h_2 = 1$. On the other hand if we only focus on the $(0, 1, \ldots, N-1) \times (0, 1, \ldots, N-1)$ box we get $h_2 = 0$. To avoid this problem of counting the points that do not belong to the $(0, 1, \ldots, N-1) \times (0, 1, \ldots, N-1)$ square, we need the lattice to be periodic with a period that divides N.

Lemma 3.3.6 (Lattice-Generated Quasi-Uniforms) *A lattice-generated distribution is quasi-uniform if the lattice has a period that divides N. The latter is true if, and only if, the matrix $M^{-1}N$ has integer entries.*

Proof: Assume the lattice M has a period that divides N. This is true if, and only if, for every $x \in \mathcal{L}(M)$ the point $x + Ne_i$ belongs to $\mathcal{L}(M)$ for all $i = 1, \ldots, n$, where e_i is the i-th unit vector with one in the i-th position and zeros elsewhere. In other words, if there exists an integer vector z such that $Mz = x$, there should

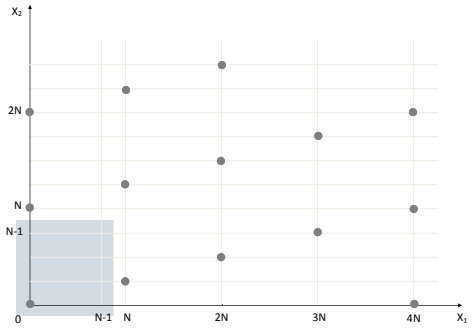


Figure 3.2: A lattice whose period does not divide N

also exist an integer vector $z^{(i)}$ such that $Mz^{(i)} = x + Ne_i$. Therefore we obtain $M(z^{(i)} - z) = Ne_i$. Letting $i = 1$ and assuming that M is lower triangular based on Lemma 3.3.5, we immediately obtain $M_{11} \neq 0$. Continuing this process for all $i \leq n$ we deduce that $\forall i,\ M_{ii} \neq 0$ which for a lower triangular matrix implies that $\det M \neq 0$ and therefore M^{-1} exists. As a result we can rewrite the above relation as $z^{(i)} = M^{-1}x + M^{-1}Ne_i = z + M^{-1}Ne_i$, which means that $M^{-1}Ne_i$ should have integer entries for all i and establishes the second claim of the lemma.

We now need to show that the resulting distribution is quasi-uniform. To this end, note that

$$p_{X_\alpha}(x_\alpha) = \sum_{x_{\mathcal{N}-\alpha}} p_{X_\mathcal{N}}(x_\mathcal{N}) = c \sum_{x_{\mathcal{N}-\alpha}, p_{X_\mathcal{N}}(x_\alpha, x_{\mathcal{N}-\alpha})\neq 0} 1,$$

which implies that a distribution taking on only the values 0 and c is quasi-uniform if, and only if, for every value x_α for which $p_{X_\alpha}(x_\alpha)$ is nonzero, the *number* of $x_{\mathcal{N}-\alpha}$ for which $p_{X_\mathcal{N}}(x_\alpha, x_{\mathcal{N}-\alpha})$ is nonzero should be *constant*. Now partitioning the lattice-generating matrix according to α and $\alpha^c = \mathcal{N} - \alpha$ and lower-triangularizing using

Lemma 3.3.5 yields

$$\begin{bmatrix} x_\alpha \\ x_{\alpha^c} \end{bmatrix} = \begin{bmatrix} M_{\alpha,\alpha} & 0 \\ M_{\alpha^c,\alpha} & M_{\alpha^c,\alpha^c} \end{bmatrix} \begin{bmatrix} z_\alpha \\ z_{\alpha^c} \end{bmatrix}. \tag{3.15}$$

Any value of x_α that yields a nonzero $p_{X_\alpha}(x_\alpha)$ is one for which $z_\alpha = M_{\alpha,\alpha}^{-1} x_\alpha$ is an integer vector. Therefore so is the vector $M_{\alpha^c,\alpha} M_{\alpha,\alpha}^{-1} x_\alpha$. Thus, the number of x_{α^c} for which $p_{X_\mathcal{N}}(x_\alpha, x_{\alpha^c})$ is nonzero is given by the number of x_{α^c} in $\{0, \ldots, N-1\}^{|\alpha^c|}$ for which $M_{\alpha^c,\alpha^c} n_{\alpha^c} = x_{\alpha^c} - M_{\alpha^c,\alpha} M_{\alpha,\alpha}^{-1} x_\alpha$ has an integer solution in n_{α^c}. However, since the lattice is periodic with period dividing N, this number is independent of the integer shift $M_{\alpha^c,\alpha} M_{\alpha,\alpha}^{-1} x_\alpha$. □

Lemma 3.3.6 tells us that we should focus on lattice-generating matrices such that $M^{-1}N$ has integer entries. The next lemma of this section shows us how to extract the entropies from the lattice-generating matrix M.

Lemma 3.3.7 (Entropy Extraction) *Consider a lattice-generated distribution with period dividing N. Then the normalized entropy of any collection of random variables X_α is given by,*

$$\underline{h}(\alpha) = |\alpha| - \frac{\log |\det M_{\alpha,\alpha}|}{\log N}, \tag{3.16}$$

where $M_{\alpha,\alpha}$ is found from the lower triangularization of the lattice-generating matrix in (3.15).

Proof: The distribution is quasi-uniform and so the entropy $h(\alpha)$ is simply the log of the number of nonzero points in the distribution $p_{X_\alpha}(x_\alpha)$. The total number of points is $N^{|\alpha|}$ and the volume of the basic volume element in the lattice corresponding to the variables x_α is well known to be $|\det M_{\alpha,\alpha}|$. This gives the total number of nonzero points in the distribution as $N^{|\alpha|}/|\det M_{\alpha,\alpha}|$, which when normalized by $\log N$ yields the desired result. □

It turns out that we can further simplify the entropy extraction formula of (3.16).

Lemma 3.3.8 (Entropy simplification) *The normalized entropy of a collection of random variables X_α of alphabet size N, with a lattice-generated distribution whose period divides N can be obtained from*

$$\underline{h}(\alpha) = |\alpha| - \frac{\log|\gcd(\text{ all } |\alpha| \times |\alpha| \text{ minors of } \mathrm{M}_\alpha)|}{\log N}, \tag{3.17}$$

where M is the lattice-generating matrix and M_α is the submatrix obtained by selecting those rows of M which are indexed by α.

Proof: In a more general setting, (3.15) can be written as follows:

$$\begin{bmatrix} x_\alpha \\ x_{\alpha^c} \end{bmatrix} = \begin{bmatrix} M_\alpha \\ M_{\alpha^c} \end{bmatrix} \begin{bmatrix} z_\alpha \\ z_{\alpha^c} \end{bmatrix} \tag{3.18}$$

where M_α is $|\alpha| \times n$ and low-rank. We can write the equivalent Smith normal forms [Smi84] of M_α and M_{α^c} as,

$$M_\alpha = U_1 D_1 V_1 \tag{3.19}$$

$$M_{\alpha^c} = U_2 D_2 V_2 \tag{3.20}$$

where U_i and V_i are unimodular matrices, and D_i are of the form

$$D_i = \begin{bmatrix} d_1^i & & & 0 & \dots & 0 \\ & \ddots & & \vdots & \ddots & \vdots \\ & & d_{|\alpha|}^i & 0 & \dots & 0 \end{bmatrix} \triangleq \begin{bmatrix} (\hat{D}_i)_{(|\alpha|\times|\alpha|)} & \mathbf{0}_{(|\alpha|\times(n-|\alpha|))} \end{bmatrix} \tag{3.21}$$

with the property that $d_j^i = \frac{\Delta_i(j)}{\Delta_i(j-1)}$ where $\Delta_i(j)$ is the gcd of the $j \times j$ minors of the corresponding matrix M_α or M_{α^c} and $\Delta_i(0) \triangleq 1$. Therefore (3.18) can be written as

follows,

$$\begin{bmatrix} x_\alpha \\ x_{\alpha^c} \end{bmatrix} = \begin{bmatrix} U_1 D_1 \\ U_2 D_2 V_2 V_1^{-1} \end{bmatrix} V_1 \begin{bmatrix} n_\alpha \\ n_{\alpha^c} \end{bmatrix}. \tag{3.22}$$

However,

$$\mathcal{L}\left(\begin{bmatrix} U_1 D_1 \\ U_2 D_2 V_2 V_1^{-1} \end{bmatrix} V_1 \right) = \mathcal{L}\left(\begin{bmatrix} U_1 D_1 \\ U_2 D_2 V_2 V_1^{-1} \end{bmatrix} \right). \tag{3.23}$$

Noting that $U_1 D_1 = \begin{bmatrix} (U_1 \hat{D}_1)_{(|\alpha| \times |\alpha|)} & \mathbf{0}_{(|\alpha| \times (n-|\alpha|))} \end{bmatrix}$ and using Lemma 3.3.7 we obtain,

$$\underline{h}(\alpha) = |\alpha| - \frac{\log |\det \hat{D}_1|}{\log N}, \tag{3.24}$$

or equivalently,

$$\underline{h}(\alpha) = |\alpha| - \frac{\log |\gcd(\text{ all } |\alpha| \times |\alpha| \text{ minors of } \mathrm{M}_\alpha)|}{\log N}, \tag{3.25}$$

which is often easier to compute. □

3.3.2 Actual Construction

In this section, we show how we can indeed simplify and calculate the lattice-derived entropies. Note that from Lemmas 3.3.5 and 3.3.6, we are assuming that the lattice-generating matrix is a lower triangular matrix whose diagonal entries are nonzero. As a matter of fact we can further assume that all the off-diagonal entries are nonzero as well. The reason is that if any of the off-diagonal entries are 0, since the whole generator matrix is full rank, we can replace M with another full rank lower-triangular generator matrix which does not have any zero entries and generates the same lattice. To be more exact, if $M_{ij} = 0$, then column j of M can be easily replaced by a linear

combination of columns j and i such that the ij entry will no longer be zero. In summary and recalling the conditions of Definition 3.3.1, we are making the following assumptions about the entries of M,

- $\forall i, j: \ M_{ij} \in \mathbb{Z}$ and $M_{ij} > 0$.

Now we should remark that for any integers M_{ij} it is always possible to find a large enough integer N and positive rational numbers γ_{ij} such that $M_{ij} = N^{\gamma_{ij}}$.

Furthermore, for large enough N it follows that $\gcd(N^{\gamma_{ij}}, N^{\gamma_{kl}}) = N^{\min(\gamma_{ij},\gamma_{kl})}$ which considerably simplifies the gcd calculation as it appears in entropy extraction formula (3.17). Since we are studying normalized entropies, increasing N comes at no cost and so we will assume all of the above.

Describing the entries of the matrix M in terms of the rational numbers γ_{ij} will eventually result in the description of the lattice region in terms of γ_{ij}'s. However before attempting to derive the joint entropies, note that the quasi-uniformity requirement of Lemma 3.3.6 also translates into a set of conditions for γ_{ij}'s.

Theorem 3.3.9 ($\boldsymbol{\gamma}$ constraints) *Enforcing quasi-uniformity on the lattice distribution, results in the following constraints on γ_{ij}'s,*

$$\begin{aligned} &\rho(i_1, \ldots, i_k) = \gamma_{i_1 i_1} + \ldots + \gamma_{i_k i_k} - \gamma_{i_k i_{k-1}} - \ldots - \gamma_{i_2 i_1} \leq 1 \\ &\forall \{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\}, \quad i_1 < i_2 < \ldots < i_k. \end{aligned} \tag{3.26}$$

Proof: Note that by Lemma 3.3.6 a lattice-generated distribution is quasi-uniform if NM^{-1} has integer entries. Let the lower-triangular matrix $\tilde{M}_p = [\tilde{m}_{jl}]_{1 \leq l \leq j \leq p}$ be the inverse of the $p \times p$ lower-triangular matrix $M_p = [N^{\gamma_{jl}}]$. Let r denote the index of the set $\{i_1, \ldots, i_{k-1}\}$ in the power set of $\{1, \ldots, i_k - 1\}$. Then we equivalently denote $\rho(i_1, \ldots, i_k)$ by $\rho_{i_k}(r)$. We show that: I) the entries of the inverse matrix, $\tilde{m}_{jl}$ $l \leq j$ are of the form $\tilde{m}_{jl} = \sum_{r=1}^{2^{j-1}} a_r^{(jl)} N^{-\rho_j(r)}$ where $a_r^{(jl)} \in \{-1, 0, 1\}$, and II) for any $1 \leq j \leq p$ and $1 \leq r \leq 2^{j-1}, \exists\ l$ such that $a_r^{(jl)} \neq 0$. Note that from I it follows

that to have integer entries for NM^{-1} we need the terms $N^{1-\rho_j(r)}$ to be integers which is possible for large enough N if the exponents are positive. This in turn gives the inequalities $\rho_j(r) \leq 1$ or equivalently (3.26). Moreover II states that we will need all such inequalities, i.e., $\forall 1 \leq j \leq p, 1 \leq r \leq 2^{j-1}$.

We proceed by induction. For $n = 1$, $M = N^{\gamma_{11}}$, and therefore NM^{-1} is simply equal to $N^{1-\gamma_{11}}$ and I and II are immediately true (note that $\gamma_{11} = \rho_1(1)$). The corresponding inequality in this case is the somewhat trivial inequality $\gamma_{11} \leq 1$ which follows from the positivity of $1-\gamma_{11}$. Next we assume that I and II are true for $n = p$ and prove them for $n = p + 1$. Consider $\tilde{M}_{p+1}$ which is essentially of the form,

$$\tilde{M}_{p+1} = \begin{pmatrix} \tilde{M}_p & \ddots & 0 \\ \tilde{m}_{p+1,1} & \dots & \tilde{m}_{p+1,p+1} \end{pmatrix}. \tag{3.27}$$

Requiring $M_{p+1}\tilde{M}_{p+1}$ to be an identity, gives,

$$\begin{pmatrix} M_p & \ddots & 0 \\ N^{\gamma_{p+1,1}} & \dots & N^{\gamma_{p+1,p+1}} \end{pmatrix} \times \begin{pmatrix} \tilde{M}_p & \ddots & 0 \\ \tilde{m}_{p+1,1} & \dots & \tilde{m}_{p+1,p+1} \end{pmatrix} = I_{p+1} \tag{3.28}$$

where I_{p+1} is the $p+1$ by $p+1$ dimensional identity matrix. Since $M_p\tilde{M}_p = I_p$ we only require,

$$\sum_{j=l}^{p+1} N^{\gamma_{p+1,j}} \tilde{m}_{jl} = 0 \quad l = 1, \dots, p \tag{3.29}$$

$$N^{\gamma_{p+1,p+1}} \tilde{m}_{p+1,p+1} = 1. \tag{3.30}$$

From (3.29) it follows that,

$$\tilde{m}_{p+1,l} = -\sum_{j=l}^{p} N^{-(\gamma_{p+1,p+1}-\gamma_{p+1,j})} \tilde{m}_{j,l} \quad l = 1, \dots, p. \tag{3.31}$$

Replacing for $\tilde{m}_{jl}$ from the induction assumption gives,

$$\tilde{m}_{p+1,l} = -\sum_{j=l}^{p}\sum_{r=1}^{2^{j-1}} a_r^{(jl)} N^{-(\gamma_{p+1,p+1}-\gamma_{p+1,j}+\rho_j(r))}, \quad l = 1,\ldots,p. \tag{3.32}$$

However,

$$\gamma_{p+1,p+1} - \gamma_{p+1,j} + \rho_j(r) = \rho_{p+1}(s)$$

for some s as an index of the elements in the power set of $\{1,\ldots,p\}$. Therefore,

$$\tilde{m}_{p+1,l} = \sum_{s=1}^{2^p} b_s^{(p+1,l)} N^{-\rho_{p+1}(s)}, \quad l = 1,\ldots,p. \tag{3.33}$$

Note that $\rho_{p+1}(1) = \gamma_{p+1,p+1}$, which does not appear in the summation (3.32) and therefore we should have $b_1^{(p+1,l)} = 0$ in (3.33). On the other hand, for any $2 \le s \le 2^p$, there exist unique $j_0 \le p$ and $1 \le r_0 \le 2^{j-1}$ such that $\rho_{p+1}(s) = \gamma_{p+1,p+1} - \gamma_{p+1,j_0} + \rho_{j_0}(r_0)$. Denoting j_0 and r_0 as functions of s by j_s and r_s, respectively, we can write $b_s^{(p+1,l)}$ in terms of $a_r^{(jl)}$,

$$b_s^{(p+1,l)} = \begin{cases} a_{r_s}^{(j_s,l)} & j_s \ge l \\ 0 & j_s < l \ \text{ or } \ s = 1. \end{cases} \tag{3.34}$$

Now for any $s \neq 1$, let $\tilde{l}$ be any column index that is less than j_s, i.e., let $\tilde{l} \in \{1,\ldots,j_s\}$. Therefore for any $\tilde{l}$ we have, $b_s^{(p+1,\tilde{l})} = a_{r_s}^{(j_s,\tilde{l})}$. By induction assumption for any j_s and r_s, $\exists\ l^* \le j_s$ such that $a_{r_s}^{(j_s,l^*)} \neq 0$. Let $\tilde{l} = l^*$ and we obtain $b_s^{(p+1,l^*)} \neq 0$. Finally for $s = 1$, from (3.30) we obtain that $\tilde{m}_{p+1,p+1} = N^{-\gamma_{p+1,p+1}} = N^{-\rho_{p+1}(1)}$ which concludes the proof. □

Remark: Note that (3.26) implies $\gamma_{ii} \le 1$. However, this need not be true for $\gamma_{ij},\ i \neq j$. Although by periodicity of M with period that divides N, one could

argue that the lattice may be generated with another matrix M' whose entries are all bounded by N (equivalently $\gamma_{ij} \leq 1$), however M' will not necessarily be lower-triangular anymore.

Recall that the normalized joint entropy $\underline{h}_\alpha$ of a lattice-based distribution can be obtained via the gcd of all the corresponding $|\alpha| \times |\alpha|$ minors as stated in (3.17). The equivalent γ_{ij} characterization of the entropies can be obtained via the replacement of $M_{ij} = N^{\gamma_{ij}}$ and using a couple of lemmas.

Lemma 3.3.10 *Let d and $e_i,\ i = 1, \ldots, m$ be fixed positive (nonzero) rational numbers. Moreover let N be an integer that can be made arbitrarily large. Then N^d and $(\sum_{i=1}^m N^{e_i} \pm 1)$ are coprime integers.*

Proof: Without loss of generality, let $d = \frac{r}{t} \leq 1$ and $e_i = \frac{s_i}{t} \leq 1$ for some $r, s_i, t \in \mathbb{Z}$ and assume that $e_1 = \min\{e_i, i = 1, \ldots, m\}$. N can be made large enough so that $N^{\frac{1}{t}}$ is an integer. It follows immediately that $N^{\frac{r}{t}}$, $N^{\frac{s_i}{t}}$, and $N^{\frac{s_i - s_1}{t}},\ i = 2, \ldots, m$ are all integers and therefore we can write $\sum_{i=1}^m N^{e_i} \pm 1 = N^{e_1}(1 + \sum_{i=2}^m N^{e_i - e_1}) \pm 1 = (1 + \omega)N^{e_1} \pm 1$ where $\omega = \sum_{i=2}^m N^{e_i - e_1}$ is an integer. Therefore we want to show that N^d and $(1 + \omega)N^{e_1} \pm 1$ are coprime. In the following we prove the lemma for $(1 + \omega)N^{e_1} - 1$, the proof for $(1 + \omega)N^{e_1} + 1$ case is similar. Two cases may be considered,

1. $d \leq e_1$: In this case, $N^{\frac{s_1 - r}{t}}$ is an integer and therefore N^d divides N^{e_1} and henceforth $(1 + \omega)N^{e_1}$. As a result, and since consecutive integers are coprime, we obtain that N^d and $(1 + \omega)N^{e_1} - 1$ are coprime.

2. $d > e_1$: As in the last case, by making $N^{\frac{1}{t}}$ an integer we can make N^d, N^{e_1}, and $N^{d - e_1}$ all integers as well,

$$N^d = c \cdot N^{e_1} \qquad c \triangleq N^{d - e_1}. \tag{3.35}$$

Now assume by contradiction that N^d and $(1+\omega)N^e - 1$ are not coprime. Therefore there exist integers $k > 1$, a and b such that,

$$N^d = k \cdot a \tag{3.36}$$

$$(1+\omega)N^{e_1} - 1 = k \cdot b. \tag{3.37}$$

From equations (3.35)–(3.37) one can easily obtain that,

$$c = k \cdot ((1+\omega)a - cb). \tag{3.38}$$

In other words k also divides N^{d-e_1} and therefore N^{d-e_1} and $(1+\omega)N^{e_1} - 1$ are not coprime either. Now the process can be repeated for $N^{d'} = N^{d-e_1}$ and $(1+\omega)N^{e_1} - 1$. If $d - e_1 \leq e_1$, by reasoning of case 1 we conclude that $N^{d'}$ and $(1+\omega)N^{e_1} - 1$ are coprime, which is a contradiction. Otherwise by repeating (3.35)–(3.37) we obtain that N^{d-2e_1} and $(1+\omega)N^{e_1} - 1$ are not coprime. This can be repeated l steps to obtain N^{d-le_1} and $(1+\omega)N^{e_1} - 1$ are not coprime. When $d - le_1$ becomes less than e_1 contradiction is reached based on case 1, establishing the result of the lemma.

□

Corollary 3.3.11 *If in Lemma 3.3.10 some of the e_i are zero (without loss of generality, e.g., $e_{m'+1}, \ldots, e_m = 0$ for $m' < m$) such that $\sum_{i=1}^{m} N^{e_i} \pm 1 = \sum_{i=1}^{m'} N^{e_i} \pm \psi$ where ψ is a nonzero integer, then there is a class of unbounded N's for which N^d and $\sum_{i=1}^{m'} N^{e_i} \pm \psi$ are coprime.*

Proof: The proof hinges on the fact that we can choose N such that for any rational number $\kappa = \frac{u}{t}$, N^κ, and ψ are coprime. Since for an integer c, ψ and $c\psi + 1$ are coprime, one trivial choice for N is $(c\psi+1)^{t\iota}$ where ι is an arbitrary integer that allows N to be arbitrary large. Therefore, assuming $d = \frac{r}{t}$ and $e_i = \frac{s_i}{t}$, we have

$N^d = (c\psi+1)^{r\iota}$ and $\sum_{i=1}^{m'} N^{e_i} \pm \psi = \sum_{i=1}^{m'}(c\psi+1)^{s_i\iota} \pm \psi$, which can also be written as, $\sum_{i=1}^{m'} N^{e_i} \pm \psi = (c\psi+1)b \pm \psi$, where b is the integer obtained by factoring out $(c\psi+1)$. Let a be the integer equal to $r\iota$ letting us denote $N^d = (c\psi+1)^a$. Now it is easy to show that $(c\psi+1)^a$ and $(c\psi+1)b \pm \psi$ are coprime. If by contradiction, there is a common divisor λ, we can write

$$(c\psi+1)^a = \lambda \cdot \phi \tag{3.39}$$

$$(c\psi+1)b \pm \psi = \lambda \cdot \delta \tag{3.40}$$

for some integers ϕ and δ. Multiplying (3.40) by $(c\psi+1)^{a-1}$, and replacing from (3.39) we obtain

$$(c\psi+1)^{a-1}\psi = \pm\lambda\left(\delta(c\psi+1)^{a-1} - \phi b\right). \tag{3.41}$$

In other words, $\lambda|(c\psi+1)^{a-1}\psi$. However, based on (3.39), λ divides $(c\psi+1)^a$, and hence, λ and ψ should be coprime.[1] From this it follows that λ should divide $(c\psi+1)^{a-1}$. Now, by replacing a with $a-1$ in (3.39), and repeating this argument, we obtain that λ should also divide $(c\psi+1)^{a-2}$. Continuing in this manner, we ultimately obtain that λ should be a common divisor of $(c\psi+1)$ as well, i.e., $c\psi+1 = \lambda \cdot \eta$. Replacing this in (3.40), immediately gives that, λ divides ψ, which is a contradiction. Therefore, $(c\psi+1)^a$ and $(c\psi+1)b \pm \psi$ are coprime, which means that there is an infinite set of choices for N that makes N^d, and $\sum_{i=1}^{m'} N^{e_i} \pm \psi$ coprime. □

Corollary 3.3.12 *Assume $N^{\upsilon}|(\sum_k N^{\sigma_k} - \sum_l N^{\tau_l})$ where σ_k and τ_l are positive rational numbers such that $\sigma_k \neq \tau_l$ $\forall k,l$ (i.e., no cancelation occurs). Then $N^{\upsilon}|N^{\sigma_k}$ and $N^{\upsilon}|N^{\tau_l}$, $\forall k,l$.*

Proof: We can equivalently write $(\sum_k N^{\sigma_k} - \sum_l N^{\tau_l}) = \sum_i (-1)^{\alpha_i} N^{f_i}$ where no two terms cancel in the latter summation and α_i are either 0 or 1. Assuming with-

[1] To see this, assume the prime factorization $c\psi+1 = \prod p_i^{q_i}$. Considering that $\lambda|(c\psi+1)^a$, we obtain $\lambda = \prod p_i^{q_i'}$, where $q_i' \leq aq_i$. However, since ψ and $c\psi+1$ are coprime, prime factorization of ψ should be of the form $\prod u_i^{v_i}$, where $\{u_i\} \cap \{p_i\} = \emptyset$. Therefore, clearly, λ and ψ are also coprime.

out loss of generality that $f_1 = \min\{f_i\}$, we can further write this summation as $\left(\sum_{i\neq 1}(-1)^{\alpha_i}N^{f_i-f_1} \pm 1\right)N^{f_1}$ where $\pm$ account for $(-1)^{\alpha_1}$. Note that if some of $f_i, i \neq 1$ are equal to f_1 so that the corresponding exponent $(f_i - f_1)$ becomes zero, those terms ought to have same sign as N^{f_1} (since there are no canceling terms). In other words if we denote $\xi = \{i|f_i = f_1\}$, then $\forall i \in \xi,\ \alpha_i = \alpha_1$. Therefore we can further write $\left(\sum_{i\neq 1}(-1)^{\alpha_i}N^{f_i-f_1} \pm 1\right)N^{f_1} = \left(\sum_{i\notin\xi}(-1)^{\alpha_i}N^{f_i-f_1} \pm |\xi|\right)N^{f_1}$. Similar to Corollary 3.3.11, it can be shown that the terms $\left(\sum_{i\notin\xi}(-1)^{\alpha_i}N^{f_i-f_1} \pm |\xi|\right)$ and N^{υ} are coprime. Thus if $N^{\upsilon}|\left(\sum_{i\notin\xi}(-1)^{\alpha_i}N^{f_i-f_1} \pm |\xi|\right)N^{f_1}$, we can conclude that N^{υ} should divide N^{f_1}. However since $f_1 = \min f_i$ we obtain that $N^{\upsilon}|N^{f_i}\ \forall i$. □

Having Corollary 3.3.12, we can now go back to the formula of (3.17) for entropy calculation. First we need a definition,

Definition 3.3.13 *Let A and B be two subsets of $\{1,\ldots,n\}$ s.t. $|A| = |B|$. Define,*

$$\delta_{A/B} = \sum_{i=1}^{|A|=|B|} \gamma_{A(i)B(i)}. \tag{3.42}$$

For example, for $A = \{2,3\}$ and $B = \{1,2\}$ we have, $\delta_{23/12} = \gamma_{21} + \gamma_{32}$.

Lemma 3.3.14 *For $\alpha, \beta \subseteq \{1,\ldots,n\}$ and $|\alpha| = |\beta|$, let $m_\alpha(\beta)$ denote the minor of M_α whose row and columns are indexed by α and β, respectively. Then $m_\alpha(\beta)$ can be expressed as,*

$$m_\alpha(\beta) = \sum_{\beta'\in\beta_{e,\alpha}} N^{\delta_{\alpha/\beta'}} - \sum_{\beta''\in\beta_{o,\alpha}} N^{\delta_{\alpha/\beta''}} \tag{3.43}$$

where $\beta_{e,\alpha} = \{\pi_e(\beta) \mid (\pi_e(\beta))_i \leq \alpha_i\}$ and $\beta_{o,\alpha} = \{\pi_o(\beta) \mid (\pi_o(\beta))_i \leq \alpha_i\}$ in which $\pi_e(\beta)$ and $\pi_o(\beta)$ represent the even and odd permutations of β, respectively, and $(.)_i$ denotes the i-th element of that set.

Proof: Using the Leibniz formula for determinants, a minor can be expressed as a polynomial of the form,

$$m_\alpha(\beta) = \sum_{\pi:\pi(\beta)_i<\alpha_i} \operatorname{sgn}(\pi) N^{\sum_{i=1}^{|\alpha|} \gamma_{\alpha_i,\pi(\beta)_i}} \tag{3.44}$$

where α_i and $\pi(\beta)_i$ denote the ith element of the sets α and $\pi(\beta)$, respectively, and π is any possible permutation on the set of column indices of the $|\alpha| \times |\alpha|$ minor. Note that only those permutations π are included for which $\pi(\beta)_i < \alpha_i$. This is due to the fact that we have assumed that the matrix is lower triangular. Since $\operatorname{sgn}(\pi)$ is $+1$ for even permutations and -1 for odd ones, we can further write (3.44) as,

$$m_\alpha(\beta) = \sum_{\pi_e:\pi_e(\beta)_i<\alpha_i} N^{\sum_i \gamma_{\alpha_i,\pi_e(\beta)_i}} - \sum_{\pi_o:\pi_o(\beta)_i<\alpha_i} N^{\sum_i \gamma_{\alpha_i,\pi_o(\beta)_i}} \tag{3.45}$$

where π_e and π_o correspond to odd and even permutations, respectively. Using Definition 3.3.13 in (3.45) concludes the proof. □

Example: Let $n = 4,\ \alpha = \{2,3,4\}$, and $\beta = \{123\}$ and assume we want to compute $m_{234}(123)$ that will be the following determinant,

$$m_{234}(123) = \begin{vmatrix} N^{\gamma_{21}} & N^{\gamma_{22}} & 0 \\ N^{\gamma_{31}} & N^{\gamma_{32}} & N^{\gamma_{33}} \\ N^{\gamma_{41}} & N^{\gamma_{42}} & N^{\gamma_{43}} \end{vmatrix}. \tag{3.46}$$

Note that $\beta_{e,\alpha} = \{(1,2,3),(2,3,1)\}$ and $\beta_{o,\alpha} = \{(1,3,2),(2,1,3)\}$. Therefore we will have

$$m_{234}(123) = N^{\delta_{234/123}} + N^{\delta_{234/231}} - N^{\delta_{234/132}} - N^{\delta_{234/213}}. \tag{3.47}$$

This can be easily verified by direct calculation as well.

Based on formula (3.17) to find the normalized joint entropy we have to calculate,

$$\underline{h}(\alpha) = |\alpha| - \frac{\log \gcd(\ m_\alpha(\beta) : \beta \subseteq \{1,\dots,n\}, |\beta| = |\alpha|\ )}{\log N} \tag{3.48}$$

which by Lemma 3.3.14, will be equal to,

$$\underline{h}(\alpha) = |\alpha| - \frac{\log \gcd(\sum_{\beta' \in \beta_{e,\alpha}} N^{\delta_{\alpha/\beta'}} - \sum_{\beta'' \in \beta_{o,\alpha}} N^{\delta_{\alpha/\beta''}},\ \forall \beta : |\beta| = |\alpha|\ )}{\log N}. \tag{3.49}$$

We may compute this using Corollary 3.3.12, however, as was also mentioned in that corollary, first we need to make sure that the terms will not cancel each other. In other words if it happens that for some $m_\alpha(\beta)$ and its respective $\beta_{e,\alpha}, \beta_{o,\alpha}$, we have $\delta_{\alpha/\beta'_i} = \delta_{\alpha/\beta''_j}$ for some $\beta'_i \in \beta_{e,\alpha},\ \beta''_j \in \beta_{o,\alpha}$ then those terms will cancel out in that $m_\alpha(\beta)$ and should not be included in the gcd calculation. When the number of terms gets large, deciding which terms will cancel and which will remain can become tricky. Therefore we need a mechanism that will allow us to simplify $m_\alpha(\beta)$ in those cases.

Lemma 3.3.15 *Consider the minor* $m_\alpha(\beta) = \sum_i N^{\delta_{\alpha/\beta'_i}} - \sum_j N^{\delta_{\alpha/\beta''_j}}$, *where we have* $\beta'_i \in \beta_{e,\alpha},\ \beta''_j \in \beta_{o,e}$. *Let* δ_{α/β'_i} *and* $\delta_{\alpha/\beta''_j}$ *be*

$$\mu(\delta_{\alpha/\beta'_i}) = \{\delta_{\alpha/\beta''_j} | \delta_{\alpha/\beta'_i} = \delta_{\alpha/\beta''_j}\} \tag{3.50}$$

$$\nu(\delta_{\alpha/\beta''_j}) = \{\delta_{\alpha/\beta'_i} | \delta_{\alpha/\beta''_j} = \delta_{\alpha/\beta'_i}\} \tag{3.51}$$

and define,

$$P_{\beta_{e,\alpha},\beta_{o,\alpha}}(\delta_{\alpha/\beta''_j}) \triangleq \begin{cases} \infty & \text{if } |\nu(\delta_{\alpha/\beta''_j})| \neq 0 \ \ \& \ \ |\nu(\delta_{\alpha/\beta''_j})| \geq |\mu(\nu(\delta_{\alpha/\beta''_j}))| \\ \infty & \text{if } |\nu(\delta_{\alpha/\beta''_j})| < |\mu(\nu(\delta_{\alpha/\beta''_j}))| \\ & \ \ \& \ \ \delta_{\alpha/\beta''_j} \in \mu(\nu(\delta_{\alpha/\beta''_j}))_{1:|\nu(\delta_{\alpha/\beta''_j})|} \\ 0 & \text{otherwise} \end{cases}. \tag{3.52}$$

Also define $P_{\beta_{e,\alpha},\beta_{o,\alpha}}(\delta_{\alpha/\beta_i'})$ *similarly by exchanging* μ *and* ν *in (3.52). Let* t *be an exponent, i.e.,* $t = \delta_{\alpha/\beta_i'}$ *or* $\delta_{\alpha/\beta_j''}$*. Then the term* N^t *does not cancel with any other terms in* $m_\alpha(\beta)$ *if* $P_{\beta_{e,\alpha},\beta_{o,\alpha}}(t) = 0$*.*

Proof: Let $\varepsilon_{ij} = 1$ if $\delta_{\alpha/\beta_i'} = \delta_{\alpha/\beta_j''}$ and 0 otherwise. Consider the bipartite graph with the left node set of $\{A_i\}$ assigned to $\delta_{\alpha/\beta_i'}$ and the right node set of $\{B_j\}$ assigned to $\delta_{\alpha/\beta_j''}$ (Fig 3.3) such that A_i and B_j are connected if and only if $\delta_{\alpha/\beta_i'} = \delta_{\alpha/\beta_j''}$ (i.e., $\varepsilon_{ij} = 1$). Therefore we can use (3.50) and (3.51) in a similar fashion, i.e., $\mu(A_i) = \{B_j | A_i = B_j\},\ \nu(B_j) = \{A_i | B_j = A_i\}$ to denote the neighbors of A_i and B_j, respectively. Similarly if S is a set, define $\mu(A_S) = \cup_{i\in S}\mu(A_i)$ and $\nu(B_S) = \cup_{j\in S}\nu(B_j)$. Since equality is a transitive property, we can readily see that, if $\mu(A_i) \cap \mu(A_j) \neq \varnothing$ then $\mu(A_i) = \mu(A_j)$ (the same property holds for $\nu(B_j)$ as well). This tells us that a valid set for ε_{ij} are the ones that partition the bipartite graph into bicliques. To simplify $m_\alpha(\beta)$, we can decide for each term $N^{\delta_{\alpha/\beta_i'}}$ (or $N^{\delta_{\alpha/\beta_j''}}$) if it will cancel with one of the $N^{\delta_{\alpha/\beta_j''}}$ (or $N^{\delta_{\alpha/\beta_i'}}$). On the bipartite graph this is equivalent to finding a matching where the matching nodes are the ones whose corresponding terms cancel out in $m_\alpha(\beta)$. Here is a simple method to find such matching.

Assume we want to find out if B_j gets matched with any of the A_i's (the argument for A_i would be similar). Let C be the biclique within the bipartite graph to which B_j belongs. By definition the number of left nodes of C is $|\nu(B_j)|$ and the number of right nodes of C is $|\mu(\nu(B_j))|$. The following can be deduced,

1. $|\nu(B_j)| = \varnothing$: This means that B_j does not match with any of the A_i's.

2. $|\nu(B_j)| \neq \varnothing$ and $|\nu(B_j)| \geq |\mu(\nu(B_j))|$: In this case, since the number of A_i's that match with B_j is more than the number of right nodes of S, we can be sure that B_j will get paired with one of the A_i's on the left of C and as a result we can discard B_j.

3. $|\nu(B_j)| < |\mu(\nu(B_j))|$: In this case not all the right nodes of C will pair with its

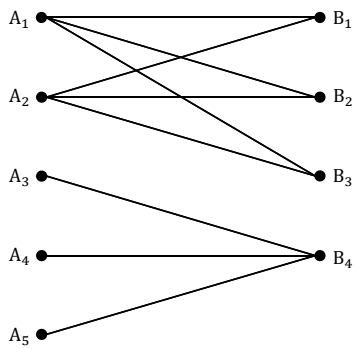


Figure 3.3: Example of a graph representing equality relations between $\delta_{\alpha/\beta_i'}$ and $\delta_{\alpha/\beta_j''}$. Note that A_i, and B_j represent $\delta_{\alpha/\beta_i'}$, and $\delta_{\alpha/\beta_j''}$, respectively.

left nodes A_i and therefore we choose to pair B_j (discard) with a node on the left of C, only if it is among the first $|\nu(B_j)|$ nodes on the right-hand side of C (first $|\nu(B_j)|$ of $\mu(\nu(B_j))$).

Therefore we can define a parameter P as,

$$P_{\{A_{i'}\},\{B_{j'}\}}(B_j) \triangleq \begin{cases} \infty & \text{if } |\nu(B_j)| \neq 0 \ \ \& \ \ |\nu(B_j)| \geq |\mu(\nu(B_j))| \\ \infty & \text{if } |\nu(B_j)| < |\mu(\nu(B_j))| \ \ \& \ \ B_j \in \mu(\nu(B_j))_{1:|\nu(B_j)|} \\ 0 & \text{otherwise} \end{cases} \tag{3.53}$$

which is 0 only when the node is not matched with another node or equivalently its corresponding term does not cancel in $m_\alpha(\beta)$. Note that for $P(A_i)$ the role of μ and ν should be exchanged in the above. Replacing $A_i,\ B_j$ with their equivalent $\delta_{\alpha/\beta_i'},\ \delta_{\alpha/\beta_j''}$ gives (3.52). □

Example: An example of such a bipartite graph can be seen in Fig. 3.3. Note that, e.g., $\mu(A_1) = \{B_1, B_2, B_3\},\ \nu(B_3) = \{A_1, A_2\}$ and $\mu(\nu(B_3)) = \{B_1, B_2, B_3\}$. Based on (3.52) we obtain that $P_{\{A_i\},\{B_j\}}(A_1) = P_{\{A_i\},\{B_j\}}(A_2) = P_{\{A_i\},\{B_j\}}(A_3) = \infty$

and $P_{\{A_i\},\{B_j\}}(A_4) = P_{\{A_i\},\{B_j\}}(A_5) = 0$. Also $P_{\{A_i\},\{B_j\}}(B_1) = P_{\{A_i\},\{B_j\}}(B_2) = P_{\{A_i\},\{B_j\}}(B_4) = \infty$ and $P_{\{A_i\},\{B_j\}}(B_3) = 0$.

Theorem 3.3.16 (Entropy in terms of γ_{ij}'s) *Let $\alpha \subseteq \{1, \ldots, n\}$. The lattice-derived normalized joint entropy $\underline{h}_\alpha$ is expressed as,*

$$\underline{h}(\alpha) = |\alpha| - \min\Big(\delta_{\alpha/\tilde{\beta}} + P_{\beta_{e,\alpha},\beta_{o,\alpha}}(\delta_{\alpha/\tilde{\beta}}), \quad \forall \tilde{\beta} \in (\beta_{e,\alpha} \cup \beta_{o,\alpha}),\ \forall \beta :\ |\beta| = |\alpha|\Big) \tag{3.54}$$

where $\delta_{\alpha/\tilde{\beta}}$ is as defined in (3.42) and $P_{\beta_{e,\alpha},\beta_{o,\alpha}}(\delta_{\alpha/\tilde{\beta}})$ as defined in (3.52).

Proof: Based on formula (3.17) the normalized joint entropy can be written as,

$$\underline{h}(\alpha) = |\alpha| - \frac{\log \gcd(\ m_\alpha(\beta),\ |\beta| = |\alpha|)}{\log N}. \tag{3.55}$$

Using Lemma 3.3.14, this can further be written as,

$$\underline{h} = |\alpha| - \frac{\log \gcd(\sum_i N^{\delta_{\alpha/\beta_i'}} - \sum_j N^{\delta_{\alpha/\beta_j''}},\ \beta_i' \in \beta_{e,\alpha},\ \beta_j'' \in \beta_{o,\alpha}\ \ \forall \beta :\ |\beta| = |\alpha|)}{\log N}. \tag{3.56}$$

By using Corollary 3.3.12 and Lemma 3.3.15, this can be written as,

$$\underline{h} = |\alpha| - \frac{\log \gcd(N^{\delta_{\alpha/\tilde{\beta}}},\ \tilde{\beta} \in (\beta_{e,\alpha} \cup \beta_{o,\alpha}),\ P_{\beta_{e,\alpha},\beta_{o,\alpha}}(\delta_{\alpha/\tilde{\beta}}) = 0)}{\log N}. \tag{3.57}$$

Assuming $\gcd\{N^{\delta_{\alpha/\tilde{\beta}}}\} = N^{\min(\delta_{\alpha/\tilde{\beta}})}$, and noting from (3.52) that when P is not 0 it is infinity we obtain the result of the theorem. □

Let Δ_n denote the space obtained from equations (3.26) and (3.54), i.e., the space of entropy vectors of n random variables obtained from lattice-generated quasi-uniform distributions. Clearly Δ_n is a nonconvex space obtained from the union of some polytopic regions. In fact each fixed ordering of σ_k and τ_l's (and therefore γ_{ij}'s) with respect to each other, defines a possibly new set of linear equations for h_α in terms of γ_{ij} and therefore results in a polytope. We denote the closure of the convex

hull of all these polytopic regions by $\overline{\text{con}}(\Delta_n)$.

3.3.3 Characterizing the Lattice-Based Entropy Region

Theorem 3.3.17 (An Inner Region for Entropic Vectors) $\overline{con}(\Delta_n) \subseteq \bar{\Omega}_n^*$ *where* $\overline{con}(\cdot)$ *represents the convex closure.*

Proof: Follows straightforwardly from the convexity of $\bar{\Omega}_n^*$. □

This inner region is a polytope,

Theorem 3.3.18 *The region* $\overline{con}(\Delta_n)$ *is a polytope for all* n*.*

Proof: Each ordering of γ_{ij}'s (e.g. $\gamma_{11} \leq \gamma_{21} \leq \gamma_{22}$, $\gamma_{21} + \gamma_{32} \leq \gamma_{22} + \gamma_{31}$, etc.) defines a polytope region for the set of entropy vectors. Δ_n is the convex hull of all these polytopes and therefore a polytope itself. □

As the number of random variables grows, the number of polytopic regions of Δ_n can grow very large and therefore computing the innerbound for Ω_n^* based on Theorem 3.3.17 becomes tricky. The following theorem gives a computable method for obtaining an inner region of $\overline{\text{con}}(\Delta_n)$ and therefore $\bar{\Omega}_n^*$.

Theorem 3.3.19 *Consider the hypercube of* $0 \leq \gamma_{ij} \leq 1, \forall 1 \leq j \leq i \leq n$ *whose faces are chopped off by* γ_{ij} *constraints in (3.26). Each corner point of this chopped hypercube is a valid* γ *point and its corresponding entropy vector can be easily calculated from (3.54). Let* $\mathfrak{R}_{\mathfrak{n}}$ *denote the convex hull of all these entropy vectors. Then* $\mathfrak{R}_{\mathfrak{n}} \subseteq \overline{con}(\Delta_n) \subseteq \bar{\Omega}_n^*$.

Proof: While the statement may seem rather obvious, in the following we explain the rationale for choosing the corner points for calculating the innerbound $\mathfrak{R}$. Note that each polytopic region of Δ_n is obtained from the corresponding entropies of the γ region defined by (3.26) and a specific order relation for γ_{ij}'s. If we add the $\gamma_{ij} \leq 1$ for $i \neq j$ to the constraints in (3.26), then we obtain the mentioned chopped off

unit hypercube. Each order relation for γ_{ij}'s, shown by $O(\gamma)$ defines a homogenous inequality and therefore a cone in the γ space which intersects the unit-chopped off hypercube in a set of points say, $B(O(\gamma))$ on its boundary. Since for the specific $O(\gamma)$, the entropies are determined linearly from γ_{ij}'s, the corresponding polytopic entropy region of Δ_n will be obtained from the convex hull of the corresponding entropy vectors of $B(O(\gamma))$. It follows that each polytopic region of Δ_n is obtained from the convex hull of a collection of boundary points on the chopped off unit hypercube and therefore the convex hull of all the entropy vectors corresponding to *all* boundary points of the chopped off hypercube in γ region will give a fair innerbound for $\overline{\text{con}}(\Delta_n)$. Since it is impossible to compute the entropies for all the boundary points of the chopped off γ region, we will only consider the corner points of the chopped off cube and compute the corresponding entropy vectors. The convex hull of these entropies is $\mathfrak{R}_{\mathfrak{n}}$ and is an inner bound for $\overline{\text{con}}(\Delta_n)$ and $\bar{\Omega}_n^*$. □

Remark: Note that in obtaining $\mathfrak{R}_{\mathfrak{n}}$ in Theorem 3.3.19, a couple of compromises have been made. First we have assumed that $\gamma_{ij} \leq 1$ which is not necessarily true for $i \neq j$. Moreover we have replaced the convex hull of the entropy vectors corresponding to all the boundary points of γ region by just the corner points. And last but not least, one should be aware that in some cases $O(\gamma)$ may involve some strict inequalities in terms of γ's and therefore its corresponding entropy region say $h(O(\gamma))$ will not be closed either. To obtain an accurate convex closure of the lattice entropy region, this fact should also be considered, whereas in computing $\mathfrak{R}_{\mathfrak{n}}$ in Theorem 3.3.19 this is ignored.

In Section 3.4 we perform the explicit constructions of the lattice region and obtain $\mathfrak{R}_{\mathfrak{n}}$ (and $\overline{\text{con}}(\Delta_n)$ where possible) for $n = 2, 3, 4$, and 5 random variables.

3.3.4 Vectorized Lattices

The entropy region inner bound generated by lattice-derived distributions can be expanded by considering generalized vector lattices,

Definition 3.3.20 *Let M be a block matrix where each block is a $q \times q$ square matrix and M is of size $nq \times nq$. Moreover assume that M satisfies the quasi-uniformity condition in Lemma 3.3.6 that NM^{-1} has integer entries. We call M a vectorized lattice-generating matrix whose corresponding entropies following (3.17) will be calculated as,*

$$h_\alpha = |\alpha| - \frac{\log \gcd(\text{ all q}|\alpha| \times \text{q}|\alpha| \text{ minors of } \text{M}_{[\alpha]})}{q \log N} \tag{3.58}$$

where $M_{[\alpha]}$ denotes the $|\alpha|q \times nq$ submatrix of M whose block rows are indexed by α.

When $q = 1$, as in the previous sections, the lattice is scalar and each column of generating matrix is a generating vector of the lattice in n dimensions. Although we will not delve into the space of vector lattices here, we will show later in Section 3.4 that some entropy vectors which do not fall in the entropy region obtained from scalar lattices can be obtained from vector lattices.

3.3.5 Connection to Groups and Linear Representable Region

As discussed previously, the entropy region of n random variables can be obtained from the region of group-derived entropies [CY02]. Here we show that our lattice construction can in fact be viewed as a quasi-uniform construction corresponding to an Abelian group.

Theorem 3.3.21 *Lattice construction is the quasi-uniform distribution obtained from*

an Abelian group.

Proof: *Consider a particular lattice distribution with a relative $n \times n$ matrix generator M. It is straightforward to see that the lattice-generated points inside the n dimensional hypercube of length N together with the addition* $\bmod N$ *operation form an Abelian group G. For $i = 1, \ldots, n$ consider the $X_i = 0$ hyperplane (note that since the origin is always a lattice point this hyperplane includes at least one lattice point and is nonempty). It is easy to see that the set of lattice points on this hyperplane forms a subgroup, say G_i of G. The cosets of G_i will be all nonempty hyperplanes of the form $X_i = c_{ij}$, for some constant $1 \leq c_{ij} \leq N$. It is then easy to see that the lattice-generated distribution coincides with the quasi-uniform distribution derived from group G and its subgroups $G_1, \ldots, G_n$. Hence for any $\alpha \subseteq \{1, \ldots, n\}$ the joint entropy of $X_\alpha = \{X_i,\ i \in \alpha\}$ will simply be the* $\log$ *of the number of nonempty hyperplanes $X_\alpha = c_{\alpha,j}$. Each hyperplane $X_\alpha = c_{\alpha,j}$ is an intersection of cosets of $G_i,\ i \in \alpha$ and therefore a coset of $\cap_{i \in \alpha} G_i$. As a result the number of hyperplanes $X_\alpha = c_{\alpha,j}$ is the number of cosets of $\cap_{i \in \alpha} G_i$ in G. Therefore $h_\alpha = \log \frac{|G|}{|G_\alpha|}$.* □

In [Cha07a], it is shown that Abelian group-derived entropies all satisfy the Ingleton inequality. We conclude that the region of lattice-derived entropies is an inner-region of Γ_n^* bounded by the Ingleton inequality.

On the other hand it is also known that linearly representable entropy vectors satisfy the Ingleton inequality. To make a comparison to the lattice construction, note that any linear representable vector is an entropy vector constructed as stated in the following theorem,

Theorem 3.3.22 (Linear representables and entropy vectors) *[YLCZ06] Let g be a $2^n - 1$ dimensional representable vector and let the set of p dimensional vectors $\{v_1, \ldots, v_n\}$ over GF(q) form a representation for it such that $g_\alpha = rank(\oplus_{i \in \alpha} v_i)$,*

then $\log(q)g$ *is an entropy vector of random variables* $X_1, \ldots, X_n$ *where,*

$$\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = \left(\begin{array}{c} v_1 \\ \hline \vdots \\ \hline v_n \end{array}\right) \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix}. \tag{3.59}$$

Proof: For a particular $\tilde{X}_i$, let $S^i_{\tilde{X}_i}$ denote the set of solutions (a's) that yield $\tilde{X}_i$. In particular let S^i_0 denote the set of solutions that yield $X_i = 0$. If $\tilde{a}$ is a particular solution, then clearly $S^i_{\tilde{X}_i} = \tilde{a} + S^i_0$. Now if we let G be the group of all points in $(\mathrm{GF}(q))^p$ together with the vector addition operation over $GF(q)$, then clearly $G_i = S^i_0,\ i = 1, \ldots, n$ form a set of n subgroups for G and $S^i_{\tilde{X}_i}$ will be the respective cosets. Moreover for any $\alpha \subseteq \{1, \ldots, n\}$, we can write the above equation as,

$$\begin{pmatrix} X_\alpha \\ X_{\alpha^c} \end{pmatrix} = \begin{pmatrix} V_\alpha \\ V_{\alpha^c} \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix} \tag{3.60}$$

and likewise define $S^\alpha_{\tilde{X}_\alpha}$ as the set of a's that yield $\tilde{X}_\alpha$ and in particular denote the null space of V_α by S^α_0. Note that since $S^\alpha_{\tilde{X}_\alpha}$ is a coset of S^α_0 we have $|S^\alpha_{\tilde{X}_\alpha}| = |S^\alpha_0|$ which is independent of X_α. It is then clear that $X_1, \ldots, X_n$ will be in a one-one correspondence with quasi-uniform distribution derived from the group G along with its subgroups $G_1, \ldots, G_n$ and hence $X_1, \ldots, X_n$ will have a uniform distribution over the range of $\left[v_1^T | \ldots | v_n^T\right]$. Therefore the joint entropy of a set of random variables indexed by $\alpha \subseteq \{1, \ldots, n\}$ is obtained from $\log \frac{|G|}{|\cap_{i\in\alpha} S^i_0|}$. However $\cap_{i\in\alpha} S^i_0 = S^\alpha_0$ and we have $|S^\alpha_0| = q^{\mathrm{nullity}(V_\alpha)} = q^{(p - \mathrm{rank}(V_\alpha))}$. Noting that $|G| = q^p$ we obtain that $h_\alpha = \mathrm{rank}(V_\alpha) \cdot \log q$. □

Due to the similarity of the lattice-generated distributions and the distribution of representable vectors, one might suspect that the two regions are equal. However comparison of the two regions becomes tricky for a number of reasons. First note

that while in the linear representable case the field size q can be either a prime or a power of a prime, in the lattice constructions, there is no such constraint. Moreover as opposed to the linear representable case where for any fixed field size the entropy (rank) can be calculated, in the entropy-derivation of the lattice construction we have assumed that the alphabet size N is arbitrarily large. In fact if we fix N to some known value then the set of γ's that will be consistent with the derivations will be restricted. Therefore if a vector is linearly representable over some finite field q there can be no guarantee that the vector may be constructed via the lattice over alphabet size q. An example is the entropy vector of 4 random variables s.t. $h_i = 1,\ h_{ij} = h_{ijk} = h_{1234} = 2$, which is linearly representable over a field with odd characteristic, however is not constructible with the scalar lattice. This vector can however be constructed by a vector lattice (see Subsection 3.4.3). The other difficulty is that in the lattice construction, the operation is always addition module N, whereas in the linear representable case when the field size is a power of a prime, e.g., $q = p^r$, operations are not simply module q. This however can be partly fixed by replacing entries of the lattice generator matrix by $r \times r$ blocks with elements over $GF(p)$ representing the elements of the field and the elements of the coefficient vector z (see (3.11)) by sub-vectors of size r and elements over $GF(p)$ also representing the elements of the field[1]. Nonetheless since we can not fix N to some prime p, again we cannot simply argue that any linearly representable vector over $q = p^r$ can be achieved via the lattice construction. Therefore the comparison of the regions is somewhat difficult. What can be said for sure however, is that both regions are inner-regions of Γ_n^* and they satisfy the Ingleton inequality. For 4 random variables (as will be discussed in Subsection 3.4.3) both regions turn out to be equal and defined by

[1]In other words, defining a vector and a matrix representation for each element of the field and operations over $GF(p)$ among them such that they respect the addition and multiplication of the field, e.g., for $Gf(4)$ such representation would be, $0 : \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ $1 : \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ $\omega : \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ $\omega + 1 : \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$.

the Shannon and the Ingleton inequalities.

3.4 Explicit Constructions

Now that we have described the general construction, we will study the explicit results for $n = 2, 3, 4$, and 5 random variables.

3.4.1 Two Random Variables

As described in Subsections 3.3.1 and 3.3.2, we can assume that the 2-by-2 lattice-generating matrix is lower triangular of the following form,

$$M = \begin{bmatrix} M_{11} & 0 \\ M_{21} & M_{22} \end{bmatrix} = \begin{bmatrix} N^{\gamma_{11}} & 0 \\ N^{\gamma_{21}} & N^{\gamma_{22}} \end{bmatrix}. \tag{3.61}$$

We first need to enforce the condition that the generated distribution be quasi-uniform. From Lemma 3.3.6 this means that the matrix

$$NM^{-1} = \begin{bmatrix} N^{1-\gamma_{11}} & 0 \\ -N^{1-\gamma_{11}-\gamma_{22}+\gamma_{21}} & N^{1-\gamma_{22}} \end{bmatrix} \tag{3.62}$$

must have integer entries. Since N is large enough this implies the inequalities,

$$\gamma_{11} \leq 1 \quad , \quad \gamma_{22} \leq 1 \quad , \quad \gamma_{11} + \gamma_{22} \leq 1 + \gamma_{21}. \tag{3.63}$$

Note that these are the inequalities obtained from 3.26. Using Lemma 3.3.7 or equivalently Theorem 3.3.16, the corresponding entropies are readily seen to be,

$$\begin{aligned} \underline{h}_1 &= 1 - \gamma_{11} \\ \underline{h}_2 &= 1 - \min(\gamma_{21}, \gamma_{22}) \\ \underline{h}_{12} &= 2 - \gamma_{11} - \gamma_{22}. \end{aligned} \tag{3.64}$$

Thus the space Δ_2 is described by (3.64) along with the constraints (3.63). The region Δ_2 may not be convex, due to the $\min(\cdot)$ operator in $\underline{h}_2$. However, it is not hard to show that the convex hull of (3.63–3.64) is,

$$\begin{cases} \underline{h}_1 = 1 - \gamma_{11} \ , \ \underline{h}_2 = 1 - \gamma_{21} \ , \ \underline{h}_{12} = 2 - \gamma_{11} - \gamma_{22} \\ 0 \leq \gamma_{11}, \gamma_{22} \leq 1 \ , \ 0 \leq \gamma_{21} \leq \gamma_{22} \ , \ \gamma_{11} + \gamma_{22} \leq 1 + \gamma_{21} \end{cases} \tag{3.65}$$

Theorem 3.4.1 (Lattice entropy region for $n = 2$) *con*$(\Delta_2) = \bar{\Omega}_2^*$ *where con is the convex hull operation.*

Proof: Any $\underline{h}_{ij}$ obtained from (3.65) is a normalized entropy vector by construction. Conversely, for any entropy vector satisfying $0 \leq \underline{h}_1, \underline{h}_2 \leq 1$ and $\underline{h}_1, \underline{h}_2 \leq \underline{h}_{12} \leq \underline{h}_1 + \underline{h}_2$, a valid set of γ_{ij}s from (3.65) can be found and therefore any entropy satisfying these conditions will belong to the region of normalized entropies. Therefore $\text{con}(\Delta_2) = \bar{\Omega}_2^*$ and this region will be given by (3.65). □

Remark: Theorem 3.4.1 also proves that for 2 random variables, any 3-dimensional entropy vector $(h_1,\ h_2,\ h_{12}) \in \bar{\Gamma}_n^*$ when normalized by $\max(h_1, h_2)$ will belong to $\bar{\Omega}_2^*$. For comparison we can also find $\mathfrak{R}_2$ based on Theorem 3.3.19.

Theorem 3.4.2 $\mathfrak{R}_2 = \overline{\Omega}_2^*$

Proof: For $n = 2$ there are three γ_{ij} namely γ_{11}, γ_{21}, and γ_{22}. Therefore the hypercube will be the 3-dimensional unit hypercube, $0 \leq \gamma_{11} \leq 1, \quad 0 \leq \gamma_{21} \leq 1, \quad 0 \leq \gamma_{22} \leq 1$

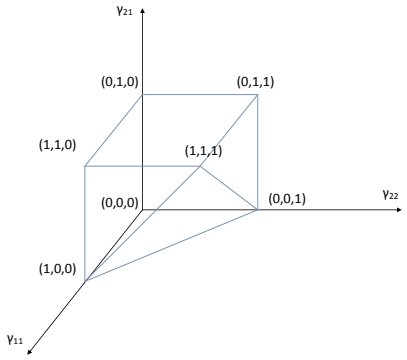


Figure 3.4: γ corner points for $n = 2$

which will be chopped off by the constraints of (3.26) which are, $\gamma_{11} \leq 1, \quad \gamma_{22} \leq 1, \quad \gamma_{11} + \gamma_{22} - \gamma_{21} \leq 1$. The resulting region is a 3-dimensional object whose corner points can be easily computed (Fig. 3.4). These points, along with their corresponding entropy vectors are shown in Table 3.1. It can be shown that the convex hull of these entropy points gives the following region,

$$\begin{aligned} 0 \leq \underline{h}_i \leq 1 \quad & i = 1, 2 \\ \underline{h}_i \leq \underline{h}_{12} \leq \underline{h}_1 + \underline{h}_2 \quad & i = 1, 2 \end{aligned} \tag{3.66}$$

which is known to be equal to $\overline{\Omega}_2^*$. It follows that $\mathfrak{R}_2 = \overline{\Omega}_2^*$. □

Table 3.1: Entropy region corner points for 2 random variables obtained through the lattice construction (γ corner points and the corresponding $\underline{h}$)

$(\gamma_{11}, \gamma_{21}, \gamma_{22})$	$(\underline{h}_1, \underline{h}_2, \underline{h}_{12})$
(1, 1, 1)	(0, 0, 0)
(1, 0, 0) (1, 1, 0)	(0, 1, 1)
(0, 1, 1)	(1, 0, 1)
(0, 0, 0) (0, 1, 0)	(1, 1, 2)
(0, 0, 1)	(1, 1, 1)

3.4.2 Three Random Variables

Again, without loss of generality we may assume

$$M = \begin{bmatrix} N^{\gamma_{11}} & 0 & 0 \\ N^{\gamma_{21}} & N^{\gamma_{22}} & 0 \\ N^{\gamma_{31}} & N^{\gamma_{32}} & N^{\gamma_{33}} \end{bmatrix}. \tag{3.67}$$

Insisting on quasi-uniformity by forcing the elements of $M^{-1}N$ to be integers (Lemma 3.3.6) yields the linear constraints given by (3.26),

$$\begin{aligned} &0 \leq \gamma_{ij} \leq 1 \\ &\gamma_{ii} + \gamma_{jj} - \gamma_{ij} \leq 1 \quad i > j \\ &\gamma_{11} + \gamma_{22} + \gamma_{33} - \gamma_{21} - \gamma_{32} \leq 1. \end{aligned} \tag{3.68}$$

Using Theorem 3.3.16 all the joint entropies for 3 variables can be easily computed. However in order to show how Theorem 3.3.16 follows from Lemma 3.3.7 or equivalently (3.17), we explain the extraction of $\underline{h}_{23}$ in the following. By (3.17), we have,

$$\underline{h}_{23} = 2 - \frac{1}{\log N} \log \gcd(N^{\gamma_{21}+\gamma_{32}} - N^{\gamma_{22}+\gamma_{31}}, N^{\gamma_{21}+\gamma_{33}}, N^{\gamma_{22}+\gamma_{33}}). \tag{3.69}$$

Let $\sigma = \gamma_{21} + \gamma_{32}$, $\tau = \gamma_{22} + \gamma_{31}$, $\omega_1 = \gamma_{33} + \gamma_{21}$, and $\omega_2 = \gamma_{33} + \gamma_{22}$. We can rewrite $\underline{h}_{23}$ as,

$$\underline{h}_{23} = 2 - \frac{1}{\log N} \log \gcd(N^{\sigma} - N^{\tau}, N^{\omega_1}, N^{\omega_2}).$$

In calculating $\gcd(N^{\sigma} - N^{\tau}, N^{\omega_1}, N^{\omega_2})$, two cases are possible,

1. If $\sigma = \tau$, then $\gcd(N^{\sigma} - N^{\tau}, N^{\omega_1}, N^{\omega_2}) = \gcd(N^{\omega_1}, N^{\omega_2}) = N^{\min(\omega_1,\omega_2)}$ where the latter equality is justified for large N as explained before.

2. If $\sigma \neq \tau$, then there is a positive rational υ such that $\gcd(N^\sigma - N^\tau, N^{\omega_1}, N^{\omega_2}) = N^\upsilon$. Therefore, N^υ divides $N^\sigma - N^\tau$ or if without loss of generality $\tau \leq \sigma$, then $N^\upsilon | N^\tau(N^{\sigma-\tau} - 1)$. However, based on Lemma 3.3.10, N^υ and $N^{\sigma-\tau} - 1$ are coprime and therefore we conclude that N^υ divides N^τ. Moreover since $\tau \leq \sigma$, by making N large enough we can make N^τ divide N^σ. As a result N^υ will divide N^σ, N^τ and also N^{ω_1} and N^{ω_2}. In other words, $N^\upsilon = \gcd(N^\sigma, N^\tau, N^{\omega_1}, N^{\omega_2})$. As before, for large enough N, we will have $N^\upsilon = N^{\min(\sigma,\tau,\omega_1,\omega_2)}$.

Now we can write the expressions for all the normalized joint entropies:

$$
\begin{aligned}
&\underline{h}_1 = 1 - \gamma_{11} \\
&\underline{h}_2 = 1 - \min(\gamma_{21}, \gamma_{22}) \\
&\underline{h}_3 = 1 - \min(\gamma_{31}, \gamma_{32}, \gamma_{33}) \\
&\underline{h}_{12} = 2 - \gamma_{11} - \gamma_{22} \\
&\underline{h}_{13} = 2 - \gamma_{11} - \min(\gamma_{32}, \gamma_{33}) \\
&\underline{h}_{23} = \begin{cases} \underline{\gamma_{21} + \gamma_{32} \neq \gamma_{22} + \gamma_{31}:} & 2 - \min(\gamma_{21} + \gamma_{32}, \gamma_{22} + \gamma_{31}, \gamma_{33} + \min(\gamma_{21}, \gamma_{22})) \\ \underline{\gamma_{21} + \gamma_{32} = \gamma_{22} + \gamma_{31}:} & 2 - \gamma_{33} - \min(\gamma_{21}, \gamma_{22}) \end{cases} \\
&\underline{h}_{123} = 3 - \gamma_{11} - \gamma_{22} - \gamma_{33}.
\end{aligned}
\tag{3.70}
$$

The space Δ_3 is defined by (3.68) and (3.70) which is clearly nonconvex. On the other hand, each ordering of γ_{ij}'s (e.g., $\gamma_{21} \leq \gamma_{22}, \gamma_{21} + \gamma_{32} \leq \gamma_{22} + \gamma_{31}$, etc.) results in a polytope that is at most six-dimensional (since there are six γ_{ij} parameters). For $n = 3$, there are 30 regions, all of which are shown in Table 3.3 in the Appendix.

Obtaining the convex hull of all these regions seems rather hard. Therefore we pursue the approach of Theorem 3.3.19 to find an inner bound. Using the software package PORTA [POR], we obtain the corner points of the 6-dimensional hypercube $0 \leq \gamma_{ij} \leq 1$ intersected with the constraints of (3.68). Having these corner points we can easily compute their corresponding entropy vectors from (3.70). There are 44

such corner points all of which, along with their corresponding entropy vectors, are shown in Table 3.4 in the Appendix. From the table, it can be seen that they result in 16 different entropy vectors.

Theorem 3.4.3 $\mathfrak{R}_3 = \overline{\Omega}_3^*$

Proof: The convex hull of the 16 vectors of Table 3.4 yields $\mathfrak{R}_3$ and is computed via the package [POR]. The convex hull is obtained to be,

$$\begin{aligned} &0 \leq \underline{h}_i \leq 1 \quad i = 1, 2, 3 \\ &\underline{h}_{ij} \leq \underline{h}_{123} \quad i, j \in \{1, 2, 3\} \\ &\underline{h}_i + \underline{h}_{123} \leq \underline{h}_{ij} + \underline{h}_{ik} \quad i, j, k \in \{1, 2, 3\} \end{aligned} \tag{3.71}$$

which in fact implies that every normalized entropy vector that satisfies (3.71) is achievable by lattice-derived entropies. It follows that the convex hull is 7-dimensional and $\mathfrak{R}_3 = \overline{\Omega}_3^*$. □

Remark: Theorem 3.4.3 also proves that any entropy vector of 3 random variables $h \in \bar{\Gamma}_3^*$ when normalized by $\max(h_1, h_2, h_3)$ will belong to $\bar{\Omega}_3^*$.

Theorem 3.4.4 *$con(\Delta_3) = \overline{\Omega}_3^*$ where con refers to the convex hull.*

Proof: It follows immediately from Theorems 3.3.19 and 3.4.3. Moreover note that if we pick wisely among those γ's that lead to the same entropy vector in Table 3.4, then all 16 entropy vectors can be attributed solely to 5 of the 30 regions of Table 3.3, namely, regions 1, 2, 3, 4, and 23 in Table 3.3 [1]. This shows that the convex hull of all the 30 regions of Table 3.3 also give $\overline{\Omega}_3^*$ and again proves the statement. □

Corollary 3.4.5 *Convex cone of Δ_3 gives $\bar{\Gamma}_3^*$.*

Proof: This readily follows from the fact that $con(\Delta_3) = \bar{\Omega}_3^*$ and that $\bar{\Gamma}_3^* = ray(\bar{\Omega}_3^*)$ (see Theorem 2.2.4). However this result can also be proved independently as follows.

[1] Instead of region 23 any of the regions 13, 14, or 18 could also work.

We show that 8 of the 16 vectors of Table 3.4 are enough to show that the convex cone of Δ_3 generates $\bar{\Gamma}_3^*$. First note that since, by construction, all vectors in con(Δ_3) are entropic, clearly con(Δ_3) $\subseteq \bar{\Omega}_3^*$ and therefore convex cone of Δ_3 is a subset of $\bar{\Gamma}_3^*$. To prove the other direction consider the region defined by,

$$\begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_{12} \\ h_{23} \\ h_{31} \\ h_{123} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 & 1 & 2 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 2 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 2 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \\ k_3 \\ k_4 \\ k_5 \\ k_6 \\ k_7 \\ k_8 \end{bmatrix}, \tag{3.72}$$

where $k_i \geq 0$. Each column vector in the matrix on the RHS can be seen to be generated by a lattice-generated distribution as it belongs to Table 3.4. Therefore the convex cone of these vectors must be a subset of $con(\Delta_3)$. If we write the above matrix equation as

$$h = \begin{bmatrix} A & a \end{bmatrix} \begin{bmatrix} k \\ k_8 \end{bmatrix} = Ak + ak_8,$$

then since the first seven columns of the matrix on the RHS is invertible, we can further write, $k = A^{-1}h - A^{-1}ak_8 \geq 0$ where we are enforcing $k \geq 0$ as we would like to form the convex cone of the columns of the matrix. Computing A^{-1} and $A^{-1}ak_8$,

yields

$$\begin{bmatrix} 0 & 0 & 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 & 0 & 0 & 1 \\ -1 & 0 & 0 & 1 & 0 & 1 & -1 \\ 0 & -1 & 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & -1 & 0 & 1 & 1 & -1 \\ 1 & 1 & 1 & -1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_{12} \\ h_{23} \\ h_{31} \\ h_{123} \end{bmatrix} \geq \begin{bmatrix} 0 \\ 0 \\ 0 \\ k_8 \\ k_8 \\ k_8 \\ -k_8 \end{bmatrix}. \tag{3.73}$$

The point is to show that for any entropic vector $h \in \bar{\Gamma}_3^*$, one can find a non-negative k_8 such that the inequality (3.73) will be satisfied. The inequality for the first 3 rows is clearly satisfied. For the next three rows it is satisfied provided that,

$$k_8 \leq \min_{i,j,k}(-h_i + h_{ij} + h_{ki} - h_{ijk}), \tag{3.74}$$

and for the last row if,

$$k_8 \geq -\sum_i h_i + \sum_{i,j} h_{ij} - h_{ijk}. \tag{3.75}$$

It is straightforward to show that the upper bound on k_8 exceeds the lower bound and so the region for k_8 is non-empty. Furthermore, by the submodularity of the entropy function, the upper bound is non-negative, which implies that a non-negative k_8 can always be found. This concludes the proof. □

Remark: It is interesting to note that, had we not included the vector $(1, 1, 1, 2, 2, 2, 2)$ among the 8 vectors, the resulting region would have the property that $I(X_1; X_2; X_3) = \sum h_i - \sum h_{ij} + h_{123} \geq 0$, which is not necessarily true for 3 variables. Likewise exclusion of $(1, 1, 1, 1, 1, 1, 1)$ results in $I(X_1; X_2; X_3) \leq 0$ and therefore it is the combination of all 8 vectors that give the whole entropy region for 3 random variables.

3.4.3 Four Random Variables

For four random variables, we assume

$$M = \begin{bmatrix} N^{\gamma_{11}} & 0 & 0 & 0 \\ N^{\gamma_{21}} & N^{\gamma_{22}} & 0 & 0 \\ N^{\gamma_{31}} & N^{\gamma_{32}} & N^{\gamma_{33}} & 0 \\ N^{\gamma_{41}} & N^{\gamma_{42}} & N^{\gamma_{43}} & N^{\gamma_{44}} \end{bmatrix}. \tag{3.76}$$

The conditions from Lemma 3.3.6 translate to:

$$\begin{aligned} &0 \leq \gamma_{ii} \leq 1, \\ &\gamma_{ii} + \gamma_{jj} - \gamma_{ij} \leq 1, \quad i > j \\ &\gamma_{ii} + \gamma_{jj} + \gamma_{kk} - \gamma_{ij} - \gamma_{jk} \leq 1, \quad i > j > k \\ &\gamma_{11} + \gamma_{22} + \gamma_{33} + \gamma_{44} - \gamma_{21} - \gamma_{32} - \gamma_{43} \leq 1. \end{aligned} \tag{3.77}$$

Extraction of the entropies can be done via Theorem 3.3.16. Here we state it for the normalized joint entropy $\underline{h}_{234}$. The complete list of entropy expressions for 4 random variables obtained via using Theorem 3.3.16 are shown in Table 3.5.

Deriving $\mathbf{h_{234}}$: Note that based on (3.17), we should obtain all 3×3 minors of the following sub-matrix,

$$\begin{bmatrix} N^{\gamma_{21}} & N^{\gamma_{22}} & 0 & 0 \\ N^{\gamma_{31}} & N^{\gamma_{32}} & N^{\gamma_{33}} & 0 \\ N^{\gamma_{41}} & N^{\gamma_{42}} & N^{\gamma_{43}} & N^{\gamma_{44}} \end{bmatrix}. \tag{3.78}$$

Recalling that $m_\alpha(\beta)$ denotes a minor whose rows and columns are indexed by α and β, respectively (see Lemma 3.3.14), then one can obtain the following either via using

Lemma 3.3.14 or direct calculation,

$$\begin{aligned} m_{234}(123) &= N^{\gamma_{21}+\gamma_{32}+\gamma_{43}} + N^{\gamma_{22}+\gamma_{33}+\gamma_{41}} - N^{\gamma_{21}+\gamma_{33}+\gamma_{42}} - N^{\gamma_{22}+\gamma_{31}+\gamma_{43}} \\ &= N^{\delta_{234/123}} + N^{\delta_{234/231}} - N^{\delta_{234/132}} - N^{\delta_{234/213}} \qquad (3.79) \\ m_{234}(124) &= N^{\gamma_{21}+\gamma_{32}+\gamma_{44}} - N^{\gamma_{22}+\gamma_{31}+\gamma_{44}} = N^{\delta_{234/124}} - N^{\delta_{234/214}} \qquad (3.80) \\ m_{234}(134) &= N^{\gamma_{21}+\gamma_{33}+\gamma_{44}} = N^{\delta_{234/134}} \qquad (3.81) \\ m_{234}(234) &= N^{\gamma_{22}+\gamma_{33}+\gamma_{44}} = N^{\delta_{234/234}}. \qquad (3.82) \end{aligned}$$

Based on (3.17), h_{234} can be written as,

$$\underline{h}_{234} = 3 - \frac{\log \gcd(m_{234}(123), m_{234}(124), m_{234}(134), m_{234}(234))}{\log N}. \qquad (3.83)$$

In order to simplify this expression, we need to examine the cases when some of the terms in $m_{234}(123)$ or $m_{234}(124)$ may cancel. When no cancelation occurs,

$$h_{234} = 3 - \min(\delta_{234/123}, \delta_{234/231}, \delta_{234/132}, \delta_{234/213}, \delta_{234/124}, \delta_{234/214}, \delta_{234/134}, \delta_{234/234}). \qquad (3.84)$$

However, some terms of $m_{234}(123)$ or $m_{234}(124)$ may cancel out with each other in which case they should not be considered in the above minimization. To address this issue consider $m_{234}(123)$. Four conditions are imaginable:

1. $\delta_{234/123} = \delta_{234/132}$
2. $\delta_{234/123} = \delta_{234/213}$
3. $\delta_{234/231} = \delta_{234/132}$
4. $\delta_{234/231} = \delta_{234/213}$.

$m_{234}(123)$ will be simplified if any of the above conditions hold. However some of these conditions may happen concurrently and therefore we need a method to compute

$m_{234}(123)$ in such cases. Considering this fact is the what lead to Lemma 3.3.15 and then to Theorem 3.3.16. Following the same steps we can rewrite $\underline{h}_{234}$ as,

$$\begin{aligned}\underline{h}_{234} \;=\; & 3 - \min\Big(\delta_{234/134}, \delta_{234/234}, \delta_{234/123} + P_{(123,231)(132,213)}(234/123), \\ & \delta_{234/231} + P_{(123,231)(132,213)}(234/231), \delta_{234/132} + P_{(123,231)(132,213)}(234/132), \\ & \delta_{234/213} + P_{(123,231)(132,213)}(234/213), \delta_{234/124} + P_{(124)(214)}(234/124), \\ & \delta_{234/214} + P_{(124)(214)}(234/214)\Big). \end{aligned} \tag{3.85}$$

Note that blowing up of a $P(\delta_{\alpha/\tilde{\beta}})$ in the above is equivalent to discarding the corresponding $\delta_{\alpha/\tilde{\beta}}$. Table 3.5 along with relations (3.77) give the whole set of equations for 4 variables. The space obtained from Table 3.5 is clearly nonconvex and so, as in the case with three random variables, we must focus on its convex hull, which from Table 3.5 seems rather hard to obtain analytically. Therefore we use Theorem 3.3.19 to calculate the innerbound $\mathfrak{R}_4$.

3.4.3.1 Calculating $\mathfrak{R}_4$

Based on Theorem 3.3.19, we obtain the chopped-hypercube defined by

$$\begin{aligned} & 0 \leq \gamma_{ij} \leq 1, \quad \forall 1 \leq j \leq i \leq 4 \\ & I(i_1, \ldots, i_k) = \gamma_{i_1 i_1} + \ldots + \gamma_{i_k i_k} - \gamma_{i_k i_{k-1}} - \ldots - \gamma_{i_2 i_1} \leq 1 \\ & \{i_1, \ldots, i_k\} \subseteq \{1, \ldots, 4\}, \quad i_1 < i_2 < \ldots < i_k. \end{aligned} \tag{3.86}$$

Using the software package PORTA [POR], we obtain 508 corner points of this 10-dimensional object in the γ_{ij} space. Calculating the corresponding entropies of these corner points based on Table 3.5 gives 67 distinct entropy vectors all of which are corner points. Some of these corner points are obtained from each other through a permutation of the underlying random variables. Excluding these permutations, 16

Table 3.2: Linear representable rays missing from the scalar lattice region of 4 random variables

	$(h_1, h_2, h_3, h_4, h_{12}, h_{13}, h_{14}, h_{23}, h_{24}, h_{34}, h_{123}, h_{124}, h_{134}, h_{234}, h_{1234})$
1	(1,1,1,1, 2,2,2,2,2,2, 2,2,2,2, 2)
2	(1,1,1,2, 2,2,2,2,2,2, 2,2,2,2, 2)
3	(1,1,2,1, 2,2,2,2,2,2, 2,2,2,2, 2)
4	(1,2,1,1, 2,2,2,2,2,2, 2,2,2,2, 2)
5	(2,1,1,1, 2,2,2,2,2,2, 2,2,2,2, 2)
6	(1,1,1,2, 2,2,3,2,3,3, 3,3,3,3, 3)
7	(1,1,2,1, 2,3,2,3,2,3, 3,3,3,3, 3)
8	(1,2,1,1, 3,2,2,3,3,2, 3,3,3,3, 3)
9	(2,1,1,1, 3,3,3,2,2,2, 3,3,3,3, 3)

entropy vectors remain, which are shown in Table 3.6 in the Appendix. Computing the convex cone of the 67 points via a linear program, gives 26 of these vectors as the rays of the region. These 26 vectors are listed in Table 3.7 in the Appendix.

3.4.3.2 Achieving Ingleton Inner Bound

In [HRSV00] it is shown that the linear representable region for 4 random variables is completely determined by Shannon and the Ingleton inequalities and that the region has 35 rays. Comparing those rays with the 26 rays of Table 3.7 we see that all 26 vectors are included in those 35 vectors. However the remaining 9 vectors (see Table 3.2) are missing in the rays of convex cone of lattice-derived entropies for 4 random variables. However we show that the vector $[1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2]$ of Table 3.2 can be achieved asymptotically with a scalar lattice construction, and the other ones in Table 3.2 can be obtained by a vector lattice.

Lemma 3.4.6 *Entropy vector* $[1,1,1,1,2,2,2,2,2,2,2,2,2,2,2]$ *for 4 random variables can be asymptotically achieved with a scalar lattice construction.*

Proof: Consider the following $[\gamma_{ij}]$ vector,

$$[\gamma_{11},\gamma_{21},\gamma_{22},\gamma_{31},\gamma_{32},\gamma_{33},\gamma_{41},\gamma_{42},\gamma_{43},\gamma_{44}] = [0,0,0,\epsilon,0,1,\epsilon^{+},0,1,1] \tag{3.87}$$

where $\epsilon > 0$ is an arbitrary small number and ϵ^{+} is any positive number greater than ϵ. It is easy to show that the above $[\gamma_{ij}]$ vector yields the lattice-derived entropy vector $[1,1,1,1,2,2,2,2,2,2-\epsilon,2,2,2,2,2]$. Taking $\epsilon \to 0$ we get the desired entropy vector. □

Next we show that the vector $[1,1,1,1,2,2,2,2,2,2,2,2,2,2,2]$ of Table 3.2 can be obtained via a vector lattice.

Lemma 3.4.7 *The vector* $[1,1,1,1,2,2,2,2,2,2,2,2,2,2,2]$ *can be obtained via a vector lattice.*

Proof: It can be easily shown that the following lattice-generating matrix yields the entropy vector $\frac{1}{2}[1,1,1,1,2,2,2,2,2,2,2,2,2,2,2]$,

$$\left[\begin{array}{ccc|cc|cc|cc} 1 & 0 & | & 0 & 0 & | & 0 & 0 & | & 0 & 0 \\ 0 & 0 & | & N & 0 & | & 0 & 0 & | & 0 & 0 \\ - & - & - & - & - & - & - & - & - & - & - \\ 0 & 1 & | & 0 & 0 & | & 0 & 0 & | & 0 & 0 \\ 0 & 0 & | & 0 & N & | & 0 & 0 & | & 0 & 0 \\ - & - & - & - & - & - & - & - & - & - & - \\ 1 & 1 & | & 0 & 0 & | & N & 0 & | & 0 & 0 \\ 0 & 0 & | & 0 & 0 & | & 0 & N & | & 0 & 0 \\ - & - & - & - & - & - & - & - & - & - & - \\ 1 & 1 & | & 0 & 0 & | & 0 & 0 & | & N & 0 \\ 0 & 0 & | & 0 & 0 & | & 0 & 0 & | & 0 & N \end{array}\right]. \tag{3.88}$$

It can be easily checked (e.g., using Mathematica) that this matrix also satisfies the quasi-uniformity condition and that N times its inverse has integer entries (see Lemma 3.3.6). The desired entropy vector therefore falls in the convex cone of the vectorized lattice-derived entropies. □

Next we show that the other vectors of Table 3.2 can also be obtained by a vector lattice.

Lemma 3.4.8 *The entropy vectors 2–9 of Table 3.2 can be obtained by a vector lattice.*

Proof: Consider the vector $[1,1,1,2,2,2,2,2,2,2,2,2,2,2,2]$. It is easy to show that this vector divided by 2, can be obtained by the vector lattice generated by the following matrix generator which satisfies the quasi-uniformity condition as well,

$$\left[\begin{array}{cc|cc|cc|cc|cc}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & N & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 1 & 0 & 0 & N & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & N & 0 & 0 & 0 \\
\hline
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & N & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
1 & 0 & 0 & 0 & 0 & N & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & N & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & N & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & N
\end{array}\right]. \tag{3.89}$$

Vectors 3, 4, and 5 of Table 3.2, scaled by a factor of 2, can be obtained by permuting the block rows of the above matrix correspondingly. Therefore again these entropy vectors of Table 3.2, fall in the convex cone of the vectorized lattice-derived entropies. Finally consider the vector $[1,1,1,2,2,2,3,2,3,3,3,3,3,3,3]$. This entropy vector divided by 2 can be obtained from the following generator which satisfies the quasi-

uniformity condition as well,

$$\left[\begin{array}{cc|cc|cc|cc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & N & 0 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & N & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & N & 0 & 0 \\ \hline 1 & 1 & 0 & 0 & 0 & 0 & N & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & N \end{array}\right]. \tag{3.90}$$

Vectors 7, 8, and 9 of Table 3.2, scaled by a factor of 2, can likewise be obtained by permuting the block rows of the above matrix, and therefore all the desired vectors will lie in the convex cone of the vectorized lattice-derived entropy vectors. □

3.4.4 Five Random Variables

We obtain the entropy expressions for 5 random variables by using Theorem 3.3.16. Since the expressions for the first 4 random variables were already reported in Table 3.5 we only give the expressions for the joint entropies which involve the 5th random variable. These can be found in Table 3.8. Together with Table 3.5 they give the whole set of normalized joint entropies for 5 lattice-derived random variables. As in the case of 4 random variables, obtaining the analytical convex hull is very difficult and we only compute the inner bound $\mathfrak{R}_5$.

3.4.4.1 Calculating $\mathfrak{R}_5$

Similar to the case of 4 random variables, we consider the following 15-dimensional polytope in the γ_{ij} space,

$$\begin{aligned} & 0 \leq \gamma_{ij} \leq 1, \quad \forall 1 \leq j \leq i \leq 5 \\ & I(i_1, \ldots, i_k) = \gamma_{i_1 i_1} + \ldots + \gamma_{i_k i_k} - \gamma_{i_k i_{k-1}} - \ldots - \gamma_{i_2 i_1} \leq 1 \\ & \{i_1, \ldots, i_k\} \subseteq \{1, \ldots, 5\}, \quad i_1 < i_2 < \ldots < i_k \end{aligned} \tag{3.91}$$

which using PORTA [POR], is found to have 10976 corner points. Computing the corresponding entropies based on Tables 3.5 and 3.8, gives 380 distinct 31-dimensional entropy vectors, all of which, are corner points of their convex hull. Their convex cone has 133 rays (clearly, all-zero vector is not counted as a ray).

Due to the particular technique of the lattice construction, the set of lattice-derived entropy vectors may not be closed under permutation of the random variables. In other words, if v is a $2^n - 1$ dimensional lattice-derived entropy vector of random variables $X_1, \ldots, X_n$, then every permutation of $X_1, \ldots, X_n$ results in a potentially different entropy vector (in fact a specific permutation of v), which may not be obtainable from the lattice construction. Therefore, in order to include these, we add all the missing permutations of the entropy vectors to the set of the rays that we found above. In this case, this addition results in 4 more entropy vectors, giving a total of 137 rays which is now closed under permutations of the underlying random variables. We find that these 137 rays are in fact, generated from the permutations (of the underlying random variables) of only 14 of them. These 14 rays are given in Table 3.9 in the Appendix.

On the other hand, the linear representable region of 5 random variables as obtained by Dougherty et al. [DFZ10], excluding the permutations, has 162 rays, of which only 18 have the property $h_{X_i} \leq 1, i = 1, \ldots, 5$. Noting that, we have consid-

ered scalar lattice-derived entropies, for which $h_{X_i} \leq 1, i = 1, \ldots, 5$ as well, we make a comparison between the lattice-derived rays and these 18 vectors, and determine that all the 14 lattice-derived rays belong to the set of these 18 vectors. The 4 missing rays are given in Table 3.10 in the Appendix.

3.5 Quasi-Uniforms of Alphabet Size 2

Among the quasi-uniform structures we have found that quasi-uniform distributions of alphabet size 2 have some nice properties which we outline here. In particular we have obtained the entropy region of these structures for 2, 3, and 4 variables. Interestingly, this region is tight for 2 and 3 variables and gives an inner bound for 4 random variables.

Theorem 3.5.1 *Let $\mathcal{Q}$ denote a quasi-uniform distribution of n random variables $X_1, \ldots, X_n$ over binary alphabet (i.e., $X_i = 0, 1 \quad \forall i$). If $\mathcal{Q}$ is such that $\mathcal{Q}(x_1 = 0, \ldots, x_n = 0) \neq 0$, then $\mathcal{Q}$ is representable over a scalar lattice.*

Proof: We prove the statement by induction. First note that the theorem is trivially true for $n = 1$ random variable. To show that this structure is constructible by a lattice, we need to show that if two points $(x_1, \ldots, x_n)$ and $(x'_1, \ldots, x'_n)$ belong to $\mathcal{Q}$, so does $(x_1 \oplus x'_1, \ldots, x_n \oplus x'_n)$ where $\oplus$ denotes addition mod 2. For $n > 1$ assume the theorem is valid for all $k < n$. If all points in $\mathcal{Q}$ are such that $x_1 = x'_1$ then the distribution can essentially be considered as an $n-1$ dimensional quasi-uniform over alphabet size 2, otherwise for $c \neq 0$ we can consider there are two points such that,

$$\mathcal{Q}(0, a_2, \ldots, a_n) = c \tag{3.92}$$

$$\mathcal{Q}(1, a'_2, \ldots, a'_n) = c. \tag{3.93}$$

Now if $\forall i > 1, a_i = a'_i$, we need to show that $\mathcal{Q}(1, a_2 \oplus a'_2, \ldots, a_n \oplus a'_n) = \mathcal{Q}(x_1 = 1, x_2 = 0, \ldots, x_n = 0) = c$. Since $\forall i > 1$ $a_i = a'_i$, that means the marginal $\mathcal{Q}(x_2 = a_2, \ldots, x_n = a_n) = 2c$ and by quasi-uniformity of the distribution all such nonzero marginals of $x_2, \ldots, x_n$ are equal to $2c$ and therefore $\mathcal{Q}(x_2 = 0, \ldots, x_n = 0) = \mathcal{Q}(x_1 = 0, x_2 = 0, \ldots, x_n = 0) + \mathcal{Q}(x_1 = 1, x_2 = 0, \ldots, x_n) = 2c$. Since $\mathcal{Q}(x_1 = 0, \ldots, x_n = 0) = c$ by assumption, we conclude that $\mathcal{Q}(x_1 = 1, x_2 = 0, \ldots, x_n = 0) = c$ which is what we needed. Therefore assume that $\exists i,\ a_i \neq a'_i$ so that the projections $(x_2 = a_2, \ldots, x_n = a_n)$ and $(x_2 = a'_2, \ldots, x_n = a'_n)$ are distinct and therefore for $c' \neq 0$ we can assume,

$$\mathcal{Q}(a_2, \ldots, a_n) = c' \tag{3.94}$$

$$\mathcal{Q}(a'_2, \ldots, a'_n) = c'. \tag{3.95}$$

Since the projection of the distribution on the $x_2, \ldots, x_n$ plane is an $n-1$ dimensional quasi-uniform structure by induction assumption, we have that $\mathcal{Q}(a_2 \oplus a'_2, \ldots, a_n \oplus a'_n) = c'$. On the other hand,

$$\mathcal{Q}(a_2 \oplus a'_2, \ldots, a_n \oplus a'_n) = \mathcal{Q}(1, a_2 \oplus a'_2, \ldots, a_n \oplus a'_n) + \mathcal{Q}(0, a_2 \oplus a'_2, \ldots, a_n \oplus a'_n) = c'. \tag{3.96}$$

Note that $c' \neq 0$ and each of the additive terms in (3.96) can be either 0 or c and therefore c' can assume values c or $2c$. If $c' = 2c$ then both term are equal to c and therefore $\mathcal{Q}(1, a_2 \oplus a'_2, \ldots, a_n \oplus a'_n) = c$. Otherwise if $c' = c$ one of the above terms is zero. Assume by contradiction that $\mathcal{Q}(1, a_2 \oplus a'_2, \ldots, a_n \oplus a'_n) = 0$ which in turn means that $\mathcal{Q}(0, a_2 \oplus a'_2, \ldots, a_n \oplus a'_n) = c$. However by (3.92) we have, $\mathcal{Q}(0, a_2, \ldots, a_n) = c$. Since the cross section $x_1 = 0$ is also an $n-1$ dimensional quasi-uniform structure, we conclude that,

$$\mathcal{Q}(0, (a_2 \oplus a'_2) \oplus a_2, \ldots, (a_n \oplus a'_n) \oplus a_n) = \mathcal{Q}(0, a'_2, \ldots, a'_n) = c. \tag{3.97}$$

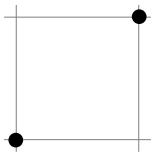


Figure 3.5: An example of a quasi-uniform distribution of two random variables with alphabet-size 2

However (3.93) and (3.97), give that, $\mathcal{Q}(a'_2, \ldots, a'_n) = 2c$ and quasi-uniformity imposes $c' = 2c$ which is a contradiction and therefore we should have $\mathcal{Q}(1, a_2 \oplus a'_2, \ldots, a_n \oplus a'_n) = c$. □

Remark: Note that the assumption $\mathcal{Q}(x_1 = 0, \ldots, x_n = 0) \neq 0$ is critical in the above theorem, since if a structure is a lattice over $GF(2)$ then for any point $(x_1, \ldots, x_n)$, the point $\mathcal{Q}(x_1 \oplus x_1, \ldots, x_n \oplus x_n) \neq 0$ should also belong to the lattice which means that the lattice should include the origin.

Next we obtain the entropy region obtained from quasi-uniform structures on alphabet size 2 for 2, 3, and 4 random variables. An example of a quasi-uniform structure with alphabet size 2 is shown in Figure 3.5 where probability is uniformly distributed among all dots.

Theorem 3.5.2 *The entropy region of 2 random variables obtained from quasi-uniform distribution of alphabet size 2 is equal to* Γ_2^*.

Proof: The set of all possible quasi-uniform distributions of 2 random variables of alphabet size 2 along with their corresponding entropy vector is shown in Figure 3.5. Obtaining the convex hull of these entropy vectors either independently or by comparison to Table 3.1 proves the theorem. □

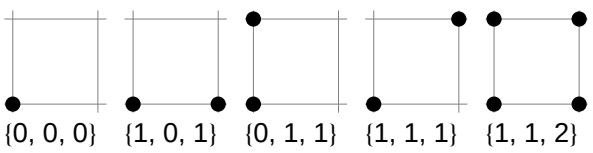


Figure 3.6: Quasi-uniform distributions of two random variables and their corresponding entropy vectors

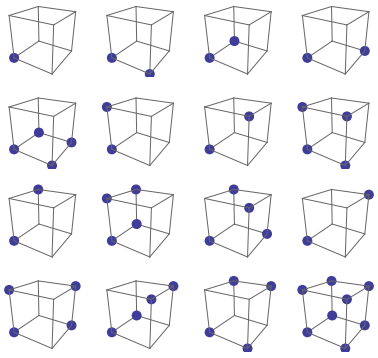


Figure 3.7: Quasi-uniform distributions of three random variables with alphabet-size 2

Theorem 3.5.3 *The entropy region of 3 random variables obtained from quasi-uniform distribution of alphabet size 2 is equal to* $\bar{\Gamma}_3^*$.

Proof: The set of all possible quasi-uniform structures on 3 random variables of alphabet size 2 is shown in Figure 3.7. Computing their corresponding entropy vectors and comparing them with the 16 entropy vectors of Table 3.4 or obtaining their convex hull independently proves the result. □

Theorem 3.5.4 *The entropy region of 4 quasi-uniform random variables over alphabet size 2 gives a strict inner bound for* Γ_4^* *and the linear representable region of 4 random variables.*

Proof: The investigation of 4 quasi-uniform random variables over alphabet size 2 gives 67 different entropy vectors corresponding to a 15-dimensional polytope with 94 facets. These 67 points match the entropy vectors obtained from the corner point of the lattice inner bound $\mathfrak{R}_4$ derived in Subsection 3.4.3, and listed in Table 3.6 in the Appendix. As we saw, those 67 corner points provide a strict inner-region for the scalar linear representable region, and hence for Γ_4^* as well. □

We should mention that there has been some nice work on identifying whether an entropy vector can be constructed from random variables of alphabet size 2 [WW09].

3.6 Conclusions

In this chapter we presented a new scheme for inner bound construction of entropy region. This method was based on determining the entropy region of random variables whose underlying joint probability distribution is uniformly spread over the points of a lattice structure in the Euclidean space. These probability distributions were assured to be quasi-uniform. We gave the principles of construction that can be generalized for any number of random variables. As any other technique the complexity increases rapidly with the number of random variables, nonetheless we obtained an explicit inner bound characterization for up to 5 random variables, and determined that the region is tight for up to 3 random variables and gives the linear representable region of 4 random variables. However we also determined that the lattice-based entropy region always satisfies the Ingleton inequality, which does not hold for all entropies in general. At the end we made comparisons with the entropy region of binary quasi-uniform random variables.

3.7 Appendix

Table 3.3: Lattice-derived regions for 3 random variables in terms of γ entries

	Conditions	h_{23}	h_{13}	h_3	h_2
1	$\gamma_{21} < \gamma_{22},\ \gamma_{31} \leq \gamma_{32} \leq \gamma_{33},$ $\gamma_{21} + \gamma_{32} < \gamma_{22} + \gamma_{31}$	$2 - \gamma_{21} - \gamma_{32}$	$2 - \gamma_{11} - \gamma_{32}$	$1 - \gamma_{31}$	$1 - \gamma_{21}$
2	$\gamma_{21} \leq \gamma_{22},\ \gamma_{31} < \gamma_{32} \leq \gamma_{33},$ $\gamma_{21} + \gamma_{32} > \gamma_{22} + \gamma_{31}$	$2 - \gamma_{22} - \gamma_{31}$			
3	$\gamma_{21} < \gamma_{22},\ \gamma_{31} < \gamma_{32} \leq \gamma_{33},$ $\gamma_{21} + \gamma_{32} = \gamma_{22} + \gamma_{31}$	$2 - \gamma_{21} - \gamma_{33}$			
4	$\gamma_{21} = \gamma_{22},\ \gamma_{31} = \gamma_{32} \leq \gamma_{33},$ $\gamma_{21} + \gamma_{32} = \gamma_{22} + \gamma_{31}$				
5	$\gamma_{21} < \gamma_{22},\ \gamma_{31} \leq \gamma_{33} \leq \gamma_{32},$ $\gamma_{21} + \gamma_{32} = \gamma_{22} + \gamma_{31}$	$2 - \gamma_{21} - \gamma_{33}$	$2 - \gamma_{11} - \gamma_{33}$	$1 - \gamma_{31}$	$1 - \gamma_{21}$
6	$\gamma_{21} < \gamma_{22},\ \gamma_{31} < \gamma_{33} < \gamma_{32}$ $\gamma_{21} + \gamma_{32} > \gamma_{22} + \gamma_{31}$ $\gamma_{22} + \gamma_{31} \geq \gamma_{33} + \gamma_{21}$				
7	$\gamma_{21} < \gamma_{22},\ \gamma_{31} < \gamma_{33} \leq \gamma_{32},$ $\gamma_{21} + \gamma_{32} = \gamma_{22} + \gamma_{31}$				
8	$\gamma_{21} < \gamma_{22},\ \gamma_{31} = \gamma_{33} < \gamma_{32},$ $\gamma_{21} + \gamma_{32} \geq \gamma_{22} + \gamma_{31}$				
9	$\gamma_{21} < \gamma_{22},\ \gamma_{31} < \gamma_{33} < \gamma_{32},$ $\gamma_{21} + \gamma_{32} > \gamma_{22} + \gamma_{31},$ $\gamma_{22} + \gamma_{31} \leq \gamma_{33} + \gamma_{21}$	$2 - \gamma_{22} - \gamma_{31}$			
10	$\gamma_{21} < \gamma_{22},\ \gamma_{31} < \gamma_{33} = \gamma_{32},$ $\gamma_{21} + \gamma_{32} > \gamma_{22} + \gamma_{31}$				
11	$\gamma_{21} = \gamma_{22},\ \gamma_{31} \leq \gamma_{33} \leq \gamma_{32}$	$2 - \gamma_{21} - \gamma_{31}$			

Table 3.3: (Continued)

	Conditions	h_{23}	h_{13}	h_3	h_2
12	$\gamma_{21} < \gamma_{22},\ \gamma_{32} \le \gamma_{31} \le \gamma_{33}$	$2-\gamma_{21}-\gamma_{32}$	$2-\gamma_{11}-\gamma_{32}$	$1-\gamma_{32}$	$1-\gamma_{21}$
13	$\gamma_{21} = \gamma_{22},\ \gamma_{32} < \gamma_{31} \le \gamma_{33}$				
14	$\gamma_{21} \le \gamma_{22},\ \gamma_{32} \le \gamma_{33} \le \gamma_{31}$				
15	$\gamma_{21} = \gamma_{22},\ \gamma_{32} = \gamma_{31} \le \gamma_{33}$	$2-\gamma_{21}-\gamma_{33}$			
16	$\gamma_{21} \le \gamma_{22},\ \gamma_{33} \le \gamma_{31} \le \gamma_{32}$	$2-\gamma_{21}-\gamma_{33}$	$2-\gamma_{11}-\gamma_{33}$	$1-\gamma_{33}$	$1-\gamma_{21}$
17	$\gamma_{21} \le \gamma_{22},\ \gamma_{33} \le \gamma_{32} \le \gamma_{31}$				
18	$\gamma_{22} < \gamma_{21},\ \gamma_{31} \le \gamma_{32} \le \gamma_{33}$	$2-\gamma_{22}-\gamma_{31}$	$2-\gamma_{11}-\gamma_{32}$	$1-\gamma_{31}$	$1-\gamma_{22}$
19	$\gamma_{22} < \gamma_{21},\ \gamma_{31} \le \gamma_{33} \le \gamma_{32}$	$2-\gamma_{22}-\gamma_{31}$	$2-\gamma_{11}-\gamma_{33}$	$1-\gamma_{31}$	$1-\gamma_{22}$
20	$\gamma_{22} < \gamma_{21},\ \gamma_{32} < \gamma_{31} \le \gamma_{33},$ $\gamma_{21}+\gamma_{32} < \gamma_{22}+\gamma_{31}$	$2-\gamma_{21}-\gamma_{32}$	$2-\gamma_{11}-\gamma_{32}$	$1-\gamma_{32}$	$1-\gamma_{22}$
21	$\gamma_{22} < \gamma_{21},\ \gamma_{32} < \gamma_{33} = \gamma_{31},$ $\gamma_{21}+\gamma_{32} < \gamma_{22}+\gamma_{31}$				
22	$\gamma_{22} < \gamma_{21},\ \gamma_{32} < \gamma_{33} < \gamma_{31},$ $\gamma_{21}+\gamma_{32} < \gamma_{22}+\gamma_{31}$ $\gamma_{21}+\gamma_{32} < \gamma_{22}+\gamma_{33}$				
23	$\gamma_{22} < \gamma_{21},\ \gamma_{32} \le \gamma_{31} \le \gamma_{33},$ $\gamma_{21}+\gamma_{32} > \gamma_{22}+\gamma_{31}$	$2-\gamma_{22}-\gamma_{31}$			
24	$\gamma_{22} < \gamma_{21},\ \gamma_{32} < \gamma_{31} \le \gamma_{33},$ $\gamma_{21}+\gamma_{32} = \gamma_{22}+\gamma_{31}$	$2-\gamma_{22}-\gamma_{33}$			
25	$\gamma_{22} < \gamma_{21},\ \gamma_{32} < \gamma_{33} < \gamma_{31}$ $\gamma_{22}+\gamma_{33} \le \gamma_{21}+\gamma_{32}$ $\gamma_{21}+\gamma_{32} < \gamma_{22}+\gamma_{31}$				
26	$\gamma_{22} < \gamma_{21},\ \gamma_{32} < \gamma_{33} < \gamma_{31},$ $\gamma_{21}+\gamma_{32} \ge \gamma_{22}+\gamma_{31}$				
27	$\gamma_{22} < \gamma_{21},\ \gamma_{32} < \gamma_{33} = \gamma_{31},$				

Table 3.3: (Continued)

	Conditions	h_{23}	h_{13}	h_3	h_2
	$\gamma_{21} + \gamma_{32} \geq \gamma_{22} + \gamma_{31}$				
28	$\gamma_{22} < \gamma_{21},\ \gamma_{32} = \gamma_{33} \leq \gamma_{31}$				
29	$\gamma_{22} < \gamma_{21},\ \gamma_{33} \leq \gamma_{31} \leq \gamma_{32}$	$2 - \gamma_{22} - \gamma_{33}$	$2 - \gamma_{11} - \gamma_{33}$	$1 - \gamma_{33}$	$1 - \gamma_{22}$
30	$\gamma_{22} < \gamma_{21},\ \gamma_{33} \leq \gamma_{32} \leq \gamma_{31}$				

Table 3.4: Entropy corner points for 3 random variables obtained through lattice construction (corner points of $\mathfrak{R}_3$, and the corresponding γ)

	$(\gamma_{11}, \gamma_{21}, \gamma_{22}, \gamma_{31}, \gamma_{32}, \gamma_{33})$	$(h_1, h_2, h_3, h_{12}, h_{13}, h_{23}, h_{123})$
1	(1, 1, 1, 1, 1, 1)	(0, 0, 0, 0, 0, 0, 0)
2	(1, 1, 1, 0, 0, 0) (1, 1, 1, 0, 1, 0) (1, 1, 1, 1, 0, 0) (1, 1, 1, 1, 1, 0)	(0, 0, 1, 0, 1, 1, 1)
3	(1, 0, 0, 1, 1, 1) (1, 1, 0, 1, 1, 1)	(0, 1, 0, 1, 0, 1, 1)
4	(1, 1, 0, 1, 0, 1)	(0, 1, 1, 1, 1, 1, 1)
5	(1, 0, 0, 0, 0, 0) (1, 0, 0, 0, 1, 0) (1, 0, 0, 1, 0, 0) (1, 1, 0, 0, 0, 0) (1, 0, 0, 1, 1, 0) (1, 1, 0, 0, 1, 0) (1, 1, 0, 1, 0, 0) (1, 1, 0, 1, 1, 0)	(0, 1, 1, 1, 1, 2, 2)
6	(0, 1, 1, 1, 1, 1)	(1, 0, 0, 1, 1, 0, 1)
7	(0, 1, 1, 0, 1, 1)	(1, 0, 1, 1, 1, 1, 1)
8	(0, 1, 1, 0, 0, 0) (0, 1, 1, 0, 1, 0) (0, 1, 1, 1, 0, 0) (0, 1, 1, 1, 1, 0)	(1, 0, 1, 1, 2, 1, 2)
9	(0, 0, 1, 1, 1, 1)	(1, 1, 0, 1, 1, 1, 1)
10	(0, 0, 0, 1, 1, 1)	(1, 1, 0, 2, 1, 1, 2)

Table 3.4: (Continued)

	$(\gamma_{11}, \gamma_{21}, \gamma_{22}, \gamma_{31}, \gamma_{32}, \gamma_{33})$	$(h_1, h_2, h_3, h_{12}, h_{13}, h_{23}, h_{123})$
	(0, 1, 0, 1, 1, 1)	
11	(0, 0, 1, 0, 1, 1)	(1, 1, 1, 1, 1, 1, 1)
12	(0, 0, 1, 0, 0, 0) (0, 0, 1, 0, 1, 0) (0, 0, 1, 1, 0, 0) (0, 0, 1, 1, 1, 0)	(1, 1, 1, 1, 2, 2, 2)
13	(0, 0, 0, 0, 1, 1) (0, 1, 0, 0, 1, 1)	(1, 1, 1, 2, 1, 2, 2)
14	(0, 0, 0, 0, 0, 1) (0, 1, 0, 1, 0, 1)	(1, 1, 1, 2, 2, 1, 2)
15	(0, 0, 0, 1, 0, 1) (0, 1, 0, 0, 0, 1)	(1, 1, 1, 2, 2, 2, 2)
16	(0, 0, 0, 0, 0, 0) (0, 0, 0, 0, 1, 0) (0, 0, 0, 1, 0, 0) (0, 1, 0, 0, 0, 0) (0, 0, 0, 1, 1, 0) (0, 1, 0, 0, 1, 0) (0, 1, 0, 1, 0, 0) (0, 1, 0, 1, 1, 0)	(1, 1, 1, 2, 2, 2, 3)

Table 3.5: Entropy descriptions of 4 scalar lattice-derived random variables

Entropy	Description
h_1	$1 - \gamma_{11}$
h_2	$1 - \min(\gamma_{21}, \gamma_{22})$
h_3	$1 - \min(\gamma_{31}, \gamma_{32}, \gamma_{33})$
h_4	$1 - \min(\gamma_{41}, \gamma_{42}, \gamma_{43}, \gamma_{44})$
h_{12}	$2 - \gamma_{11} - \gamma_{22}$
h_{13}	$2 - \gamma_{11} - \min(\gamma_{32}, \gamma_{33})$
h_{14}	$2 - \gamma_{11} - \min(\gamma_{42}, \gamma_{43}, \gamma_{44})$
h_{23}	$2 - \min(\delta_{23/12} + P_{(12)(21)}(23/12), \delta_{23/21} + P_{(12)(21)}(23/21), \delta_{23/13}, \delta_{23/23})$
h_{24}	$2 - \min(\delta_{24/13}, \delta_{24/23}, \delta_{24/14}, \delta_{24/24}, \delta_{24/12} + P_{(12)(21)}(24/12),$ $\delta_{24/21} + P_{(12)(21)}(24/21))$
h_{34}	$2 - \min(\delta_{34/12} + P_{(12)(21)}(34/12), \delta_{34/21} + P_{(12)(21)}(34/21),$ $\delta_{34/13} + P_{(13)(31)}(34/13), \delta_{34/31} + P_{(13)(31)}(34/31),$ $\delta_{34/23} + P_{(23)(32)}(34/23), \delta_{34/32} + P_{(23)(32)}(34/32), \delta_{34/14}, \delta_{34/24}, \delta_{34/34})$
h_{123}	$3 - \gamma_{11} - \gamma_{22} - \gamma_{33}$
h_{124}	$3 - \gamma_{11} - \gamma_{22} - \min(\gamma_{43}, \gamma_{44})$
h_{134}	$3 - \gamma_{11} - \min(\delta_{34/23} + P_{(23)(32)}(34/23), \delta_{34/32} + P_{(23)(32)}(34/32),$ $\delta_{34/24}, \delta_{34/34})$
h_{234}	$3 - \min(\delta_{234/123} + P_{(123,231)(132,213)}(234/123), \delta_{234/134}, \delta_{234/234},$ $\delta_{234/231} + P_{(123,231)(132,213)}(234/231), \delta_{234/132} + P_{(123,231)(132,213)}(234/132),$ $\delta_{234/213} + P_{(123,231)(132,213)}(234/213), \delta_{234/124} + P_{(124)(214)}(234/124),$ $\delta_{234/214} + P_{(124)(214)}(234/214))$
h_{1234}	$4 - \gamma_{11} - \gamma_{22} - \gamma_{33} - \gamma_{44}$

Table 3.6: Corner points of the scalar lattice-derived region for 4 random variables, excluding the permutations

	$(h_1\ h_2\ h_3\ h_4, h_{12}\ h_{13}\ h_{14}\ h_{23}\ h_{24}\ h_{34}, h_{123}\ h_{124}\ h_{134}\ h_{234}, h_{1234})$
1	(0 0 0 0, 0 0 0 0 0 0, 0 0 0 0, 0)
2	(0 0 0 1, 0 0 1 0 1 1, 0 1 1 1, 1)
3	(0 0 1 1, 0 1 1 1 1 1, 1 1 1 1, 1)
4	(0 0 1 1, 0 1 1 1 1 2, 1 1 2 2, 2)
5	(0 1 1 1, 1 1 1 1 1 1, 1 1 1 1, 1)
6	(0 1 1 1, 1 1 1 1 2 2, 1 2 2 2, 2)
7	(0 1 1 1, 1 1 1 2 2 2, 2 2 2 2, 2)
8	(0 1 1 1, 1 1 1 2 2 2, 2 2 2 3, 3)
9	(1 1 1 1, 1 1 1 1 1 1, 1 1 1 1, 1)
10	(1 1 1 1, 1 1 2 1 2 2, 1 2 2 2, 2)
11	(1 1 1 1, 1 2 2 2 2 1, 2 2 2 2, 2)
12	(1 1 1 1, 1 2 2 2 2 2, 2 2 2 2, 2)
13	(1 1 1 1, 1 2 2 2 2 2, 2 2 3 3, 3)
14	(1 1 1 1, 2 2 2 2 2 2, 2 3 3 3, 3)
15	(1 1 1 1, 2 2 2 2 2 2, 3 3 3 3, 3)
16	(1 1 1 1, 2 2 2 2 2 2, 3 3 3 3, 4)

Including permutations of these 16 vectors (obtained from permutations of the underlying random variables), yields a total of 67 vectors for the corner points.

Table 3.7: Rays of the scalar lattice-derived region for 4 random variables

	$(h_1\ h_2\ h_3\ h_4, h_{12}\ h_{13}\ h_{14}\ h_{23}\ h_{24}\ h_{34}, h_{123}\ h_{124}\ h_{134}\ h_{234}, h_{1234})$
1	(0 0 0 1, 0 0 1 0 1 1, 0 1 1 1, 1)
2	(0 0 1 0, 0 1 0 1 0 1, 1 0 1 1, 1)
3	(0 1 0 0, 1 0 0 1 1 0, 1 1 0 1, 1)
4	(1 0 0 0, 1 1 1 0 0 0, 1 1 1 0, 1)
5	(0 0 1 1, 0 1 1 1 1 1, 1 1 1 1, 1)
6	(0 1 0 1, 1 0 1 1 1 1, 1 1 1 1, 1)
7	(1 0 0 1, 1 1 1 0 1 1, 1 1 1 1, 1)
8	(0 1 1 0, 1 1 0 1 1 1, 1 1 1 1, 1)
9	(1 0 1 0, 1 1 1 1 0 1, 1 1 1 1, 1)
10	(1 1 0 0, 1 1 1 1 1 0, 1 1 1 1, 1)
11	(0 1 1 1, 1 1 1 1 1 1, 1 1 1 1, 1)
12	(1 0 1 1, 1 1 1 1 1 1, 1 1 1 1, 1)
13	(1 1 0 1, 1 1 1 1 1 1, 1 1 1 1, 1)
14	(1 1 1 0, 1 1 1 1 1 1, 1 1 1 1, 1)
15	(0 1 1 1, 1 1 1 2 2 2, 2 2 2 2, 2)
16	(1 0 1 1, 1 2 2 1 1 2, 2 2 2 2, 2)
17	(1 1 0 1, 2 1 2 1 2 1, 2 2 2 2, 2)
18	(1 1 1 0, 2 2 1 2 1 1, 2 2 2 2, 2)
19	(1 1 1 1, 1 1 1 1 1 1, 1 1 1 1, 1)
20	(1 1 1 1, 1 2 2 2 2 2, 2 2 2 2, 2)
21	(1 1 1 1, 2 1 2 2 2 2, 2 2 2 2, 2)
22	(1 1 1 1, 2 2 1 2 2 2, 2 2 2 2, 2)
23	(1 1 1 1, 2 2 2 1 2 2, 2 2 2 2, 2)
24	(1 1 1 1, 2 2 2 2 1 2, 2 2 2 2, 2)

Table 3.7: (Continued)

	$(h_1\ h_2\ h_3\ h_4, h_{12}\ h_{13}\ h_{14}\ h_{23}\ h_{24}\ h_{34}, h_{123}\ h_{124}\ h_{134}\ h_{234}, h_{1234})$
25	(1 1 1 1, 2 2 2 2 2 1, 2 2 2 2, 2)
26	(1 1 1 1, 2 2 2 2 2 2, 3 3 3 3, 3)

Table 3.8: Joint entropy descriptions of 5 scalar lattice-derived random variables involving the 5th variable

Entropy	Description
h_5	$1 - \min(\gamma_{51}, \gamma_{52}, \gamma_{53}, \gamma_{54}, \gamma_{55})$
h_{15}	$2 - \gamma_{11} - \min(\gamma_{52}, \gamma_{53}, \gamma_{54}, \gamma_{55})$
h_{25}	$2 - \min(\delta_{25/12} + P_{(12)(21)}(25/12), \delta_{25/21} + P_{(12)(21)}(25/21), \delta_{25/13}, \delta_{25/14}, \delta_{25/15}, \delta_{25/23}, \delta_{25/24}, \delta_{25/25})$
h_{35}	$2 - \min(\delta_{35/12} + P_{(12)(21)}(35/12), \delta_{35/21} + P_{(12)(21)}(35/21), \delta_{35/13} + P_{(13)(31)}(35/13), \delta_{35/31} + P_{(13)(31)}(35/31), \delta_{35/23} + P_{(23)(32)}(35/23), \delta_{35/32} + P_{(23)(32)}(35/32), \delta_{35/14}, \delta_{35/15}, \delta_{35/24}, \delta_{35/25}, \delta_{35/34}, \delta_{35/35})$
h_{45}	$2 - \min(\delta_{45/12} + P_{(12)(21)}(45/12), \delta_{45/21} + P_{(12)(21)}(45/21), \delta_{45/13} + P_{(13)(31)}(45/13), \delta_{45/31} + P_{(13)(31)}(45/31), \delta_{45/14} + P_{(14)(41)}(45/14), \delta_{45/41} + P_{(14)(41)}(45/41), \delta_{45/23} + P_{(23)(32)}(45/23), \delta_{45/32} + P_{(23)(32)}(45/32), \delta_{45/24} + P_{(24)(42)}(45/24), \delta_{45/42} + P_{(24)(42)}(45/42), \delta_{45/34} + P_{(34)(43)}(45/34), \delta_{45/43} + P_{(34)(43)}(45/43), \delta_{45/15}, \delta_{45/25}, \delta_{45/35}, \delta_{45/45})$
h_{125}	$3 - \gamma_{11} - \gamma_{22} - \min(\gamma_{53}, \gamma_{54}, \gamma_{55})$
h_{135}	$3 - \gamma_{11} - \min(\delta_{35/23} + P_{(23)(32)}(35/23), \delta_{35/32} + P_{(23)(32)}(35/32), \delta_{35/24}, \delta_{35/25}, \delta_{35/34}, \delta_{35/35})$
h_{145}	$3 - \gamma_{11} - \min(\delta_{45/23} + P_{(23)(32)}(45/23), \delta_{45/32} + P_{(23)(32)}(45/32), \delta_{45/24} + P_{(24)(42)}(45/24), \delta_{45/42} + P_{(24)(42)}(45/42), \delta_{45/34} + P_{(34)(43)}(45/34), \delta_{45/43} + P_{(34)(43)}(45/43), \delta_{45/25}, \delta_{45/35}, \delta_{45/45})$
h_{235}	$3 - \min(\delta_{235/123} + P_{(123,231)(132,213)}(235/123), \delta_{235/231} + P_{(123,231)(132,213)}(235/231), \delta_{235/132} + P_{(123,231)(132,213)}(235/132), \delta_{235/213} + P_{(123,231)(132,213)}(235/213), \delta_{235/124} + P_{(12)(21)}(23/12), \delta_{235/214} + P_{(12)(21)}(23/21), \delta_{235/125} + P_{(12)(21)}(23/12), \delta_{235/215} + P_{(12)(21)}(23/21), \delta_{235/134}, \delta_{235/135}, \delta_{235/234}, \delta_{235/235})$

Table 3.8: (Continued)

Entropy	Description
h_{245}	$3 - \min(\delta_{245/123} + P_{(123,231)(132,213)}(245/123),$ $\delta_{245/231} + P_{(123,231)(132,213)}(245/231), \delta_{245/132} + P_{(123,231)(132,213)}(245/132),$ $\delta_{245/213} + P_{(123,231)(132,213)}(245/213), \delta_{245/124} + P_{(124,241)(142,214)}(245/124),$ $\delta_{245/241} + P_{(124,241)(142,214)}(245/241), \delta_{245/142} + P_{(124,241)(142,214)}(245/142),$ $\delta_{245/214} + P_{(124,241)(142,214)}(245/214), \delta_{245/125} + P_{(12)(21)}(24/12),$ $\delta_{245/215} + P_{(12)(21)}(245/215), \delta_{245/134} + P_{(34)(43)}(45/34),$ $\delta_{245/143} + P_{(34)(43)}(45/43), \delta_{245/234} + P_{(34)(43)}(45/34),$ $\delta_{245/243} + P_{(34)(43)}(45/43), \delta_{245/135}, \delta_{245/145}, \delta_{245/235}, \delta_{245/245})$
h_{345}	$3 - \min(\delta_{345/123} + P_{(123,231,312)(132,213,321)}(345/123),$ $\delta_{345/231} + P_{(123,231,312)(132,213,321)}(345/231),$ $\delta_{345/312} + P_{(123,231,312)(132,213,321)}(345/312),$ $\delta_{345/132} + P_{(123,231,312)(132,213,321)}(345/132),$ $\delta_{345/213} + P_{(123,231,312)(132,213,321)}(345/213),$ $\delta_{345/321} + P_{(123,231,312)(132,213,321)}(345/321),$ $\delta_{345/124} + P_{(124,241)(214,142)}(345/124), \delta_{345/241} + P_{(124,241)(214,142)}(345/241),$ $\delta_{345/214} + P_{(124,241)(214,142)}(345/214), \delta_{345/142} + P_{(124,241)(214,142)}(345/142),$ $\delta_{345/125} + P_{(12)(21)}(34/12), \delta_{345/215} + P_{(12)(21)}(34/21),$ $\delta_{345/134} + P_{(134,341)(314,143)}(345/134), \delta_{345/341} + P_{(134,341)(314,143)}(345/341),$ $\delta_{345/314} + P_{(134,341)(314,143)}(345/314), \delta_{345/143} + P_{(134,341)(314,143)}(345/143),$ $\delta_{345/135} + P_{(13)(31)}(34/13), \delta_{345/315} + P_{(13)(31)}(34/31),$ $\delta_{345/234} + P_{(234,342)(324,243)}(345/234), \delta_{345/342} + P_{(234,342)(324,243)}(345/342),$ $\delta_{345/324} + P_{(234,342)(324,243)}(345/324), \delta_{345/243} + P_{(234,342)(324,243)}(345/243),$ $\delta_{345/235} + P_{(23)(32)}(34/23), \delta_{345/325} + P_{(23)(32)}(34/32),$ $\delta_{345/145}, \delta_{345/245}, \delta_{345/345})$

Table 3.8: (Continued)

Entropy	Description
h_{1235}	$4 - \min(\delta_{1235/1234}, \delta_{1235/1235})$
h_{1245}	$4 - \gamma_{11} - \gamma_{22} - \min(\delta_{45/34} + P_{(34)(43)}(45/34), \delta_{45/43} + P_{(34)(43)}(45/43),$ $\delta_{45/35}, \delta_{45/45})$
h_{1345}	$4 - \min(\delta_{1345/1234} + P_{(1234,1342)(1243,1324)}(1345/1234), \delta_{1345/1342}+$ $+P_{(1234,1342)(1243,1324)}(1345/1342), \delta_{1345/1243}+$ $+P_{(1234,1342)(1243,1324)}(1345/1243), \delta_{1345/1324}+$ $+P_{(1234,1342)(1243,1324)}(1345/1324), \delta_{1345/1235} + P_{(23)(32)}(34/23),$ $\delta_{1345/1325} + P_{(23)(32)}(34/32), \delta_{1345/1245}, \delta_{1345/1345})$
h_{2345}	$4 - \min(\delta_{2345/1234} + P_{(1234,1342,2143,2314)(1243,1324,2134,2341)}(2345/1234),$ $\delta_{2345/1342} + P_{(1234,1342,2143,2314)(1243,1324,2134,2341)}(2345/1342),$ $\delta_{2345/2143} + P_{(1234,1342,2143,2314)(1243,1324,2134,2341)}(2345/2143),$ $\delta_{2345/2314} + P_{(1234,1342,2143,2314)(1243,1324,2134,2341)}(2345/2314),$ $\delta_{2345/1243} + P_{(1234,1342,2143,2314)(1243,1324,2134,2341)}(2345/1243),$ $\delta_{2345/1324} + P_{(1234,1342,2143,2314)(1243,1324,2134,2341)}(2345/1324),$ $\delta_{2345/2134} + P_{(1234,1342,2143,2314)(1243,1324,2134,2341)}(2345/2134),$ $\delta_{2345/2341} + P_{(1234,1342,2143,2314)(1243,1324,2134,2341)}(2345/2341),$ $\delta_{2345/1235} + P_{(1235,2315)(1325,2135)}(2345/1235),$ $\delta_{2345/2315} + P_{(1235,2315)(1325,2135)}(2345/2315),$ $\delta_{2345/1325} + P_{(1235,2315)(1325,2135)}(2345/1325),$ $\delta_{2345/2135} + P_{(1235,2315)(1325,2135)}(2345/2135),$ $\delta_{2345/1245} + P_{(12)(21)}(23/12), \delta_{2345/2145} + P_{(12)(21)}(23/21),$ $\delta_{2345/1345}, \delta_{2345/2345})$
h_{12345}	$5 - \gamma_{11} - \gamma_{22} - \gamma_{33} - \gamma_{44} - \gamma_{55}$

Table 3.9: Rays of scalar lattice-derived region of 5 random variables ($\mathfrak{R}_5$), excluding the permutations

	h
1	(0 0 0 0 1, 0 0 0 1 0 0 1 0 1 1, 0 0 1 0 1 1 0 1 1 1, 0 1 1 1 1, 1)
2	(0 0 0 1 1, 0 0 1 1 0 1 1 1 1 1, 0 1 1 1 1 1 1 1 1 1, 1 1 1 1 1, 1)
3	(0 0 1 1 1, 0 1 1 1 1 1 1 1 1 1, 1 1 1 1 1 1 1 1 1 1, 1 1 1 1 1, 1)
4	(0 0 1 1 1, 0 1 1 1 1 1 2 2 2, 1 1 1 2 2 2 2 2 2 2, 2 2 2 2 2, 2)
5	(0 1 1 1 1, 1 1 1 1 1 1 1 1 1 1, 1 1 1 1 1 1 1 1 1 1, 1 1 1 1 1, 1)
6	(0 1 1 1 1, 1 1 1 1 2 2 2 2 2 1, 2 2 2 2 2 1 2 2 2 2, 2 2 2 2 2, 2)
7	(0 1 1 1 1, 1 1 1 1 2 2 2 2 2 2, 2 2 2 2 2 2 3 3 3 3, 3 3 3 3 3, 3)
8	(1 1 1 1 1, 1 1 1 1 1 1 1 1 1 1, 1 1 1 1 1 1 1 1 1 1, 1 1 1 1 1, 1)
9	(1 1 1 1 1, 1 2 2 2 2 2 2 2 2 2, 2 2 2 3 3 3 3 3 3 3, 3 3 3 3 3, 3)
10	(1 1 1 1 1, 2 2 2 2 2 2 1 1 2 2, 2 2 2 2 2 2 2 2 2 2, 2 2 2 2 2, 2)
11	(1 1 1 1 1, 2 2 2 2 2 2 2 1 1 1, 2 2 2 2 2 2 2 2 2 1, 2 2 2 2 2, 2)
12	(1 1 1 1 1, 2 2 2 2 2 2 2 2 2 2, 3 3 3 3 3 2 3 2 3 3, 3 3 3 3 3, 3)
13	(1 1 1 1 1, 2 2 2 2 2 2 2 2 2 2, 3 3 3 3 3 3 3 3 3 2, 3 3 3 3 3, 3)
14	(1 1 1 1 1, 2 2 2 2 2 2 2 2 2 2, 3 3 3 3 3 3 3 3 3 3, 4 4 4 4 4, 4)

The entries of the entropy vector h are ordered as follows:

$$\begin{aligned} h \;=\; & (h_1\ h_2\ h_3\ h_4\ h_5, h_{12}\ h_{13}\ h_{14}\ h_{15}\ h_{23}\ h_{24}\ h_{25}\ h_{34}\ h_{35}\ h_{45}, h_{123}\ h_{124}\ h_{125} \\ & h_{134}\ h_{135}\ h_{145}\ h_{234}\ h_{235}\ h_{245}\ h_{345}, h_{1234}\ h_{1235}\ h_{1245}\ h_{1345}\ h_{2345}, h_{12345}). \end{aligned}$$

Table 3.10: Rays of the linear representable region of 5 random variables (having $h_{X_i} \leq 1$), missing from the lattice-derived region

	h
1	(0 1 1 1 1, 1 1 1 1 2 2 2 2 2 2, 2 2 2 2 2 2 2 2 2 2, 2 2 2 2 2, 2)
2	(1 1 1 1 1, 2 2 2 2 2 2 2 2 2 1, 2 2 2 2 2 2 2 2 2 2, 2 2 2 2 2, 2)
3	(1 1 1 1 1, 2 2 2 2 2 2 2 2 2 2, 2 2 2 2 2 2 2 2 2 2, 2 2 2 2 2, 2)
4	(1 1 1 1 1, 2 2 2 2 2 2 2 2 2 2, 3 3 3 3 3 3 3 3 3 3, 3 3 3 3 3, 3)

The entries of the entropy vector h are ordered as follows:

$$\begin{aligned} h \;=\; & (h_1\ h_2\ h_3\ h_4\ h_5, h_{12}\ h_{13}\ h_{14}\ h_{15}\ h_{23}\ h_{24}\ h_{25}\ h_{34}\ h_{35}\ h_{45}, h_{123}\ h_{124}\ h_{125} \\ & h_{134}\ h_{135}\ h_{145}\ h_{234}\ h_{235}\ h_{245}\ h_{345}, h_{1234}\ h_{1235}\ h_{1245}\ h_{1345}\ h_{2345}, h_{12345}). \end{aligned}$$

Chapter 4

Linear Representable Entropy Vectors

4.1 Introduction

There recently has been a great deal of effort to determine the information-theoretic capacity of networks. However, for a long time, the basis of the main technique for sending information over networks was to consider information as a fluid which could only be routed (or replicated). Network coding, first introduced in [ACLY00], showed that for networks with two destinations or more, routing is not optimal and coding at the nodes of the network can in general increase the throughput and save bandwidth. Nonetheless the optimal coding strategy remains as a topic of research.

In the multicast scenario, where all the destinations desire the same set of source messages, linear network coding is proven to achieve the cut-set bound. For the general multi-source multi-sink networks where sinks can have arbitrary demands, the capacity region is expressed in terms of the space of *entropy vectors* Γ_n^* which for the case of networks is the entropy region of random variables associated with the network [YYZ07][HS07a]. This characterization yields the best rates possible, independent of the coding used to achieve them (see Chapter 2). Since the characterization of Γ_n^* is an open problem for more than 3 variables, explicit computation of the capacity region remains unsolved. However one might be interested in obtaining the capacity

region of the network when using a specific group of codes.

Linear codes in particular are very interesting. In fact, although they are proven to be suboptimal in general [DFZ05], they are appealing due to their simple structure. However determining the linear coding capacity of an arbitrary network also remains an open problem. As a matter of fact just as the coding capacity of networks is deeply connected to the characterization of the entropy region, the linear coding capacity can be shown to be related the characterization of a subset of the entropy region known as the "linear representable region". Lack of a full characterization for the linear representable region indeed accounts for the absence of a complete solution to the determination of the linear coding capacity of arbitrary networks.

In this chapter, we will focus on characterizing the scalar linear representable region (a subset of the linear representable region) and give a complete solution for the case of 4 random variables. Moreover we study linear network coding under different assumptions, such as restricting the number of sources to 2 or linear network coding over binary alphabet size.

The rest of this chapter is organized as follows. First we will state some known results about how the scalar linear coding capacity is related to the region of scalar linear representable entropy vectors. Next we give an algorithm and explicitly compute the scalar linear representable entropy region for 4 random variables. The method is in principle extendable to greater number of random variables as well. We then turn our attention to networks with 2 sources and show the optimality of linear codes among all scalar codes for the network. Finally we study binary capacity of networks by appealing to the linear representability results for small finite fields in the matroid theory.

4.2 Preliminaries

In this section we review some definitions and important theorems about linear codes and linear representability of vectors.

Definition 4.2.1 (Linear Code) *Let $G = (V, E)$ denote a network graph with the node set V and the edge set E. A subset of V at which source messages are generated is called the source nodes and those nodes of V which demand some of the sources are called sinks. Moreover it is assumed that each edge of the network carries a message. Denote the random variable associated with the j-th source generated at node i by ${X_S}^i(j)$ and the random variable on the edge e of the network by X_e. Furthermore assume that the source and edge variables are vector valued random variables of size m_s and m_v over some finite field $\mathcal{F}$. Let $X^i_{in}(j)$ and $X^i_{out}(j)$ be the jth input and output variables for node i, respectively. If there is a set of matrices F^v_{jk} and F^s_{jk} over $\mathcal{F}$ such that,*

1. *every output of node i is obtained by a linear transformation from its inputs and possible source messages generated at i, i.e., $X^i_{out}(k) = \sum_j F^v_{kj} {X_{in}}^i(j) + \sum_j F^s_{kj} X^i_s(j)$*

2. *and moreover every sink node l whose j-th demand is denoted by $X^l_d(j)$ can reconstruct its demands from a linear combination of its inputs, i.e., there exists a set of matrices G^v_{kj} and G^s_{kj} over $\mathcal{F}$ such that, $X^l_d(k) = \sum_j G^v_{kj} {X_{in}}^l(j) + \sum_j G^s_{kj} X^l_s(j)$*

then we say that the set of $F^v_{jk}, F^s_{jk}, G^v_{kj}$, and G^s_{kj} constitute a linear code for the network.

Remark: In Definition 4.2.1 when $m_s = m_v$ the network is called solvable. Furthermore if $m_s = m_v = 1$ then the linear code is called scalar linear and the network is therefore scalar linear solvable [DFZ05].

It is shown that every solvable multicast network has a scalar linear solution over a sufficiently large alphabet size [LYC03, KM03]. Although later it was proved that for the case of non-multicast networks even vector linear coding is sub-optimal [DFZ05], linear codes are of particular interest due to their simplicity.

Recall from Chapter 2 that determining the capacity of an acyclic memoryless wired network can be reduced to a convex optimization problem as follows:

$$
\begin{aligned}
&\max \alpha^T h && (4.1)\\
&\text{s.t. } h \in \bar{\Omega}_n^* \\
&h_{X_e} \le C_e \qquad \text{for any edge } e \\
&h(X_{S_1}, \ldots, X_{S_s}) = \sum_{i=1}^{s} h(X_{S_i}) \\
&h(X_{in}^i, X_{out}^i(j)) = h(X_{in}^i) \quad j = 1, \ldots, X_{out}^i(j) \ \text{ for each nonsource node } i \\
&h(X_{in}^l, X_d^l(j)) = h(X_d^l(j)) \ \ l \text{ is a sink node}
\end{aligned}
$$

where s is the total number of sources in the network, X_{in}^i represent all the inputs of node i, and C_e is the link capacity for edge e.

If one is interested in finding the linear coding capacity of a network then all entropies of network random variables in the optimization formulation (4.1), can be replaced by rank of the matrices that relate the relevant variables to the sources. In such cases the following definition turns out to be useful [Cha07b, YLCZ06].

Definition 4.2.2 (Linear representable vectors) *A $2^n - 1$ dimensional vector g whose entries are indexed by subsets of $\mathcal{N} = \{1, \ldots, n\}$ is called a linear representable vector (also a linear rank vector[1]) if there exist n matrices $\{v_1, \cdots, v_n\}$, each of*

[1]We use the term "linear rank vector" as to avoid confusion with the matroid rank functions which will be discussed later in this chapter. We may drop the word "linear" when the meaning is clear from the context.

dimension $\sigma \times \tau$ over a finite field $GF(q)$ such that for any $\alpha \subseteq \mathcal{N}$,

$$g_\alpha = rank(\oplus_{i\in\alpha} v_i) \tag{4.2}$$

where $\oplus_{i\in\alpha} v_i$ denotes the space spanned by rows of $\{v_i,\ i \in \alpha\}$[2]. If in particular $\sigma = 1$, then g is called "scalar-representable" while if $\sigma > 1$ it is called "vector-representable" or "multilinear representable". We denote the space of all linear representable vectors of dimension $2^n - 1$ by Γ_n^r and the region of all scalar-representable rank vectors by Γ_n^{sr}. Moreover we call $\frac{1}{\sigma}g$ a normalized representable vector and denote the corresponding space of all normalized representable vectors by Ω_n^r.

It turns out that every linear representable vector is a multiple of an entropy vector. The following theorem whose proof we explained in Chapter 3 states this fact.

Theorem 4.2.3 (Linear representables and entropy vectors) *[YLCZ06] Define g to be a $2^n - 1$ dimensional rank vector and let the set of $\sigma \times \tau$ matrices $\{v_1, \ldots, v_n\}$ over GF(q) form a representation for it such that $g_\alpha = rank(\oplus_{i\in\alpha} v_i)$, then $\log(q)g$ is an entropy vector of random variables $X_1, \ldots, X_n$ where,*

$$\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} - & v_1 & - \\ & \vdots & \\ - & v_n & - \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix}. \tag{4.3}$$

The random variables $X_1, \ldots, X_n$ constructed as such are called "GF(q) linearly-related" random variables.

Noting that $\bar{\Gamma}_n^*$ is a convex cone, it immediately follows from Theorem 4.2.3 that $\Gamma_n^r \subseteq \text{con}(\Gamma_n^r) \subseteq \bar{\Gamma}_n^*$ where con(.) denotes the convex hull.

[2]Sometimes a "linear representable vector" is defined as a vector g whose entries are $g_\alpha = \log q \cdot \text{rank}(\oplus_{i\in\alpha} v_i)$ [Cha07b]. However here we will stick to (4.2).

Remark: Note that a normalized representable vector is in fact an entropy vector normalized by the log of the alphabet size since the alphabet size of each random variable in (4.3) is q^σ. In the context of network coding capacity and parallel to the arguments of Chapter 2, the entropy of a vector-valued network random variable X that is linearly related to the s-dimensional vector of sources X_S through a $\sigma \times s$ matrix A, i.e., $X = AX_S$ over T channel uses, will be $\log q \cdot T \cdot \text{rank}(A)$, which when normalized by $\log(q^{\sigma T})$ gives $\frac{1}{\sigma}\text{rank}(A)$ as the normalized entropy.

As stated in Section 4.1, the linear coding capacity of a network is connected to the region of linear representable vectors. This connection can be seen collectively from Definition 4.2.1, optimization formulations (4.1), Definition 4.2.2, and Theorem 4.2.3. The following theorem formalizes this fact:

Theorem 4.2.4 (Linear coding capacity) *Let $X_1, \cdots, X_n$ denote the variables of a wired network. The maximum weighted sum rate achieved by linear codes can be obtained from the following optimization over the convex hull of normalized representable region:*

$$
\begin{aligned}
&\max \alpha^T h \qquad (4.4)\\
&s.t.\ h \in con(\Omega_n^r)\\
&h_{X_e} \le C_e \qquad \text{for any edge } e\\
&h(X_{S_1}, \ldots, X_{S_s}) = \sum_{i=1}^{s} h(X_{S_i})\\
&h(X_{in}^i, X_{out}^i(j)) = h(X_{in}^i) \quad j = 1, \ldots, X_{out}^i(j) \ \text{ for each nonsource node } i\\
&h(X_{in}^l, X_d^l(j)) = h(X_d^l(j)) \ \ l \text{ is a sink node}
\end{aligned}
$$

where con(.) represents the convex hull, and as in formulation (4.1), s is the total number of sources in the network, X_{in}^i represent all the inputs of node i, and C_e is the link capacity for edge e. In particular if one is interested in the best scalar linear codes, then $con(\Omega_n^r)$ should be replaced with $con(\Gamma_n^{sr})$ in (4.4).

It is known that the linear representable entropies satisfy an inequality called the *Ingleton inequality* [Ing71] which does not hold for all entropy vectors and hence it proves that the region of all linear representable entropies is a strict subset of Γ_n^*. This justifies the sub-optimality of linear codes.

Recently there has been some progress in determining the convex cone of linear representable vectors. It has been shown in [HRSV00] that not only the set of Shannon and Ingleton inequalities are necessary conditions for a vector to be a rank vector of 4 random variables but they are also sufficient conditions. In other words the convex cone of rank region (linear representable region) of 4 random variables is completely characterized by the set of Ingleton and Shannon-type inequalities. The approach of [HRSV00] is based on finding all the extreme rays of the set defined by Ingleton and Shannon-type inequalities and finding a linear representation for all those points. Although this approach finds the whole linear representable region for 4 variables it is not extendable to more than 4. Moreover it does not find the region of *scalar* representable entropies. In more recent works however [DFZ10, CGK10, Kin09], it has been shown that for more than 4 random variables, rank vectors satisfy inequalities other than Shannon and Ingleton. In particular [DFZ10] has determined the convex cone of rank region for 5 random variables by discovering new inequalities for 31-dimensional rank vectors and showing the representability of all the rays of such region. For more than 5 random variables the full characterization of (scalar or vector) linear representable entropies remains open.

In an attempt to find the capacity region of wired networks based on Theorem 4.2.4, we will study the *scalar* linear representable region under different assumptions.

4.3 Scalar Linear Representable Region

In this section we will study the scalar linear representable region and characterize it explicitly for 4 random variables. The method we present for determining this region is in principle extendable to a larger number of random variables. In the next section we will show that the same framework can be used to obtain the entropy region of all feasible entropy vectors in networks with two sources.

4.3.1 General Technique

Based on Theorem 4.2.3 one can obtain the set of all linear representable vectors g for n random variables by finding all the possible rank vectors of the matrix $M = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$.

Theorem 4.3.1 (Scalar linear representable entropies) Γ_n^{sr} *can be obtained from all the rank vectors of an* $n \times n$ *matrix* M *with entries over an arbitrary finite field* $GF(q)$.

Proof: Follows trivially from Theorem 4.2.3. □

Therefore in order to compute Γ_4^{sr} for example, we need to find all rank vectors obtained from a 4×4 matrix M for which we need to consider all relative dependencies between rows of M. In what follows this will be explained in detail.

4.3.2 Scalar Linear Representable Region of 4 Random Variables

In this section we consider characterizing the scalar linear representable region for 4 random variables. Note that based on Theorem 4.3.1 we need to obtain all rank vectors from a 4×4 matrix.

4.3.2.1 Deriving All Rank Vectors of a 4×4 Matrix

To obtain the rank vectors based on Theorem 4.3.1, we need to determine the possible dependency relations among the rows of a 4×4 matrix. The general idea for such approach is to first determine the possible pairwise dependencies of the rows of the matrix and for each such obtained structure, further determine the triple dependencies, etc. In what follows we will explain this method in detail,

Theorem 4.3.2 (Scalar linear representable vectors of 4 random variables) *There are a total of 68 different scalar linear representable vectors for 4 random variables.*

Proof: [Obtaining rank vectors from a 4×4 matrix] Let the rows of the 4×4 matrix be denoted by v_1, v_2, v_3, and v_4. We examine the dependency relations in different scenarios based on the number of zero rows and we assume that the underlying field size is arbitrary. First assume that there are no zero rows.

Scenario 1: No zero rows

We begin by determining the pairwise dependencies. Note that for a 4×4 matrix whose rows are v_1, v_2, v_3, and v_4, five cases are possible. By denoting the pairwise dependent rows inside parenthesis, we obtain the following cases,

1. $(v_1)(v_2)(v_3)(v_4)$: no two rows are pairwise dependent (i.e., no two rows are aligned).

2. $(v_i, v_j)(v_k)(v_l)$: v_i and v_j are aligned and the other two rows v_k and v_l are independent.

3. $(v_i, v_j)(v_k, v_l)$: rows v_i, v_j are aligned and the other two rows v_k, v_l are also aligned.

4. $(v_i, v_j, v_k)(v_l)$: rows v_i, v_j, v_k are aligned to each other and independent from v_l.

5. (v_1, v_2, v_3, v_4): all four rows are aligned.

Note that these 5 cases correspond to different partitions of v_1, v_2, v_3, v_4. Now we will examine each case separately,

1. $(v_1)(v_2)(v_3)(v_4)$: Since all pairs of rows are independent, we have $r_{ij} = 2,\ \forall i, j$, and we can consider the following possibilities for the triple dependencies:

   (a) there is no set of dependent triplets, i.e., $r_{ijk} = 3,\ \forall i, j, k$, then the whole set of rows can be either dependent or independent,

      i. the four rows are independent and therefore $r_{1234} = 4$. In this case the corresponding rank vector will be $(1\ 1\ 1\ 1, 2\ 2\ 2\ 2\ 2\ 2, 3\ 3\ 3\ 3, 4)$.

      ii. the four rows are linearly dependent and therefore $r_{1234} = 3$. Hence, the corresponding rank vector will be $(1\ 1\ 1\ 1, 2\ 2\ 2\ 2\ 2\ 2, 3\ 3\ 3\ 3, 3)$.

   (b) there is one triple set that is linearly dependent and independent of the 4-th row. Without loss of generality assume that that linear dependent triplet is v_1, v_2, v_3. Therefore we obtain that $r_{123} = 2$. However rank of the other triplets will be still 3. Clearly we also have $r_{1234} = 3$. Therefore the resulting rank vector in this case will be, $(1\ 1\ 1\ 1, 2\ 2\ 2\ 2\ 2\ 2, 2\ 3\ 3\ 3, 3)$. Note that considering the permutations of this vector, this case results in 4 vectors in total.

   (c) more than one triple set is linearly dependent. This essentially means that all triples are dependent and therefore $r_{ijk} = 2,\ \forall i, j, k$. Moreover this also gives $r_{1234} = 2$. In other words the rank vector corresponding to this case is obtained to be $(1\ 1\ 1\ 1, 2\ 2\ 2\ 2\ 2\ 2, 2\ 2\ 2\ 2, 2)$.

2. $(v_i, v_j)(v_k)(v_l)$: Without loss of generality assume that we have $(v_1, v_2)(v_3)(v_4)$ which immediately gives $r_{12} = 1$ and $\forall (i, j) \neq (1, 2), r_{ij} = 2$. Since v_1 and v_2 are aligned, we can represent them by a unit vector v_1' in their direction. For

the matter of determining triplet dependencies, therefore, we need to examine the dependencies of v_1', v_3, v_4. Two cases are feasible,

(a) there is no set of dependent triplets in v_1', v_3, v_4. Therefore we have $r_{123} = r_{124} = 2$ and $r_{134} = r_{234} = 3$. Clearly $r_{1234} = 3$ and we obtain the rank vector $(1\ 1\ 1\ 1, 1\ 2\ 2\ 2\ 2\ 2, 2\ 2\ 3\ 3, 3)$. Considering permutations of this vector this case yields 6 vectors in total.

(b) the triplet v_1', v_3, v_4 is linearly dependent. In this case we have $r_{ijk} = 2,\ \forall i, j, k$ and $r_{1234} = 2$ giving the rank $(1\ 1\ 1\ 1, 1\ 2\ 2\ 2\ 2\ 2, 2\ 2\ 2\ 2, 2)$. Considering the permutations, we obtain a total of 6 vectors from this case as well.

3. $(v_i, v_j)(v_k, v_l)$: Without loss of generality assume we have $(v_1, v_2)(v_3, v_4)$. This assumption alone determines all the rank vector entries. In particular $r_{12} = r_{34} = 1$, $r_{13} = r_{14} = r_{23} = r_{24} = 2$. Moreover $r_{ijk} = 2,\ \forall i, j, k$ and $r_{1,2,3,4} = 2$, yielding the rank vector $(1\ 1\ 1\ 1, 1\ 2\ 2\ 2\ 2\ 1, 2\ 2\ 2\ 2, 2)$. Three permutations for i, j, k are possible in total, i.e.—the structures $(v_1, v_2)(v_3, v_4)$, $(v_1, v_3)(v_2, v_4)$, and $(v_1, v_4)(v_2, v_3)$—and therefore this case gives 3 rank vectors in total.

4. $(v_i, v_j, v_k)(v_l)$: Without loss of generality assume that we have $(v_1, v_2, v_3)(v_4)$. This implies that $r_{12} = r_{13} = r_{23} = 1$ and $r_{14} = r_{24} = r_{34} = 2$. Moreover $r_{123} = 1$, $r_{124} = r_{134} = r_{234} = 2$, and $r_{1234} = 2$. Therefore we get the rank vector $(1\ 1\ 1\ 1, 1\ 1\ 2\ 1\ 2\ 2, 1\ 2\ 2\ 2, 2)$. Considering the permutations, this case results in a total of 4 rank vectors.

5. (v_1, v_2, v_3, v_4): In this case all the rows are aligned and the rank vector can be easily obtained to be $(1\ 1\ 1\ 1, 1\ 1\ 1\ 1\ 1\ 1, 1\ 1\ 1\ 1, 1)$.

Scenario 2: One zero row

Assume without loss of generality that the first row is the all zero vector, i.e., $v_1 = 0$. This immediately gives $r_1 = 0$. For v_2, v_3, and v_4 the following cases are possible:

1. $(v_2)(v_3)(v_4)$: none of the other three rows are aligned. Therefore we have $r_{12} = r_{13} = r_{14} = 1$ and $r_{23} = r_{24} = r_{34} = 2$. In this case these three rows can either be linearly dependent or independent,

   (a) v_2, v_3, v_4 are linearly independent. Therefore we have $r_{123} = r_{124} = r_{134} = 2$ and $r_{234} = r_{1234} = 3$. The rank vector is $(0\ 1\ 1\ 1, 1\ 1\ 1\ 2\ 2\ 2, 2\ 2\ 2\ 3, 3)$. Taking into account the possible permutations we obtain 4 vectors in total.

   (b) v_2, v_3, v_4 are linearly dependent. The difference with the case where v_2, v_3, v_4 are independent will be that $r_{234} = r_{1234} = 2$. Therefore the resulting rank vector is $(0\ 1\ 1\ 1, 1\ 1\ 1\ 2\ 2\ 2, 2\ 2\ 2\ 2, 2)$. Considering the possible permutations again gives 4 vectors.

2. $(v_j, v_k)(v_l),\ j, k, l \neq 1$: In this case two of the nonzero rows are aligned and independent from the last nonzero row. Without loss of generality assume $(v_2, v_3)(v_4)$. The rank vector is obtained to be $(0\ 1\ 1\ 1, 1\ 1\ 1\ 1\ 2\ 2, 1\ 2\ 2\ 2, 2)$. Note that the zero row could be any of the four rows, and for any chosen zero row, one can consider any two of the three remaining nonzero rows aligned. Therefore there are a total of 12 vectors resulting from this case (counting the permutations).

3. (v_2, v_3, v_4): The rank vector in this case will be $(0\ 1\ 1\ 1, 1\ 1\ 1\ 1\ 1\ 1, 1\ 1\ 1\ 1, 1)$. Taking into account the feasible permutations we obtain a total of 4 vectors from this case.

Scenario 3: Two zero rows

Assume without loss of generality that $v_1 = v_2 = 0$. Therefore $r_1 = r_2 = 0$. Now the remaining two rows can be either aligned or independent,

1. $(v_3)(v_4)$: In this case the other two rows are independent. Therefore the resulting rank vector will be $(0\ 0\ 1\ 1, 0\ 1\ 1\ 1\ 1\ 2, 1\ 1\ 2\ 2, 2)$. Note that the two zero

rows can be chosen in six different ways and hence there are a total of 6 vectors in this case.

2. (v_3, v_4): In this case the two nonzero rows are aligned. Therefore we can easily obtain the rank vector for this case as $(0\ 0\ 1\ 1, 0\ 1\ 1\ 1\ 1\ 1, 1\ 1\ 1\ 1, 1)$. Again considering permutations we obtain a total of 6 vectors from this case.

Scenario 4: Three zero rows

Without loss of generality assume the nonzero row is v_4. The obtained rank vector in this case will be $(0\ 0\ 0\ 1, 0\ 0\ 1\ 0\ 1\ 1, 0\ 1\ 1\ 1, 1)$ and considering permutations we get a total of 4 rank vectors from this case (the assumption of three zero rows).

Scenario 5: All rows are zero

This will be the trivial all-zero rank vector.

Putting together all the rank vectors obtained in Scenarios 1–5, we obtain 68 rank vectors. □

Remark: An alternative approach for determining the rank vectors resulting from a 4×4 matrix M is to first lower triangularize M, i.e.,[1]

$$M = \begin{bmatrix} \mathtt{x} & & & \\ \mathtt{x} & \mathtt{x} & & \\ \mathtt{x} & \mathtt{x} & \mathtt{x} & \\ \mathtt{x} & \mathtt{x} & \mathtt{x} & \mathtt{x} \end{bmatrix} \tag{4.5}$$

and then consider all the relations between rows. In particular we can define p_i to denote the largest column index for which row i has a nonzero entry, therefore $0 \leq p_i \leq i$. Then, e.g., $p_i = 0$ would imply that row i is totally zero. Moreover if $p_i \neq p_j$ we can immediately conclude that rows i and j are independent. However if two or more p_i's are equal then the relations could be more complicated and all of them should be considered. For example, consider $p_1 = 0, p_2 = p_3 = p_4 = 2$. That

[1]The matrix M can be assumed to be lower triangular without loss of generality due to repeated use of the Bezout identity.

would be the following structure,

$$M = \begin{bmatrix} 0 & \\ \mathtt{x} & \mathtt{x} \\ \mathtt{x} & \mathtt{x} \\ \mathtt{x} & \mathtt{x} \end{bmatrix}. \tag{4.6}$$

Now note that for the rank of singletons we have $r_1 = 0$ and $r_j = 1,\ j \neq 1$. For pairs that include row 1, $r_{1j} = 1,\ j \neq 1$. However to determine the rest of the pairwise ranks we should consider the different feasible dependencies among rows 2, 3, and 4. Denoting row i by v_i as previously stated, three different cases can be considered for v_2, v_3, and v_4, i.e., $(v_2)(v_3)(v_4)$,$(v_j, v_k)(v_l)$ $j, k, l \neq 1$, and (v_2, v_3, v_4). Each of these cases results in a different rank vector. In particular note that we always have $r_{1234} = r_{234} \leq 2$ because there are only two nonzero columns. There are a total of 5! combinations for values that $p_1, \ldots, p_4$ can take and for each combination all dependencies should be taken into account.

4.3.2.2 Characterizing Γ_4^{sr}

Based on the rank vectors that we obtained in the last part, we can now characterize the region of scalar linear representable vectors for 4 random variables.

Theorem 4.3.3 (Scalar rank region of 4 random variables) *The region of scalar representable entropies of 4 random variables is obtained from the convex hull of 68 rank vectors. The convex cone of this region has 27 rays which can be seen in Table 4.1.*

Proof: The scalar linear representable region is characterized from the convex hull of all the 68 rank vectors found in Theorem 4.3.2. Using the software package PORTA [POR] which uses a Fourier-Motzkin elimination, we compute the convex hull of these 68 vectors to obtain the region of 15-dimensional scalar representable entropy vectors. All these 68 vectors correspond to corner points and their convex hull is represented

Table 4.1: Rays of the scalar linearly representable region of 4 random variables

	$(h_1,\ h_2,\ h_3,\ h_4,\ h_{12},\ h_{13},\ h_{14},\ h_{23},\ h_{24},\ h_{34},\ h_{123},\ h_{124},\ h_{134},\ h_{234},\ h_{1234})$
(1)	(1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3)
(2)	(1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2)
(3)	(1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2)
(4)	(1, 1, 1, 1, 2, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2)
(5)	(1, 1, 1, 1, 2, 2, 1, 2, 2, 2, 2, 2, 2, 2, 2)
(6)	(1, 1, 1, 1, 2, 2, 2, 1, 2, 2, 2, 2, 2, 2, 2)
(7)	(1, 1, 1, 1, 2, 2, 2, 2, 1, 2, 2, 2, 2, 2, 2)
(8)	(1, 1, 1, 1, 2, 2, 2, 2, 2, 1, 2, 2, 2, 2, 2)
(9)	(0, 0, 0, 1, 0, 0, 1, 0, 1, 1, 0, 1, 1, 1, 1)
(10)	(0, 0, 1, 0, 0, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1)
(11)	(0, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2)
(12)	(0, 1, 0, 0, 1, 0, 0, 1, 1, 0, 1, 1, 0, 1, 1)
(13)	(1, 0, 1, 1, 1, 2, 2, 1, 1, 2, 2, 2, 2, 2, 2)
(14)	(1, 1, 0, 1, 2, 1, 2, 1, 2, 1, 2, 2, 2, 2, 2)
(15)	(1, 1, 1, 0, 2, 2, 1, 2, 1, 1, 2, 2, 2, 2, 2)
(16)	(1, 0, 0, 0, 1, 1, 1, 0, 0, 0, 1, 1, 1, 0, 1)
(17)	(0, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1)
(18)	(0, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1)
(19)	(0, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1)
(20)	(1, 0, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1)
(21)	(1, 0, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1)
(22)	(1, 1, 0, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1)
(23)	(0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1)
(24)	(1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1)
(25)	(1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1)
(26)	(1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1)
(27)	(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1)

by 86 inequalities or, equivalently, hyperplanes. Among the hyperplanes, 50 of them pass through the origin and are used for computing the rays of the convex cone of the rank vectors. Moreover computing their convex *cone* results in 27 rays which are depicted in Table 4.1. □

Comparisons:

- Recall that in Section 3.5, we obtained the corner points of 4 quasi-uniform random variables of alphabet size 2 [FSH08]. In that case there are 67 corner points (listed in Table 3.6) and, interestingly, the only differing vector between

those 67, and the corner points of the scalar representable region is the vector $(1\ 1\ 1\ 1, 2\ 2\ 2\ 2\ 2\ 2, 2\ 2\ 2\ 2, 2)$, which is not representable over binary alphabet size. In matroid theory, this vector corresponds to the rank of $U_{2,4}$ matroid. Therefore all the other corner points of scalar ranks correspond to the entropies of quasi-uniform random variables with binary alphabet size.

- Another comparison with the results of [HRSV00] that has obtained the rays of the linear representable region (either scalar or vector), shows that the rays of the scalar rank region are exactly the scalar representable rays of the vector rank region (i.e., there are not any extra rays for the scalar linear representable region). In fact [HRSV00] finds 35 rays for the set defined by Shannon-type and Ingleton inequalities and shows that they are all representable. However only 27 of those are scalar representable, which exactly correspond to the rays of the convex cone region that we have found.

4.4 Scalar Linear Codes for Networks with Two Sources

As stated in Theorem 4.2.4, the region of linearly representable entropy vectors is important in finding the capacity region of a given network. In the last section we found the scalar region for 4 random variables. However most networks involve more than 4 variables and finding the linear representable entropy region becomes computationally hard when the number of random variables grows. Therefore in an attempt to find linear solutions for a general network, one may simplify the problem by limiting the number of sources to some number s. This in effect upperbounds the values of the joint entropies of the random variables to s. In this section by letting $s = 2$, we consider networks with 2 sources.

Now we can state the following theorem about the region of such scalar repre-

sentable vectors,

Theorem 4.4.1 *The region of scalar representable vectors whose entries are bounded by 2, is obtained from all rank vectors of an $n \times 2$ matrix.*

Proof: In general from Theorem 4.2.3 any scalar linear representable vector can be obtained from the rank vector of an $n \times n$ matrix. Since the entries of the vector in this case are bounded by 2 by assumption, that means that the $n \times n$ matrix is rank 2 and therefore the scalar representable vector in this case can be considered to be the rank vector of an $n \times 2$ matrix. □

Denote the space of scalar rank vectors obtained from an $n \times 2$ matrix by $\Gamma^{sr}_{n\times 2}$.

Theorem 4.4.2 *Given a network with two sources, a scalar linear solution for it can be found by solving a convex optimization over the convex hull of $\Gamma^{sr}_{n\times 2}$.*

Proof: Follows from Theorems 4.4.1 and replacing $\text{con}(\Omega^r_n)$ by $\text{con}(\Gamma^{sr}_{n\times 2})$ in 4.2.4. □

Henceforth we will try to find the rank vectors of an $n \times 2$ matrix.

4.4.1 Rank Vectors of an $n \times 2$ Matrix

The underlying principle for finding all rank vectors of an $n \times 2$ matrix is the same as that of a 4×4 matrix which was discussed in the last section. In fact similar to the case of 4×4 matrix, in considering different structures for the $n \times 2$ matrix, we first need to consider whether each row is zero or not zero and for the nonzero rows consider the pairwise dependencies first. The following shows an example of such a structure,

$$M = \begin{bmatrix} \mathtt{x} & \mathtt{x} \\ \mathtt{x} & \mathtt{x} \\ 0 & 0 \\ \mathtt{x} & \mathtt{x} \end{bmatrix}. \tag{4.7}$$

However since the matrix is $n \times 2$ versus the $n \times n$ general matrix of the last section, we can obtain nice results and simplify things considerably in this case. As an instance an immediate consequence of having a rank 2 matrix is that knowing the nonzero rows and their pairwise ranks we can determine the whole rank vector.

Lemma 4.4.3 *The entries of the rank vector of an $n \times 2$ matrix that correspond to ranks of a set of more than 2 rows, can be obtained from the pairwise ranks.*

Proof: Follows trivially from the fact that M is rank 2. In fact if α is the set of indices of a collection of rows, s.t. $|\alpha| \geq 2$ then $r_\alpha = \max_{i,j \in \alpha} r_{ij}$. □

The following lemma lays the foundation for characterizing all rank vectors of an $n \times 2$ matrix.

Lemma 4.4.4 *Let $K = \{i_1, \cdots, i_k\}$ denote the set of indices of the nonzero rows of an $n \times 2$ matrix M. Then there is a bijection between all rank vectors of M and set partitions of K.*

Proof: Let P be the set of all partitions of K and $\mathcal{R}$ be the set of all rank vectors obtained from M. We show that there is a bijective mapping $\Pi : P \to \mathcal{R}$.

First we show the existence of such mapping by showing that for any given partition of K, e.g., $p \in P$ we can construct an $n \times 2$ matrix with $\{i_1, \ldots, i_k\}$ as the set of its nonzero rows as follows: Let $p = S_1/S_2/\ldots/S_t$ be a partition of K. Let $\{v_1, \ldots, v_t\}$ be t pairwise independent 1×2 vectors. This is always possible by assuming that the underlying finite field is sufficiently large [1]. For $l = 1, \ldots, n$ if $l \in S_j$ then set $M_l = v_j$ where M_l represents the l-th row of M. Define $\Pi(p)$ to be the rank vector of the matrix M so constructed.

To complete the proof, we show that this mapping is one-to-one and onto. Assume $\Pi(p) = \Pi(\hat{p})$ and $p \neq \hat{p}$. Therefore there exists $i, j \in K$ such that i and j belong to the same partition in p and to different partitions in $\hat{p}$. If we let $\Pi(p)_{\{i,j\}}$ denote the

[1] It can be easily shown that the total number of pairwise independent 1×2 vectors over GF(q) for a prime q is at least $q + 1$. Hence it is enough to choose $q \geq t - 1$.

rank of rows i and j, it is easy to check that $\Pi(p)_{\{i,j\}} = 1$, and $\Pi(\hat{p})_{\{i,j\}} = 2$, which contradicts the assumption. Therefore $\Pi(.)$ is one-to-one. Next we show that $\Pi(.)$ is onto by constructing a partition $p \in P$ for every $r \in \mathcal{R}$ s.t. $\Pi(p) = r$. From Lemma 4.4.3 it is enough to consider those entries of the rank vector r that correspond to the pairwise ranks. In other words, $r = \Pi(p)$ if there exists a $p \in P$ s.t. $\Pi(p)_{\{i,j\}} = r_{ij}$ for all $i, j \in K$. Let i, j be in the same set in partition p if and only if $r_{ij} = 1$. The partition p is well-defined if we can show the following relation:

$$r_{ij} = r_{jk} = 1 \Rightarrow r_{ik} = 1,$$

which holds by the definition of the rank function. This concludes the proof. □

Although Lemma 4.4.4 gives a nice algorithm for determining all the rank vectors of an $n \times 2$ matrix by determining all the set partitions of the nonzero rows of the corresponding matrix, one cannot easily determine if, for a given vector, there exists a valid partitioning. In what follows we will try to answer that question. However we first need to define the following binary relation.

Definition 4.4.5 *Assume that r is a $2^n - 1$ dimensional vector whose entries are indexed by subsets of $\{1, \cdots, n\}$. Let K be defined as, $K = \{i \mid r_i = 1,\ i \in \{1, \ldots, n\}\}$. Then for $i, j \in \{1, \ldots, n\}$ we define the binary relation $\sim$ as follows:*

$$i \sim j \ \Leftrightarrow\ \exists\ T \subset \{1, 2, ..., n\} : i, j \in T \cap K \text{ and } r_T = 1. \tag{4.8}$$

Note that the binary relation of Definition 4.4.5 is both reflexive and symmetric. Therefore it will be an equivalence relation if it is transitive as well. Note that we have the following lemma,

Lemma 4.4.6 *Let r be a $2^n - 1$ dimensional vector whose entries are indexed by subsets of $\{1, \ldots, n\}$. If entries of r satisfy the Shannon inequalities, then the binary relation defined on r as in Definition 4.4.5 is an equivalence relation.*

Proof: We only need to show that under the assumption of Shannon inequality satisfaction, the binary relation is transitive. Therefore assume that for $i, j, k \in \{1, \ldots, n\}$ where $r_i = r_j = r_k = 1$ we have $i \sim j$ and $j \sim k$ and we need to show that $j \sim k$. Since $i \sim j$, based on definition, there is a set $T_1 \subseteq \{1, \ldots, n\}$ such that $i, j \in T_1$ and $r_{T_1} = 1$. Similarly for j and k there is a set $T_2 \subseteq \{1, \ldots, n\}$ such that $j, k \in T_2$ and $r_{T_2} = 1$. Note that $T_1 \cap T_2$ at least includes j. Since the entries of r satisfy Shannon inequalities we have,

$$1 = r_{T_1} \leq r_{T_1 \cup T_2} \leq r_{T_1} + r_{T_2} - r_{T_1 \cap T_2} = 2 - r_{T_1 \cap T_2}. \tag{4.9}$$

Moreover for $r_{T_1 \cap T_2}$ we obtain that,

$$1 = r_j \leq r_{T_1 \cap T_2} \leq r_{T_1} = 1 \tag{4.10}$$

meaning that $r_{T_1 \cap T_2} = 1$ and therefore from (4.9) we get $r_{T_1 \cup T_2} = 1$. Since $T_1 \cup T_2$ includes both i and k, based on Definition 4.4.5, $i \sim k$. □

Corollary 4.4.7 *Let r be a $2^n - 1$ dimensional vector and $\sim$ be the binary relation defined on r based on Definition 4.4.5. Assume for some $1 \leq i, j, k \leq n$, where $r_i = r_j = r_k = 1$, we have $i \sim j$, i.e., $\exists T, i, j \in T, r_T = 1$ and $j \sim k$, which also means $\exists T', j, k \in T', r_{T'} = 1$. Then if the entries of r satisfy the Shannon inequalities,*

1. *For all subsets of T, i.e., $\tilde{T} \subseteq T$ such that $i \in \tilde{T}$ or $j \in \tilde{T}$, we also have $r_{\tilde{T}} = 1$.*
2. *$r_{T \cup T'} = 1$.*

Proof: Part 1 follows trivially from the Shannon inequalities and part 2 is a direct consequence of Lemma 4.4.6. □

Now we can state the main theorem that allows one to determine if a given vector is the rank vector of an $n \times 2$ matrix.

Theorem 4.4.8 *Let $r \in \{0,1,2\}^{2^n-1}$ be a vector whose entries are indexed by subsets of $\{1,\ldots,n\}$. Moreover assume that for $i = 1,\ldots,n,\ r_i \in \{0,1\}$. Then r is a rank vector of an $n\times 2$ matrix if and only if the entries of r satisfy the Shannon inequalities.*

Proof: First note that if r is a rank vector of an $n\times 2$ matrix then it is an entropy vector and its entries will satisfy the Shannon inequalities inevitably. Now assume that the entries of $r \in \{0,1,2\}^{2^n-1}$ satisfy Shannon inequalities. We want to show that r will be a rank vector. If the entries of r satisfy Shannon inequalities, then based on Lemma 4.4.6 the binary relation defined on r will be an equivalence relation and therefore induces equivalence classes. Note that if an element of the set $\{1,\ldots,n\}$, e.g., i, does not belong to any of the induced equivalence classes then it essentially means that $r_i = 0$. This is due to the fact that r_i can only be zero or one and if it is one, at least we have $i \sim i$ and therefore i will belong to an equivalence class. Therefore the equivalence classes induced by $\sim$ form a partition p for $K = \{i | r_i = 1, i = 1,\ldots,n\}$. Based on Lemma 4.4.4, there is a valid rank vector $r' = \Pi(p)$ corresponding to this partitioning. We prove that $r = r'$ by showing that for all $T \subseteq \mathcal{N}, r_T = r'_T$.

1. $T \subseteq \mathcal{N} \setminus K$: We show that for all $T \subseteq \mathcal{N} \setminus k$ we have $r_T = r'_T = 0$. Since r satisfies the Shannon inequalities $r_T \leq \sum_{i\in T} r_i$, however since if $i \in T$ it is not in the set K, then $r_i = 0$ and therefore $r_T = 0$. Moreover by construction of $r' = \Pi(p)$, r' is also zero for such T. Now note that for all other subsets $T \subseteq \mathcal{N}, T \nsubseteq (\mathcal{N} \setminus K)$, $r_T \neq 0$, and $r'_T \neq 0$. The former due to Shannon inequality and the latter by construction. Therefore for such T, r_T and r'_T are either 1 or 2.

2. $T \subseteq K$: As stated in the last case, for $T \subseteq \mathcal{N}$, r_T and r'_T can only take values 1 or 2. Therefore we prove the equality of r_T and r'_T for this class of subsets by showing that for $T \subseteq K$, $r_T = 1$ if and only if $r'_T = 1$. First assume that $r_T = 1$. Then from the definition of $\sim$, $T \cap K = T \in C$ where C is an

equivalence class under the relation $\sim$. From the construction of $r' = \Pi(p)$ in Lemma 4.4.4, C is also a partition of K and hence $r'_T = 1$. Conversely assume that $r'_T = 1$. Therefore again we have $T \cap K = T \in C$. Since r satisfies the Shannon inequalities for all equivalence classes we have $r_C = 1$ and $\forall T \subseteq C$, also $r_T = 1$.

3. $T = \alpha \cup \beta, \alpha \subseteq K, \beta \subseteq \mathcal{N} \setminus K$: Note that since r satisfies Shannon we have, $r_\alpha \leq r_T \leq r_\alpha + r_\beta$. However since $\beta \subseteq \mathcal{N} \setminus K$ based on case 1, $r_\beta = 0$ and hence $r_T = r_\alpha$. Moreover since r' is a rank vector based on construction, therefore it satisfies Shannon and therefore by a similar argument $r'_T = r'_\alpha$. However since $\alpha \subseteq K$ based on case 2, $r_\alpha = r'_\alpha$.

□

Remark 1: We want to emphasize the need for Shannon conditions in the proof of Theorem 4.4.8. In particular if r does not satisfy Shannon, entries of r could be such that $\sim$ still be an equivalence relation on r. In such case however, one does not necessarily have $r_C = 1$ and $\forall T \subseteq C, r_T = 1$. The 7-dimensional vectors $(1\ 1\ 1, 2\ 2\ 2, 1)$ and $(1\ 1\ 1, 1\ 1\ 1, 2)$ which do not satisfy the Shannon inequalities are two such examples, both having a single equivalence class $\{1, 2, 3\}$ based on Definition 4.4.5.

Remark 2: Note that Theorem 4.4.8 suggests that, in order to compute the scalar linear capacity of a network with 2 sources, we should include all the Shannon inequalities as constraints in the network optimization problem. However the number of non-redundant Shannon inequalities is $n + \binom{n}{2}2^{n-2}$ [Yeu02] which is exponential in n. Nonetheless the number of joint entropies that appear in network constraints (topology constraints of the network) are usually much less than the whole number of joint entropies of the random variables of the network. Therefore the hope is that in such cases one need not impose all the Shannon inequality constraints. Extension of the approach that resulted in Theorem 4.4.8 can be useful in determining the region

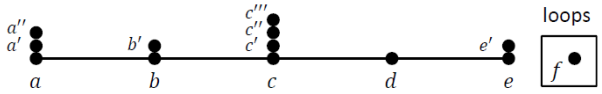


Figure 4.1: Geometric representation for a rank-2 matroid

of projection of rank vectors into particular subsets of its entries.

Remark 3: For those familiar with the matroid theory, an alternative way of expressing the result of Theorem 4.4.8 is by establishing that rank-2 matroids are representable (matroid representability is discussed in Section 4.5). In this context, a rank-2 matroid can be represented as in Fig. 4.1, where every point shows a matroid element, and pairwise dependent elements are shown by touching points grouped into a column. Assume there are t columns of touching points in Fig. 4.1. One can easily show that there are at least $q+1$ pairwise independent 2-dimensional vectors over GF(q). Hence, if we let $q \geq t-1$, we can find t pairwise independent 2-dimensional vectors over GF(q). Assigning each of the t vectors to all the elements in a column of Fig. 4.1, yields a representation for the matroid.

Corollary 4.4.9 *Among scalar codes for a network with 2 sources, linear codes are optimal.*

Proof: Recall that entropy vector of network random variables under *any* scalar network coding assumption, satisfies the Shannon inequalities. Since based on Theorem 4.4.8, Shannon inequalities are also sufficient for characterizing $\Gamma^{sr}_{n\times 2}$, the result follows immediately from Theorem 4.4.2 and Theorem 4.4.8. □

Remark: We should mention that there has been some other works on networks with two sources in the literature [WS10, EF09]. In fact, Corollary 4.4.9 has also been recognized by Wang et al. through a graph-theoretic approach [WS10].

4.4.2 Some Explicit Computations

Now we give the results of the explicit computation of the rank region of $n \times 2$ matrices for three values of n,

1. 4×2 matrix: Computing all the ranks of a 4×2 matrix results in 52 vectors. The convex hull of these vectors is represented by 156 inequalities and interestingly all the 52 points are corner points of the convex hull. Out of 156 inequalities, 50 of them are homogenous, meaning that they define hyperplanes that pass through the origin. However as opposed to the 4×4 case there are only 26 rays for the convex cone. The only missing vector compared to the 4×4 case is, (1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3). This is somehow not surprising, since this is the only ray in Table 4.1 that has rank entries greater than 2.

2. 5×2 matrix: There are a total of 203 vectors, out of which 112 are rays of the convex cone region.

3. 6×2 matrix: There are 877 rank vectors for a 6×2 matrix which are obviously 63 dimensional. Computing the convex cone of these vectors by means of a linear program, gave 575 rays.

4.5 Binary Scalar Linear Network Coding

In the previous section, we saw how constraining the number of sources to 2 could make the region of linear representable entropies more manageable. Another approach for making the problem of determining the linear solutions of a network more tractable, is to focus on the linear solutions over some fixed alphabet size. This is equivalent to determining the region of linear representable vectors over the given alphabet-size, and thus it is what we will study in this section. We examine the problem from the matroids perspective.

4.5.1 Entropy, Polymatroid, and Matroid

Recall that entropy satisfies the submodularity conditions. Indeed entropy is a polymatroid.

Definition 4.5.1 (Polymatroid) *A finite set E called the ground set along with a function r called the rank function, which maps the elements of the power set of E to non-negative real numbers, i.e., $r : 2^E \to \mathbb{R}^+$ is called a polymatroid if and only if for all $A, B \subseteq E$, r satisfies the following:*

1. *$r(\emptyset) = 0$.*
2. *If $A \subseteq B$ then $r(A) \leq r(B)$.*
3. *$r(A \cup B) + r(A \cap B) \leq r(A) + r(B)$: submodularity.*

where we have assumed that the entries of r are indexed by subsets of E.

A relevant though different with polymatroid concept is the *matroid* definition.

Definition 4.5.2 (Matroid) *A finite set E called the ground set along with a function r called the rank function, that maps the elements of the power set of E to non-negative integers, i.e., $r : 2^E \to \mathbb{Z}^+$ is called a matroid if and only if for all $A, B \subseteq E$, r satisfies the following:*

1. *$0 \leq r(A) \leq |A|$.*
2. *If $A \subseteq B$ then $r(A) \leq r(B)$.*
3. *$r(A \cup B) + r(A \cap B) \leq r(A) + r(B)$: submodularity.*

Therefore the main difference between polymatroids and matroids is the integer requirement for the entries of the rank functions. The condition $r(A) \leq |A|$ in the

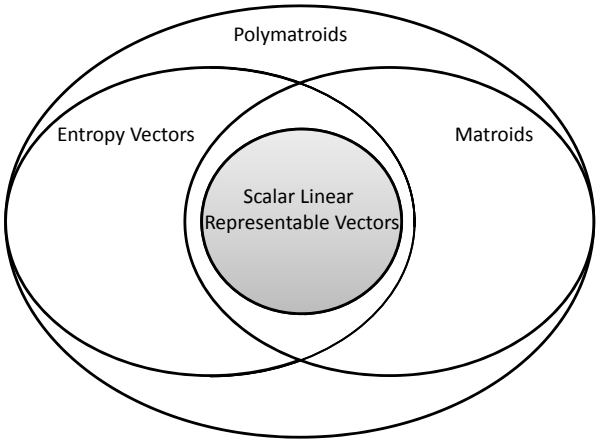


Figure 4.2: Polymatroids, matroids, entropy, and the linear representable vectors

matroid is not too critical, as one can scale the polymatroids. Therefore it is easy to see that any matroid is a polymatroid, however the converse is not true. Moreover, as mentioned, any entropy vector is also a polymatroid, whereas not all polymatroids are entropic. This is simply due to the fact that Γ_n^* is a strict subset of the polytope defined by Shannon inequalities. If we denote a matroid whose ground set elements can be considered as random variables $\{X_1, \ldots, X_n\}$ with entropy as its rank function by entropic, then comparing matroid rank functions and entropy vectors also reveals that neither all entropy vectors are matroidal nor all matroid rank functions are entropic. Not all entropy vectors are matroidal simply because there is no reason why the entries of the entropy vector should be integers, and not all matroids are entropic because there are some matroid rank vectors which are proven to violate some non-Shannon type inequalities and therefore are not entropic [DFZ07]. An important class of matroids are the "representable" matroids.

Definition 4.5.3 (Representable matroids) *A matroid is called representable if its rank function is a linear representable vector.*

Clearly due to Theorem 4.2.3 all representable matroids are entropic. Moreover one can easily see that all scalar representable vectors are also matroids. Therefore, in order to study the scalar linear representable region one can as well study the region of linear representable matroids and use the available results in the matroid theory [Oxl06]. Figure 4.2 shows all these connections.

4.5.2 Matroid Representability and Excluded Minors

The question of whether a matroid is representable or not has long been an open problem. This problem is solved only if one is interested to know whether a matroid is representable over $GF(2), GF(3)$, or $GF(4)$. Before stating those results however, we need an alternative definition of the matroid.

Definition 4.5.4 (Matroid in terms of independent sets) *Let E be a finite set called the ground set and $\mathcal{I}$ be a set of subsets of E. Then the ordered pair $(E, \mathcal{I})$ is a matroid:*

1. *If $\emptyset \in \mathcal{I}$.*
2. *If $A \subseteq B$ and $B \in \mathcal{I}$ then $A \in \mathcal{I}$.*
3. *If $A, B \subseteq \mathcal{I}$ and $|A| \leq |B|$, then there exists an element e of $B - A$ such that $A \cup e \in \mathcal{I}$.*

The elements of the set $\mathcal{I}$ are called independent sets. Moreover any subset of E that is not in $\mathcal{I}$ is called a dependent set.

It is easy to see that the Definitions 4.5.2 and 4.5.4 are equivalent. In fact if the matroid rank vector is given, the independent sets of the matroid will be those subsets $A \subseteq E$ for which $r(A) = |A|$. Conversely if the independent sets of the matroid are

given, then the rank of every subset A is the size of the largest independent set within A.

There are two ways one can remove elements from a matroid such that the resulting smaller object is still a matroid. These two operations are namely the "deletion" and "contraction" of a matroid [Oxl04].

Definition 4.5.5 (Deletion) *Let $(E, \mathcal{I})$ be a matroid and $e \in E$ be an element of the ground set. Deletion of e from the matroid is the new matroid $(E', \mathcal{I}')$ with ground set $E' = E - \{e\}$ and the independent set $\mathcal{I}' = \{I | I \in \mathcal{I},\ e \notin I\}$*

Definition 4.5.6 (Contraction) *Let $(E, \mathcal{I})$ be a matroid and $e \in E$ be an element of the ground set. The contraction of e from the matroid is the new matroid $(E', \mathcal{I}')$ with ground set $E' = E - \{e\}$ and the independent set $\mathcal{I}' = \{I - \{e\} | I \in \mathcal{I},\ e \in I\}$.*

From Definitions 4.5.5 and 4.5.6, one can easily see that deletion and contraction can be considered as dual of each other: while the independent sets of the deletion matroid are those independent sets of the original matroid that do not contain e, the independent sets of contraction matroid are obtained by removing e from those independent sets of the original matroid that did contain e. Note that while contraction always reduces the rank of a matroid, deletion does not necessarily do so. Furthermore we would like to mention that deletion and contraction commute within themselves and with each other [Oxl04], therefore we do not need to specify order when performing these operations. It turns out that deletion and contraction translate to nice operations over the matroid rank function.

- Deletion corresponds to marginalization: Since in deletion a set of elements is deleted from the ground set and the independent sets of the rest of the matroid are kept as before, deletion of a set T from the matroid is just equivalent to deleting those rank entries of the rank vector that contained any elements of T. In other words, deletion is equivalent to marginalizing the rank vector.

- Contraction corresponds to conditioning: It can be shown [Oxl06] that contracting the set T from the matroid results in the new rank function $r'(A) = r(A \cup T) - r(T)$ for all $A \subseteq E - T$. Note that if r is an entropy vector then this is just equivalent to $r'(A) = r(A|T)$.

Performing a series of deletions and contractions over a matroid results in a smaller matroid known as "minor". These objects were first introduced by Tutte in 1958.

Definition 4.5.7 (Minor) *Let $(E, \mathcal{I})$ be a matroid and assume that X, Y are two disjoint subsets of E such that either of them could be empty. The matroid obtained by deleting X and contracting Y is called a minor.*

As previously mentioned, the problem of linear representability of matroids is only solved over GF(2), GF(3), and GF(4). It turns out that in all these cases the minors of the matroid play an important role.

To explain representability results over these fields first consider GF(2). A uniform matroid denoted by $U_{m,n}$ is defined as the matroid with n elements in its ground set such that its independent sets are all subsets of size at most m. In other words the rank vector of uniform matroid is such that for all subsets A where $|A| \leq m$ we have $r(A) = |A|$ and for $|A| > m,\ r(A) = m$. For the case of $U_{2,4}$ that would mean that we have 4 elements where every two of them are independent. Since there are only 3 pairwise independent vectors over GF(2) namely $[1\ 0], [0\ 1], [1\ 1]$, one can immediately see that $U_{2,4}$ is not representable over GF(2), although it has representation over any other field. As it turns out having a U_{24} as a minor is the only reason for non-representability of a matroid over GF(2).

Theorem 4.5.8 (Tutte 1958) *A matroid is binary representable if and only if it does not have any U_{24} minors.*

For representability over GF(3) and GF(4) there are similar results in terms of the minors of the matroid.

Theorem 4.5.9 (Reid 1971; Bixby 1979; Seymour 1979) *A matroid is ternary representable if and only if it has no minors isomorphic to either of the* $U_{25}, U_{3,5}, F_7$, *or* F_7^*.

Theorem 4.5.10 (Geelen, Gerards, Kapoor 2000) *A matroid is quaternary if and only if it does not have any minors isomorphic to either of the* U_{26}, U_{46}, F_7^-, $(F_7^-)^*$, P_6, P_8, *or* P_8''.

In the above theorems, F_7 and F_7^* denote the Fano matroid and its dual, F_7^- and $(F_7^-)^*$ are the non-Fano and its dual, and P_8'' is a special relaxation of P_8 matroid. However we shall not further delve into these here.

The set of minors which forbid a matroid to be representable over a certain field GF(q) are called the *excluded minors* for GF(q) representability. The representability problem over any other field than those stated above is pretty much open. As an instance it is not even known whether the set of excluded minors for GF(q) representability $q \geq 5$ is finite or not. Although Rota in 1970 conjectured that they are finite. There are many more interesting questions related to representability of matroids over finite fields and one may consult [Oxl06] or [Oxl04] for more details.

4.5.3 Binary Capacity of Networks and Binary Representability

We will now focus on binary representable matroids to study scalar linear codes over GF(2) in networks.

Theorem 4.5.11 (Binary representable vectors) *A* $2^n - 1$ *dimensional vector* h *is scalar binary representable if and only if:*

1. h *is a rank function of a matroid on* n *elements as stated in Definition 4.5.2.*

2. *For any four elements* $i, j, k, l \in \{1, \ldots, n\}$ *and any subset* $T \subseteq \{1, \ldots, n\} - \{i, j, k, l\}$ *the 15-dimensional vector* h' *whose entries are defined as* $h'(A) = h(A \cup T) - h(T), \forall A \subseteq \{i, j, k, l\}$ *is not* U_{24}.

Proof: The first condition can be seen easily from the fact that a binary representable vector is a binary matroid and vice versa. Moreover second statement is equivalent to obtaining 4 element minors of the matroid on n elements by contracting the set $T \subseteq \{1, \ldots, n\} - \{i, j, k, l\}$ and deleting $S = \{1, \ldots, n\} - \{i, j, k, l\} - T$ and enforcing those minors not to be U_{24}, which based on Theorem 4.5.8 is necessary and sufficient for binary representability. □

Recall from Theorem 4.2.4 that to obtain the best scalar linear code in a network, we need to perform the optimization over the convex hull of the linear representable vectors. Therefore we need the following theorem:

Theorem 4.5.12 (Convex hull of binary entropic vectors) *A* $2^n - 1$ *dimensional vector* h *is in the convex hull of entropy region of* n *binary linearly-related random variables if and only if:*

1. $h \in \Gamma_n^{mat}$ *where* Γ_n^{mat} *is the convex hull of matroid rank functions over* n *elements.*

2. *For any four elements* $i, j, k, l \in \{1, \ldots, n\}$ *and any subset* $T \subseteq \{1, \ldots, n\} - \{i, j, k, l\}$ *the 15-dimensional vector* h' *whose entries are defined as* $h'(A) = h(A \cup T) - h(T), \forall A \subseteq \{i, j, k, l\}$ *is in the convex hull of the entropy region of 4 binary linearly-related random variables.*

Proof: Follows readily from Theorem 4.5.11. □

Note that, as stated earlier, due to the integer constraint on the matroid rank elements, the convex hull/cone of matroids is a subset of the convex hull/cone of polymatroids. In particular for 4 elements, the convex cone of polymatroid cone has

41 rays and the convex cone of matroids has 35 rays (these 35 rays are the same as the ones given by Table 3.7, and Table 3.2 together).

We can now state the optimization formulation for determining the capacity of networks over binary operations.

Theorem 4.5.13 (Binary capacity of networks) *Let $X_1, \ldots, X_n$ denote the random variables of a acyclic memoryless wired network. The maximum weighted sum rate achieved by scalar linearly-related binary random variables and binary operations can be obtained from the following optimization problem:*

$$\max \alpha^T h \tag{4.11}$$

subject to $h \in con(\Gamma_n^{mat})$ and,

1. $h_{X_e} \leq C_e \qquad$ *for any edge e*
2. $h(X_{S_1}, \ldots, X_{S_s}) = \sum_{i=1}^{s} h(X_{S_i})$
3. $h(X_{in}^i, X_{out}^i(j)) = h(X_{in}^i) \quad j = 1, \ldots, X_{out}^i(j)$ *for each nonsource node i*
4. $h(X_{in}^l, X_d^l(j)) = h(X_d^l(j))$ *l is a sink node*
5. *For any four elements $i, j, k, l \in \{1, \ldots, n\}$ and any subset $T \subseteq \{1, \ldots, n\} - \{i, j, k, l\}$ the 15-dimensional vector h' whose entries are defined as $h'(A) = h(A \cup T) - h(T), \forall A \subseteq \{i, j, k, l\}$ is in the convex hull of the entropy region of 4 binary linearly-related random variables.*

Remark: Note that although Theorem 4.5.13 gives a linear programming approach for optimizing the achievable rates in a network, its complexity could be exponential in n the number of random variables of the network. In fact the last condition of the Theorem 4.5.13 requires one to check all the 4 element minors. The number of these minors alone is $\binom{n}{4}2^{n-4}$. Moreover as we know that the number of inequalities that

define the polymatroid cone is $n + \binom{n}{2}2^{n-2}$, it is likely that the number of inequalities that are required for characterizing the convex hull/cone of matroids is exponential as well. However, some assumptions about the network may reduce the number of conditions significantly. As an instance if know that there are r sources in the network then all representable vectors corresponding to network solutions will be from rank r matroids. The number of minors that need to be checked in a rank r matroid for binary representability is simply $\binom{n}{4}\binom{n-4}{r-2} = O(n^{r+2})$. This is due to the fact that $r-2$ contractions are needed to make the obtained minor rank 2 for the comparison with the U_{24}.

4.6 Vector Linear Codes

Thus far we only considered the scalar linear codes and hence the scalar linear representable region. Note that a multilinear representable vector can still be seen as a bigger scalar representable vector where we would only be interested in a subset of its entries. Compared to scalar linear representability, vector linear representability has not been studied as much. It turns out that there are matroids which are not scalar representable but have a vector representation. $U_{2,4}$ for instance has a vector representation over GF(2). Here is one representation for it,

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \end{bmatrix}. \tag{4.12}$$

Another example of a matroid that does not have scalar linear representation is the non-Pappus matroid, which turns out to be vector representable over GF(3) [SA98, Mat99].

We will talk about numerical methods for obtaining scalar and vector representations of matroids in Chapter 6. In that chapter we will come back to the U_{24} and non-Pappus matroids.

4.7 Conclusions

In this chapter we studied the linear representable entropy region along with its connections to the linear networks codes. In particular for the most part we focused on the scalar representable entropy region (entropy vectors with scalar underlying random variables). First we presented a method for obtaining the linear representable entropy region and we explicitly computed it for 4 random variables. Next we turned our attention to networks with two sources and particularly showed the optimality of linear codes among all scalar codes for such networks. Finally we studied the binary capacity of networks, and by appealing to the subject of excluded minors of a matroid, gave necessary and sufficient conditions for a vector to be binary representable.

Chapter 5

Gaussian Entropy Region

5.1 Introduction

The effort to characterize the entropy region has mostly focused on discrete random variables, ostensibly because the study of discrete random variables is simpler. However, continuous random variables are as important, where now for any collection of random variables X_α, with joint probability density function $f_{X_\alpha}(x_\alpha)$, the differential entropy is defined as

$$h_\alpha = -\int f_{X_\alpha}(x_\alpha) \log f_{X_\alpha}(x_\alpha) dx_\alpha. \tag{5.1}$$

Let $\sum_\alpha a_\alpha H_\alpha \geq 0$ be a valid discrete information inequality. This inequality is called *balanced* if for all $i \in \mathcal{N}$ we have $\sum_{\alpha: i \in \alpha} a_\alpha = 0$. Using this notion Chan [Cha03] has shown a correspondence between discrete and continuous information inequalities, which allows us to compute the entropy region for one from the other.

Theorem 5.1.1 (Discrete/continuous information inequalities)

1. *A linear continuous information inequality* $\sum_\alpha a_\alpha h_\alpha \geq 0$ *is valid if and only if its discrete counterpart* $\sum_\alpha a_\alpha H_\alpha \geq 0$ *is balanced and valid.*

2. *A linear discrete information inequality* $\sum_\alpha a_\alpha H_\alpha \geq 0$ *is valid if and only if it can be written as* $\sum_\alpha \beta_\alpha h_\alpha + \sum_{i=1}^n r_i (h_{i,i^c} - h_{i^c})$ *for some* $r_i \geq 0$*, where*

$\sum_\alpha \beta_\alpha h_\alpha \geq 0$ *is a valid continuous information inequality (i^c denotes the complement of i in $\mathcal{N}$).*

The above theorem suggests that one can also study continuous random variables to determine Γ_n^*. Among all continuous random variables, the most natural ones to study first (for many of the reasons further described below) are Gaussians. This will be the main focus of this chapter.

Let $X_1, \cdots, X_n \in \mathcal{R}^T$ be n jointly distributed zero-mean[1] vector valued Gaussian random variables with covariance matrix $R \in \mathcal{R}^{nT \times nT}$. Clearly, R is symmetric, positive semidefinite, and consists of block matrices of size $T \times T$ (corresponding to each random variable). We will allow T to be arbitrary and will therefore consider the *normalized* joint entropy of any subset $\alpha \subseteq \mathcal{N}$ of these random variables

$$\underline{h}_\alpha = \frac{1}{T} \cdot \frac{1}{2} \log \Big((2\pi e)^{T|\alpha|} \det R_\alpha \Big), \tag{5.2}$$

where $|\alpha|$ denotes the cardinality of the set α and R_α is the $|\alpha|T \times |\alpha|T$ matrix obtained by keeping those block rows and block columns of R that are indexed by α. Note that our normalization is by the dimensionality of the X_i, i.e., by T, and that we have used $\underline{h}$ to denote normalized entropy.

Normalization has the following important consequence:

Theorem 5.1.2 (Convexity of the region for $\underline{h}$) *The closure of the region of normalized Gaussian entropy vectors is convex.*

Proof: Let $\underline{h}^x$ and $\underline{h}^y$ be two normalized Gaussian entropy vectors. This means that the first corresponds to some collection of Gaussian random variables $X_1, \ldots, X_n \in \mathcal{R}^{T_x}$ with the covariance matrix R^x, for some T_x, and the second to some other collection $Y_1, \ldots, Y_n \in \mathcal{R}^{T_y}$ with the covariance matrix R^y, for some T_y. Now generate

[1] Since differential entropy is invariant to shifts there is no point in assuming nonzero means for the X_i.

N_x copies of jointly Gaussian random variables $X_1, \ldots, X_n$ and N_y copies of $Y_1, \ldots, Y_n$ and define the new set of random variables $Z_i = [(X_i^1)^t, \ldots, (X_i^{N_x})^t, (Y_i^1)^t, \ldots, (Y_i^{N_y})^t]^t$ where $(.)^t$ denotes the transpose, by stacking N_x and N_y independent copies of each, respectively, into a $N_x T_x + N_y T_y$ dimensional vector. Clearly the Z_i are jointly-Gaussian. Due to the independencies of the X_i^k and Y_i^l, $k = 1, \ldots N_x$, $l = 1, \ldots, N_y$, the non-normalized entropy of the collection of random variables Z_α is

$$h_\alpha^z = N_x T_x \underline{h}_\alpha^x + N_y T_y \underline{h}_\alpha^y.$$

To obtain the normalized entropy we should divide by $N_x T_x + N_y T_y$

$$\underline{h}_\alpha^z = \frac{N_x T_x}{N_x T_x + N_y T_y} \underline{h}_\alpha^x + \frac{N_y T_y}{N_x T_x + N_y T_y} \underline{h}_\alpha^y,$$

which, since N_x and N_y are arbitrary, implies that every vector that is a convex combination of $\underline{h}^x$ and $\underline{h}^y$ is entropic and generated by a Gaussian. □

Note that $\underline{h}_\alpha$ can also be written as follows:

$$\underline{h}_\alpha = \frac{1}{2T} \log \det R_\alpha + \frac{|\alpha|}{2} \log 2\pi e. \tag{5.3}$$

Therefore if we define

$$g_\alpha = \frac{1}{T} \log \det R_\alpha, \tag{5.4}$$

it is obvious that g_α can be obtained from $\underline{h}_\alpha$ and vice versa. All that is involved is a scaling of the covariance matrix R. For balanced inequalities there is the additional property,

Lemma 5.1.3 *If $\sum_\alpha a_\alpha H_\alpha$ is balanced then $\sum_\alpha |\alpha| a_\alpha = 0$.*

Proof: We can simply write,

$$\sum_{\alpha} |\alpha| a_\alpha = \sum_{\alpha} \sum_{i \in \alpha} a_\alpha = \sum_{i} \left(\sum_{\alpha : i \in \alpha} a_\alpha \right) = 0. \tag{5.5}$$

□

Therefore since inequalities for continuous entropies are balanced, any inequality satisfied by h is also satisfied by g and vice versa. As a result the space for g and $\underline{h}$ are the same. For simplicity, we will therefore use g_α instead of $\underline{h}_\alpha$ throughout the chapter and use the term entropy for both g and $\underline{h}$ interchangeably.

In this chapter we characterize the entropy region of 3 jointly Gaussian random variables and study the minimal set of necessary and sufficient conditions for a 15-dimensional vector to represent an entropy vector of 4 jointly Gaussian random variables. As equation (5.4) suggests, entropy of any subset of random variables from a collection of Gaussian random variables is simply the "log" of the principal minor of the covariance matrix corresponding to this subset. Therefore studying entropy of Gaussian random variables involves studying the relations among principal minors of symmetric positive semi-definite matrices, i.e., the covariance matrices. It has recently been noted that one of these relations is the so-called Cayley's "hyperdeterminant" [HS07b]. Therefore along the study of entropy of 4 Gaussian random variables we also examine the hyperdeterminant relation.

The remainder of this chapter is organized as follows: In the next section we review background and some motivating results on the entropies of Gaussian random variables. Section 5.3 states the main results on the characterization of the entropy region of 3 jointly Gaussian random variables. In Section 5.4 we examine the hyperdeterminant relation in connection to the entropy region of Gaussian random variables. We give a determinant formula for calculating the special $2 \times 2 \times 2$ hyperdeterminant. Moreover we present new and transparent proof of the result of [HS07b] on why the

principal minors of a symmetric matrix satisfy the hyperdeterminant relation. In Section 5.5 we study the minimal set of necessary and sufficient condition for a $2^n - 1$ dimensional vector to be the entropy vector of n scalar jointly Gaussian random variables. For $n = 4$ there are 5 such equations and we explicitly state them. Finally we turn our attention toward the entropy region in wireless networks where the entropy power inequality plays an important role.

5.2 Some Known Results

From (5.4) it can be easily seen that any valid information inequality for entropies can be immediately converted into an inequality for the (block) principal minors of a symmetric, positive semi-definite matrix. This connection has been previously used in the literature. In fact one can study determinant inequalities by studying the corresponding entropy inequalities, see, e.g., [CT88].

Let g be the entropy vector corresponding to some vector-valued collection of random variables with an $nT \times nT$ covariance matrix R. Further, let m denote the vector of block principal minors of R. Then it is clear that $m = e^{gT}$, where the exponential acts component-wise on the entries of g. Then the submodularity of entropy translates to the following inequality for the principal minors:

$$m_{\alpha\cup\beta} \cdot m_{\alpha\cap\beta} \leq m_\alpha \cdot m_\beta. \tag{5.6}$$

In the context of determinant inequalities for a Hermitian positive semidefinite matrix, this is known as the "Koteljanskii" inequality and is a generalization of the "Hadamard-Fischer" inequalities [FJ00]. Dating back at least to Hadamard in 1893, studying the determinant inequalities is an old subject which is of interest on its own and has many applications in matrix analysis and probability theory.

Some of the interesting problems in the area of principal minor relations include

characterizing the set of bounded ratios of principal minors for a given class of matrices (e.g., the class of positive definite or the class of matrices for which all of their principal minors are positive, i.e., the P matrices) [JB93, HJ], studying the Gaussian conditional independence structure in the context of probabilistic representation [LM07] and detecting P matrices, e.g., via computation of all principal minors of a given matrix [GT06a].

Although determinant inequalities have been studied extensively on their own, and also through the entropy inequalities, the reverse approach of determining Gaussian entropies via exploration of the space of principal minors has been less considered [LM07, Lne03]. As it turns out, this approach is deeply related to the "principal minor assignment" problem, where a matrix with a set of fixed principal minors is desired. Recently there has been progress towards this area for symmetric matrices [HS07b, GT06b] and we will discuss this in more detail in Sections 5.4 and 5.5.

Apart from the result of [Lne03] which shows the tightness of the Zhang-Yeung non-Shannon inequality [ZY98] for Gaussian random variables, one of the encouraging results for studying the Gaussian random variables is that they can violate the "Ingleton bound". This bound is one of the best known inner bounds for Γ_4^* [ZY98].

Theorem 5.2.1 (Ingleton inequality) *[Ing71] Let $v_1, \cdots, v_n$ be n vector subspaces and let $\mathcal{N} = \{1, \cdots, n\}$. Further let $\alpha \subseteq \mathcal{N}$ and r_α be the rank function defined as the dimension of the subspace $\oplus_{i\in\alpha} v_i$. Then for any subsets $\alpha_1, \alpha_2, \alpha_3, \alpha_4 \subseteq \mathcal{N}$, we have*

$$\begin{aligned} &r_{\alpha_1} + r_{\alpha_2} + r_{\alpha_1\cup\alpha_2\cup\alpha_3} + r_{\alpha_1\cup\alpha_2\cup\alpha_4} + r_{\alpha_3\cup\alpha_4} \\ &-r_{\alpha_1\cup\alpha_2} - r_{\alpha_1\cup\alpha_3} - r_{\alpha_1\cup\alpha_4} - r_{\alpha_2\cup\alpha_3} - r_{\alpha_2\cup\alpha_4} \leq 0. \end{aligned} \tag{5.7}$$

The Ingleton inequality was first obtained for the rank of vector spaces. However it turns out that certain types of entropy functions, in particular all linear representable (corresponding to linear codes over finite fields) and pseudo-abelian group

characterizable entropy functions also satisfy this inequality and hence fall into this inner bound [Cha07b, Cha07a]. However if we consider 4 jointly Gaussian random variables, we find, interestingly, that they can violate the Ingleton bound. Consider the following covariance matrix:

$$\begin{bmatrix} 1 & \varepsilon & a & a \\ \varepsilon & 1 & a & a \\ a & a & 1 & 0 \\ a & a & 0 & 1 \end{bmatrix}. \tag{5.8}$$

To violate the Ingleton inequality we need to have:

$$\begin{aligned} & g_1 + g_2 + g_{123} + g_{124} + g_{34} \\ & -g_{12} - g_{13} - g_{14} - g_{23} - g_{24} \geq 0. \end{aligned} \tag{5.9}$$

Substituting for values of g and simplifying we obtain:

$$\frac{1-\varepsilon}{1+\varepsilon} \geq \left(\frac{1-2a^2+a^4}{1-2a^2+\varepsilon} \right)^2. \tag{5.10}$$

Moreover imposing positivity conditions for this matrix to correspond to a true covariance matrix gives $0 \leq a^2 \leq 0.5$, $4a^2 - 1 \leq \varepsilon \leq 1$. Solving inequality (5.10) subject to these constraints yields a region of permissible ε and a^2 (Fig 5.1). In particular the point $\varepsilon = 0.25 \;\; a = 0.5$ lies in this region. Interestingly enough, this example has also been discovered in the context of determinantal inequalities in [JB93]. Taking these results into account, we will study the Gaussian entropy region for 2,3 random variables and give the minimal number of necessary and sufficient conditions for a $2^n - 1$ dimensional vector to correspond to the entropy of n scalar jointly Gaussian random variables in the next sections.

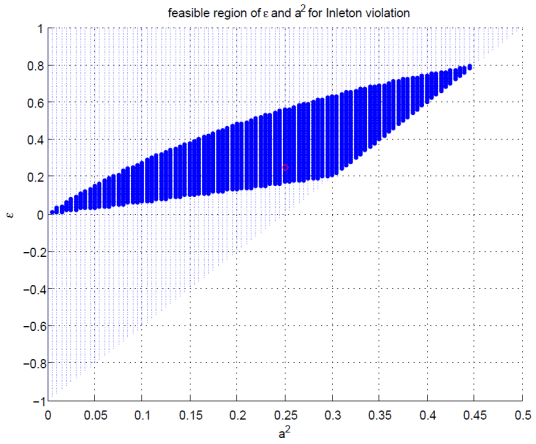


Figure 5.1: Feasible region of ε and a^2 for the specific Ingleton violating example

5.3 Entropy Region of 2 and 3 Gaussian Random Variables

The above results (violation of the Ingleton bound and tightness of the non-Shannon inequality) lead one to speculate whether the entropy region for arbitrary continuous random variables can be obtained from the entropy region of (vector-valued) Gaussian ones. Although this is the case for $n = 2$ random variables, unfortunately, it is not true in general (not for even $n = 3$).

5.3.1 $n = 2$

Entropy region of 2 jointly Gaussian random variables is trivially equal to the whole region of 2 arbitrary distributed continuous random variables.

Theorem 5.3.1 *Entropy region of 2 jointly Gaussian random variables is described by the single inequality $g_{12} \leq g_1 + g_2$ and is equal to the entropy region of 2 arbitrary distributed continuous random variables.*

Proof: Since it is known that the continuous entropy region is described by the single balanced inequality $h_{12} \leq h_1 + h_2$, to prove the theorem it is sufficient to show that

any entropy vector $[h_1, h_2, h_{12}]$ satisfying this inequality may be described by 2 jointly Gaussians and this is trivial to show. □

5.3.2 Main Results for $n = 3$

Although we consider vector-valued jointly Gaussian random variables, for $n = 3$, we find that considering the convex hull of scalar jointly Gaussian random variables is sufficient for characterizing the Gaussian entropy region,

Theorem 5.3.2 *The entropy region of 3 vector-valued Gaussian random variables can be obtained from the convex hull of scalar Gaussian random variables.*

The characterization of the entropy region of 3 jointly Gaussian random variables is formalized in the next theorem, which shows that in general the Gaussian entropy region is a strict subset of the entropy region of arbitrary distributed continuous random variables.

Theorem 5.3.3 (Entropy Region for $n = 3$ Gaussian RVs) *Let the 7-dimensional vector $g = [g_1, g_2, g_3, g_{12}, g_{23}, g_{31}, g_{123}]^t$ be an entropy vector generated by 3 Gaussian random variables. Define $x_k = e^{g_{ij} - g_i - g_j}$ and $\tilde{y} = \frac{\prod_k x_k}{\max_k x_k} + 2\max_k x_k - \sum_k x_k$. The closure of the Gaussian entropy region generated by such g vectors is characterized by,*

1. *For $\tilde{y} \leq 0$:*

$$g_{ij} \leq g_i + g_j \quad , \quad g_{123} \leq \min_j (g_{ij} + g_{jk} - g_j). \tag{5.11}$$

2. *For $\tilde{y} > 0$:*

$$g_{ij} \leq g_i + g_j \quad , \quad g_{123} \leq \sum_k g_k + \log\left(\max\left[0, -2 + \sum_k x_k + 2\sqrt{\prod_k (1 - x_k)}\right]\right) \tag{5.12}$$

The entropy region for three random variables is simply given by the above inequalities. Thus, when $\tilde{y} \leq 0$, the Gaussian entropy region coincides with the continuous entropy region; however, when $\tilde{y} > 0$ (and this can happen for some valid entropy vectors), we have the tighter upper bound (5.12) on g_{123}.

Theorem 5.3.3 implies that the Gaussian entropy region for $n = 3$ vector-valued random variables is *strictly* smaller than the actual entropy region.

Nonetheless, not all hope is lost and the next theorem shows that one can indeed construct the entropy region for $n = 3$ random variables from the entropy region generated by vector-valued Gaussians.

Theorem 5.3.4 (General and Gaussian Entropy Regions) *Let $g \in \mathcal{R}^7$ be a continuous entropy vector. Then there exists a $\theta^* > 0$, such that for all $\theta \geq \theta^*$, the vector $\frac{1}{\theta}g$ can be generated by three vector-valued jointly Gaussian random variables.*

In other words, the entropy region for $n = 3$ continuous random variables is the (convex) cone generated by the entropy region of 3 Gaussian random variables. This result gives us hope that the study of Gaussians may be fruitful for $n \geq 4$.

5.3.3 Proof of Main Results for $n = 3$

In what follows we will outline the proofs of Theorems 5.3.3 and 5.3.4. The basic idea is to determine the structure of the Gaussian random variables that generate the *boundary* of the entropy region for Gaussians, and then to determine what the boundary entropies are. We need a few lemmas:

Lemma 5.3.5 (Boundary of the Gaussian Entropy Region) *The boundary of the Gaussian entropy region is generated by the concatenation of a set of vector valued*

Gaussian random variables with covariance

$$\begin{bmatrix} \alpha_{11} I_{\hat{T}} & \alpha_{12}\Phi_{12} & \alpha_{13}\Phi_{13} \\ \alpha_{12}\Phi_{12}^t & \alpha_{22} I_{\hat{T}} & \alpha_{23}\Phi_{23} \\ \alpha_{13}\Phi_{13}^t & \alpha_{23}\Phi_{23}^t & \alpha_{33} I_{\hat{T}} \end{bmatrix}, \tag{5.13}$$

where the Φ_{ij} *are* orthogonal *matrices, and another set of independent vector-valued Gaussian random vectors with covariance*

$$\begin{bmatrix} \alpha_{11} I_{T-\hat{T}} & 0 & 0 \\ 0 & \alpha_{22} I_{T-\hat{T}} & 0 \\ 0 & 0 & \alpha_{33} I_{T-\hat{T}} \end{bmatrix}. \tag{5.14}$$

Proof: To find the boundary region for 3 jointly Gaussian random variables, we should solve the following maximization problem:

$$\max_{h} \sum_{s\subseteq\{1,2,3\}} \gamma_s h_s \tag{5.15}$$

or equivalently,

$$\max_{\tilde{R}} \sum_{s\subseteq\{1,2,3\}} \gamma_s \log\det \tilde{R}_s, \tag{5.16}$$

where $\tilde{R}$ is the $3T \times 3T$ block covariance matrix which for the moment we assume that all its principal minors are nonzero. This optimization comes about when we fix any 6 of the entropies and try to maximize the last one.

KKT conditions necessitate that the derivative of (5.16) with respect to $\tilde{R}$ be zero, i.e., $\frac{\partial}{\partial \tilde{R}} \left(\sum_{s\subseteq\{1,2,3\}} \gamma_s \log\det \tilde{R}_s \right) = 0$. To compute the derivatives we note that for a symmetric matrix X, we have $\frac{\partial}{\partial X} \log\det X = 2X^{-1} - \mathrm{diag}(X^{-1})$, where diag denotes

the diagonal elements. If we adopt the following notation,

$$\begin{pmatrix} \tilde{S}_{11} & \tilde{S}_{12} \\ \tilde{S}_{21} & \tilde{S}_{22} \end{pmatrix} = \begin{pmatrix} \tilde{R}_{11} & \tilde{R}_{12} \\ \tilde{R}_{21} & \tilde{R}_{22} \end{pmatrix}^{-1}, \begin{pmatrix} \tilde{T}_{11} & \tilde{T}_{13} \\ \tilde{T}_{31} & \tilde{T}_{33} \end{pmatrix} = \begin{pmatrix} \tilde{R}_{11} & \tilde{R}_{13} \\ \tilde{R}_{31} & \tilde{R}_{33} \end{pmatrix}^{-1}$$
$$\begin{pmatrix} \tilde{U}_{22} & \tilde{U}_{23} \\ \tilde{U}_{32} & \tilde{U}_{33} \end{pmatrix} = \begin{pmatrix} \tilde{R}_{22} & \tilde{R}_{23} \\ \tilde{R}_{32} & \tilde{R}_{33} \end{pmatrix}^{-1}, \ \tilde{V}_{ij} = (\tilde{R}^{-1})_{ij}, \tag{5.17}$$

then we obtain,

$$\begin{aligned} &\gamma_1 \begin{pmatrix} \tilde{R}_{11}^{-1} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \gamma_2 \begin{pmatrix} 0 & 0 & 0 \\ 0 & \tilde{R}_{22}^{-1} & 0 \\ 0 & 0 & 0 \end{pmatrix} + \gamma_3 \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \tilde{R}_{33}^{-1} \end{pmatrix} \\ &+2\gamma_{12} \begin{pmatrix} \tilde{S}_{11} & \tilde{S}_{12} & 0 \\ \tilde{S}_{21} & \tilde{S}_{22} & 0 \\ 0 & 0 & 0 \end{pmatrix} - \gamma_{12} \begin{pmatrix} \mathrm{diag}(\tilde{S}_{11}) & 0 & 0 \\ 0 & \mathrm{diag}(\tilde{S}_{22}) & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ &+2\gamma_{13} \begin{pmatrix} \tilde{T}_{11} & 0 & \tilde{T}_{13} \\ 0 & 0 & 0 \\ \tilde{T}_{31} & 0 & \tilde{T}_{33} \end{pmatrix} - \gamma_{13} \begin{pmatrix} \mathrm{diag}(\tilde{T}_{11}) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \mathrm{diag}(\tilde{T}_{33}) \end{pmatrix} \\ &+2\gamma_{23} \begin{pmatrix} 0 & 0 & 0 \\ 0 & \tilde{U}_{22} & \tilde{U}_{23} \\ 0 & \tilde{U}_{32} & \tilde{U}_{33} \end{pmatrix} - \gamma_{23} \begin{pmatrix} 0 & 0 & 0 \\ 0 & \mathrm{diag}(\tilde{U}_{22}) & 0 \\ 0 & 0 & \mathrm{diag}(\tilde{U}_{33}) \end{pmatrix} \\ &+2\gamma_{123}\, \tilde{R}^{-1} - \gamma_{123}\, \mathrm{diag}(\tilde{R}^{-1}) = 0. \end{aligned} \tag{5.18}$$

Multiplying (5.18) by $\tilde{R}$ from right we obtain,

$$\begin{pmatrix} \gamma_1 I & \gamma_1 \tilde{R}_{11}^{-1}\tilde{R}_{12} & \gamma_1 \tilde{R}_{11}^{-1}\tilde{R}_{13} \\ \gamma_2 \tilde{R}_{22}^{-1}\tilde{R}_{21} & \gamma_2 I & \gamma_2 \tilde{R}_{22}^{-1}\tilde{R}_{23} \\ \gamma_3 \tilde{R}_{33}^{-1}\tilde{R}_{31} & \gamma_3 \tilde{R}_{33}^{-1}\tilde{R}_{32} & \gamma_3 I \end{pmatrix} + 2\gamma_{12} \begin{pmatrix} I & 0 & \begin{pmatrix} \tilde{S}_{11} & \tilde{S}_{12} \\ \tilde{S}_{21} & \tilde{S}_{22} \end{pmatrix} \begin{pmatrix} \tilde{R}_{13} \\ \tilde{R}_{23} \end{pmatrix} \\ 0 & I & \\ 0 & 0 & 0 \end{pmatrix}$$
$$+2\gamma_{13} \begin{pmatrix} I & \tilde{T}_{11}\tilde{R}_{12} + \tilde{T}_{13}\tilde{R}_{32} & 0 \\ 0 & 0 & 0 \\ 0 & \tilde{T}_{31}\tilde{R}_{12} + \tilde{T}_{33}\tilde{R}_{32} & I \end{pmatrix} + 2\gamma_{23} \begin{pmatrix} 0 & 0 & 0 \\ \begin{pmatrix} \tilde{U}_{22} & \tilde{U}_{23} \\ \tilde{U}_{32} & \tilde{U}_{33} \end{pmatrix} \begin{pmatrix} \tilde{R}_{21} \\ \tilde{R}_{31} \end{pmatrix} & I & 0 \\ & 0 & I \end{pmatrix}$$
$$+2\gamma_{123}\, I - \operatorname{diag} \begin{pmatrix} \gamma_{12}\tilde{S}_{11} + \gamma_{13}\tilde{T}_{11} + \\ \gamma_{123}\tilde{V}_{11} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ \gamma_{12}\tilde{S}_{22} + \gamma_{23}\tilde{U}_{22} + \\ \gamma_{123}\tilde{V}_{22} \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ \gamma_{13}\tilde{T}_{33} + \gamma_{23}\tilde{U}_{33} + \\ \gamma_{123}\tilde{V}_{33} \end{pmatrix} \cdot \tilde{R} = 0 \tag{5.19}$$

Let,

$$\tilde{D}_1 \triangleq \operatorname{diag}(\gamma_{12}\tilde{S}_{11} + \gamma_{13}\tilde{T}_{11} + \gamma_{123}\tilde{V}_{11}) \tag{5.20}$$

$$\tilde{D}_2 \triangleq \operatorname{diag}(\gamma_{12}\tilde{S}_{22} + \gamma_{23}\tilde{U}_{22} + \gamma_{123}\tilde{V}_{22}) \tag{5.21}$$

$$\tilde{D}_3 \triangleq \operatorname{diag}(\gamma_{13}\tilde{T}_{33} + \gamma_{23}\tilde{U}_{33} + \gamma_{123}\tilde{V}_{33}). \tag{5.22}$$

Then by equating the diagonal term of the left-hand side of (5.19) to zero we get,

$$(\gamma_i + 2\sum_j \gamma_{ij} + 2\gamma_{123})I = \tilde{D}_i \tilde{R}_{ii} \tag{5.23}$$

where we have implicitly assumed that $\gamma_{ij} = \gamma_{ji}$. This gives,

$$\tilde{R}_{ii} = (\gamma_i + 2\sum_j \gamma_{ij} + 2\gamma_{123})\tilde{D}_i^{-1}. \tag{5.24}$$

In other words $\tilde{R}_{ii}$ should be diagonal. If we assume that $\det \tilde{R}_{ii} = \alpha_{ii}^T,\ \alpha_{ii} > 0$, then the following matrix has unit determinant,

$$L = \begin{pmatrix} \frac{1}{\sqrt{\alpha_{11}}}\tilde{R}_{11}^{1/2} & 0 & 0 \\ 0 & \frac{1}{\sqrt{\alpha_{22}}}\tilde{R}_{22}^{1/2} & 0 \\ 0 & 0 & \frac{1}{\sqrt{\alpha_{33}}}\tilde{R}_{33}^{1/2} \end{pmatrix} \tag{5.25}$$

and we can go back and multiply (5.18) from left and right by the above matrix L. Considering the fact that $\tilde{R}_{ii}$ are diagonal, and denoting

$$L_{[1,2]} = \begin{pmatrix} \frac{1}{\sqrt{\alpha_{11}}}\tilde{R}_{11}^{1/2} & 0 \\ 0 & \frac{1}{\sqrt{\alpha_{22}}}\tilde{R}_{22}^{1/2} \end{pmatrix}, \tag{5.26}$$

we note that

$$L \begin{pmatrix} \mathrm{diag}(\tilde{S}_{11}) & 0 & 0 \\ 0 & \mathrm{diag}(\tilde{S}_{22}) & 0 \\ 0 & 0 & 0 \end{pmatrix} L \ = \ \begin{pmatrix} \mathrm{diag}\left(\left[L_{[1,2]}^{-1} \begin{pmatrix} \tilde{R}_{11} & \tilde{R}_{12} \\ \tilde{R}_{21} & \tilde{R}_{22} \end{pmatrix} L_{[1,2]}^{-1} \right]^{-1} \right) & 0 \\ 0 \quad 0 & 0 \end{pmatrix}. \tag{5.27}$$

Therefore if we define,

$$R = L^{-1}\tilde{R}L^{-1} \tag{5.28}$$

and S, T, U, and V also similar to (5.17),

$$\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}^{-1}, \begin{pmatrix} T_{11} & T_{13} \\ T_{31} & T_{33} \end{pmatrix} = \begin{pmatrix} R_{11} & R_{13} \\ R_{31} & R_{33} \end{pmatrix}^{-1}$$
$$\begin{pmatrix} U_{22} & U_{23} \\ U_{32} & U_{33} \end{pmatrix} = \begin{pmatrix} R_{22} & R_{23} \\ R_{32} & R_{33} \end{pmatrix}^{-1}, \; V_{ij} = (R^{-1})_{ij} \tag{5.29}$$

it follows that (5.18) and (5.19) will be satisfied by R, S, T, U, V instead of $\tilde{R}, \tilde{S}, \tilde{T}, \tilde{U}, \tilde{V}$. However note that now we have, $R_{ii} = \alpha_{ii} I$. Next similar to (5.20), (5.21), and (5.22) we can define D_1, D_2, and D_3 for S, T, U, V. Then equating the diagonal terms of (5.19) to zero when R is used instead of $\tilde{R}$ yields,

$$(\gamma_i + 2\sum_j \gamma_{ij} + 2\gamma_{123})I = \alpha_{ii} D_i \tag{5.30}$$

which immediately gives,

$$D_i = \delta_i I \tag{5.31}$$

for some constant δ_i. Now considering elements (2,1), (3,1) together, (1,2), (3,2) with each other, and (1,3), (2,3) simultaneously in (5.19) when $\tilde{R}$ is replaced by R and using $R_{ii} = \alpha_{ii} I$, we obtain for $i, j, k \in \{1, 2, 3\}$,

$$\left(\begin{bmatrix} (\delta_i + \frac{\gamma_i}{\alpha_{ii}})I & 0 \\ 0 & (\delta_j + \frac{\gamma_j}{\alpha_{jj}})I \end{bmatrix} + 2\gamma_{ij} \begin{bmatrix} R_{ii} & R_{ij} \\ R_{ji} & R_{jj} \end{bmatrix}^{-1} \right) \begin{bmatrix} R_{ik} \\ R_{jk} \end{bmatrix} = 0. \tag{5.32}$$

Therefore simplifying condition (5.32) by multiplying it by $\begin{bmatrix} R_{ii} & R_{ij} \\ R_{ji} & R_{jj} \end{bmatrix}$, we obtain:

$$\begin{bmatrix} (\gamma_i + 2\gamma_{ij} + \delta_i\alpha_{ii})I & \frac{\gamma_j+\delta_j\alpha_{jj}}{\alpha_{jj}}R_{ij} \\ \frac{\gamma_i+\delta_i\alpha_{ii}}{\alpha_{ii}}R_{ji} & (\gamma_j + 2\gamma_{ij} + \delta_j\alpha_{jj})I \end{bmatrix} \begin{bmatrix} R_{ik} \\ R_{jk} \end{bmatrix} = 0. \tag{5.33}$$

Now if the $2T \times T$ matrix $\begin{bmatrix} R_{ik}^t & R_{jk}^t \end{bmatrix}^t$ were full rank, the rank of the left $2T \times 2T$ matrix would be T and therefore its Schur complement should be zero, i.e.:

$$(\gamma_j + 2\gamma_{ij} + \delta_j\alpha_{jj})I - \frac{(\frac{\gamma_i+\delta_i\alpha_{ii}}{\alpha_{ii}})(\frac{\gamma_j+\delta_j\alpha_{jj}}{\alpha_{jj}})}{\gamma_i + 2\gamma_{ij} + \delta_i\alpha_{ii}}R_{ji}R_{ij} = 0 \tag{5.34}$$

in other words:

$$R_{ji}R_{ij} = R_{ij}R_{ji} \triangleq \beta_{ij}I. \tag{5.35}$$

Since R is symmetric, $R_{ji} = R_{ij}^t$, this implies that off-diagonal blocks of R are multiples of an orthogonal matrix. However, in the general case $\begin{bmatrix} R_{ik}^t & R_{jk}^t \end{bmatrix}^t$ need not be full rank. Therefore there is a $T \times T$ unitary matrix θ_{ij} such that:

$$\begin{bmatrix} R_{ik} \\ R_{jk} \end{bmatrix} \theta_{ij} = \begin{bmatrix} \overline{R}_{ik} & 0 \\ \overline{R}_{jk} & 0 \end{bmatrix}. \tag{5.36}$$

Assume the column rank of $\begin{bmatrix} \overline{R}_{ik}^t & \overline{R}_{jk}^t \end{bmatrix}^t$ to be T_{ij}. This suggests doing a similarity transformation on R with the following unitary matrix without affecting the block principal minors:

$$\Theta = \begin{bmatrix} \theta_{23} & 0 & 0 \\ 0 & \theta_{31} & 0 \\ 0 & 0 & \theta_{12} \end{bmatrix}. \tag{5.37}$$

From which we obtain:

$$\Theta^* R \Theta = \begin{bmatrix} \alpha_{11} I & \theta^*_{23} R_{12} \theta_{31} & \theta^*_{23} R_{13} \theta_{12} \\ \theta^*_{31} R_{21} \theta_{23} & \alpha_{22} I & \theta^*_{31} R_{23} \theta_{12} \\ \theta^*_{12} R_{31} \theta_{23} & \theta^*_{12} R_{32} \theta_{31} & \alpha_{33} I \end{bmatrix}. \tag{5.38}$$

Considering $R_{21}\theta_{23}$ and $\theta^*_{31}R_{21}$ simultaneously and using (5.36) we have,

$$R_{21}\Theta_{23} = \begin{pmatrix} \overline{R}_{21} & 0 \end{pmatrix} \tag{5.39}$$

$$\Theta^*_{31} R_{21} = \begin{pmatrix} \overline{R}^*_{12} \\ 0 \end{pmatrix}. \tag{5.40}$$

Therefore we can simply obtain the following structure for $\theta^*_{31}R_{21}\theta_{23}$:

$$\theta^*_{31} R_{21} \theta_{23} = \begin{bmatrix} \hat{R}_{21} & 0 \\ 0 & 0 \end{bmatrix} \tag{5.41}$$

where the dimension of $\hat{R}_{21}$ is $T_{31} \times T_{23}$. A similar argument for other elements yields the following structure for $\Theta^* R \Theta$:

$$\begin{bmatrix} \alpha_{11} I_{T_{23}} & 0 & \hat{R}_{12} & 0 & \hat{R}_{13} & 0 \\ 0 & \alpha_{11} I_{T-T_{23}} & 0 & 0 & 0 & 0 \\ \hat{R}_{21} & 0 & \alpha_{22} I_{T_{31}} & 0 & \hat{R}_{23} & 0 \\ 0 & 0 & 0 & \alpha_{22} I_{T-T_{31}} & 0 & 0 \\ \hat{R}_{31} & 0 & \hat{R}_{32} & 0 & \alpha_{33} I_{T_{12}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \alpha_{33} I_{T-T_{12}} \end{bmatrix}. \tag{5.42}$$

Now we go back to (5.33) and rewrite it as follows (we assume that $\Theta_{ij} = \Theta_{ji}$),

$$\begin{pmatrix} \Theta^*_{jk} & 0 \\ 0 & \Theta^*_{ik} \end{pmatrix} \begin{pmatrix} (\gamma_i + 2\gamma_{ij} + \delta_i\alpha_{ii})I & \frac{\gamma_j+\delta_j\alpha_{jj}}{\alpha_{jj}} R_{ij} \\ \frac{\gamma_i+\delta_i\alpha_{ii}}{\alpha_{ii}} R_{ji} & (\gamma_j + 2\gamma_{ij} + \delta_j\alpha_{jj})I \end{pmatrix} \times \begin{pmatrix} \Theta_{jk} & 0 \\ 0 & \Theta_{ik} \end{pmatrix} \begin{pmatrix} \Theta^*_{jk} & 0 \\ 0 & \Theta^*_{ik} \end{pmatrix} \begin{pmatrix} R_{ik} \\ R_{jk} \end{pmatrix} \Theta_{ij} = 0 \tag{5.43}$$

which essentially gives,

$$\begin{pmatrix} (\gamma_i + 2\gamma_{ij} + \delta_i\alpha_{ii})I & \frac{\gamma_j+\delta_j\alpha_{jj}}{\alpha_{jj}} \Theta^*_{jk} R_{ij} \Theta_{ik} \\ \frac{\gamma_i+\delta_i\alpha_{ii}}{\alpha_{ii}} \Theta^*_{ik} R_{ji} \Theta_{jk} & (\gamma_j + 2\gamma_{ij} + \delta_j\alpha_{jj})I \end{pmatrix} \begin{pmatrix} \Theta^*_{jk} R_{ik} \Theta_{ij} \\ \Theta^*_{ik} R_{jk} \Theta_{ij} \end{pmatrix} = 0. \tag{5.44}$$

Using (5.42) and plugging the relevant entries back into (5.44) we obtain,

$$\begin{bmatrix} (\gamma_i + 2\gamma_{ij} + \delta_i\alpha_{ii})I_{T_{jk}} & \frac{\gamma_j+\delta_j\alpha_{jj}}{\alpha_{jj}} \hat{R}_{ij} \\ \frac{\gamma_i+\delta_i\alpha_{ii}}{\alpha_{ii}} \hat{R}_{ji} & (\gamma_j + 2\gamma_{ij} + \delta_j\alpha_{jj})I_{T_{ki}} \end{bmatrix} \begin{bmatrix} \hat{R}_{ik} \\ \hat{R}_{jk} \end{bmatrix} = 0. \tag{5.45}$$

Note that the dimension of $\begin{bmatrix} \hat{R}^t_{ik} & \hat{R}^t_{jk} \end{bmatrix}^t$ is $(T_{jk} + T_{ki}) \times T_{ij}$. If we let the rank of the left matrix in (5.45) be r we will have:

$$r \leq T_{jk} + T_{ki} - T_{ij}. \tag{5.46}$$

On the other hand it is also obvious that:

$$r \geq T_{jk},\ T_{ki}. \tag{5.47}$$

From (5.46) and (5.47) it follows that:

$$T_{ij} \leq \min(T_{jk}, T_{ki}). \tag{5.48}$$

Since a similar argument can be used for T_{jk} and T_{ki} we conclude that:

$$T_{12} = T_{23} = T_{31} \triangleq \hat{T}. \tag{5.49}$$

Now note that (5.45) is similar to (5.33) with $\hat{R}_{ij}$ instead of R_{ij}. Therefore the same argument leads to the conclusion that $\hat{R}_{ij}$ is a multiple of an orthogonal matrix, say Φ_{ij}; in other words $\hat{R}_{ij} = \alpha_{ij}\Phi_{ij}$. From which it follows that after a series of permutations, $\Theta^* R \Theta$ can be written as follows:

$$\begin{bmatrix} \alpha_{11}I_{\hat{T}} & \alpha_{12}\Phi_{12} & \alpha_{13}\Phi_{13} & 0 & 0 & 0 \\ \alpha_{12}\Phi_{12}^t & \alpha_{22}I_{\hat{T}} & \alpha_{23}\Phi_{23} & 0 & 0 & 0 \\ \alpha_{13}\Phi_{13}^t & \alpha_{23}\Phi_{23}^t & \alpha_{33}I_{\hat{T}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha_{11}I_{T-\hat{T}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \alpha_{22}I_{T-\hat{T}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \alpha_{33}I_{T-\hat{T}} \end{bmatrix} \tag{5.50}$$

which if viewed as the timeshare of a set of Gaussian random variables with an orthogonal covariance matrix and another set of independent random variables, it has the same block principal minors as (5.42), and moreover this structure is also consistent with the requirement (5.31). Note that (5.50) is an optimal solution for optimization (5.16) only if it is a positive semi-definite matrix. Therefore α_{ij}'s and Φ_{ij}'s should be such that,

$$\alpha_{ii}\alpha_{jj} - \alpha_{ij}^2 \geq 0 \tag{5.51}$$

$$\det \begin{pmatrix} \alpha_{11}I_T & \alpha_{12}\Phi_{12} & \alpha_{13}\Phi_{13} \\ \alpha_{12}\Phi_{12}^t & \alpha_{22}I_T & \alpha_{23}\Phi_{23} \\ \alpha_{13}\Phi_{13}^t & \alpha_{23}\Phi_{23}^t & \alpha_{33}I_T \end{pmatrix} \geq 0. \tag{5.52}$$

Finally since R was obtained from $\tilde{R}$ by a multiplication by a unit determinant matrix, it has the same minors as $\tilde{R}$ and therefore R is the optimal solution of the main optimization problem (5.15). □

Lemma 5.3.6 (Block Orthogonal, Block Diagonal Covariance) *Consider the covariance matrix*

$$R = \begin{bmatrix} \alpha_{11} I_T & \alpha_{12}\Phi_{12} & \alpha_{13}\Phi_{13} \\ \alpha_{12}\Phi_{12}^t & \alpha_{22} I_T & \alpha_{23}\Phi_{23} \\ \alpha_{13}\Phi_{13}^t & \alpha_{23}\Phi_{23}^t & \alpha_{33} I_T \end{bmatrix}, \tag{5.53}$$

where the Φ_{ij} are orthogonal, $\alpha_{ii} > 0$, and the 2×2 block principal minors $m_{ij} = p_{ij}^T = (\alpha_{ii}\alpha_{jj} - \alpha_{ij}^2)^T$ are such that $p_{ij} \geq 0$. Then

$$detR \leq \Big(\alpha_{11}\alpha_{22}\alpha_{33} - \alpha_{11}\alpha_{23}^2 - \alpha_{22}\alpha_{13}^2 - \alpha_{33}\alpha_{12}^2 + 2|\alpha_{12}\alpha_{13}\alpha_{23}|\Big)^T \tag{5.54}$$

with equality iff $\Phi + \Phi^t = 2I$ where $\Phi = \Phi_{13}^t \Phi_{12} \Phi_{23}$.

Proof: We can easily write the following,

$$\begin{aligned} \det R &= \frac{1}{\alpha_{11}^T} \det \left(\begin{pmatrix} \alpha_{11}\alpha_{22} I_T & \alpha_{11}\alpha_{23}\Phi_{23} \\ \alpha_{11}\alpha_{23}\Phi_{23}^t & \alpha_{11}\alpha_{33} I_T \end{pmatrix} - \begin{pmatrix} \alpha_{12}\Phi_{12}^t \\ \alpha_{13}\Phi_{13}^t \end{pmatrix} \begin{pmatrix} \alpha_{12}\Phi_{12} & \alpha_{13}\Phi_{13} \end{pmatrix} \right) \\ &= \frac{1}{\alpha_{11}^T} \det \Big((\alpha_{11}\alpha_{22} - \alpha_{12}^2)(\alpha_{11}\alpha_{33} - \alpha_{13}^2) I_T \\ &\quad -(\alpha_{11}\alpha_{23}\Phi_{23}^t - \alpha_{12}\alpha_{13}\Phi_{13}^t\Phi_{12})(\alpha_{11}\alpha_{23}\Phi_{23} - \alpha_{12}\alpha_{13}\Phi_{12}^t\Phi_{13}) \Big) \\ &= \det \Big((\alpha_{11}\alpha_{22}\alpha_{33} - \alpha_{11}\alpha_{23}^2 - \alpha_{22}\alpha_{13}^2 - \alpha_{33}\alpha_{12}^2) I_T \\ &\quad +\alpha_{12}\alpha_{13}\alpha_{23}(\Phi_{13}^t\Phi_{12}\Phi_{23} + \Phi_{23}^t\Phi_{12}^t\Phi_{13}) \Big). \end{aligned} \tag{5.55}$$

The result immediately follows from $-2I \leq \Phi + \Phi^t \leq 2I$. □

From Lemma 5.3.5 and Lemma 5.3.6, minors of the optimal points with covariance matrix (5.50) can be obtained,

$$m_i = \alpha_{ii}^T \tag{5.56}$$

$$m_{ij} = (\alpha_{ii}\alpha_{jj} - \alpha_{ij}^2)^{\hat{T}}(\alpha_{ii}\alpha_{jj})^{T-\hat{T}} \tag{5.57}$$

$$\max_{\Phi} m_{123} = (\alpha_{11}\alpha_{22}\alpha_{33})^{T-\hat{T}} \times\Big(\alpha_{11}\alpha_{22}\alpha_{33} - \alpha_{11}\alpha_{23}^2 - \alpha_{22}\alpha_{13}^2 - \alpha_{33}\alpha_{12}^2 + 2|\alpha_{12}\alpha_{13}\alpha_{23}|\Big)^{\hat{T}}. \tag{5.58}$$

However these values can also be obtained by a timeshare of 3 scalar random variables with covariance matrix,

$$\begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{12} & \alpha_{22} & \alpha_{23} \\ \alpha_{13} & \alpha_{23} & \alpha_{33} \end{pmatrix} \tag{5.59}$$

and 3 other independent scalar random variables. This suggests that the region of 3 vector-valued Gaussian random variables may be obtained from the convex hull region of 3 scalar Gaussian random variables. In other words for $n = 3$, considering vector-valued random variables will not give any entropy vector that is not obtainable from scalar valued ones. This is essentially the statement of Theorem 5.3.2 and we can now proceed to its proof.

Proof of Theorem 5.3.2: As in Lemma 5.3.5, we write the following optimization problem,

$$\max_{R} \sum_{s\subseteq\{1,2,3\}} \gamma_s \log\det R_s, \tag{5.60}$$

and follow the steps therein to obtain equation (5.45). From (5.45) and using $T_{12} = T_{23} = T_{31} = \hat{T}$, we can write the following,

$$\rho_{23}\hat{R}_{21} = -\tau_3\hat{R}_{23}\hat{R}_{31} \tag{5.61}$$

$$\rho_{32}\hat{R}_{31} = -\tau_2\hat{R}_{32}\hat{R}_{21} \tag{5.62}$$

$$\rho_{31}\hat{R}_{32} = -\tau_1\hat{R}_{31}\hat{R}_{12} \tag{5.63}$$

where,

$$\rho_{ij} = \gamma_i + 2\gamma_{ij} + \delta_i\alpha_{ii} \tag{5.64}$$

$$\tau_k = \frac{\gamma_k + \delta_k\alpha_{kk}}{\alpha_{kk}}. \tag{5.65}$$

Now if the elements $\hat{R}_{21}$ and $\hat{R}_{31}$ have the following QR factorization,

$$\hat{R}_{21} = \hat{Q}_{21}\overline{R}_{21} \tag{5.66}$$

$$\hat{R}_{31} = \hat{Q}_{31}\overline{R}_{31} \tag{5.67}$$

We can plug these back into (5.61)–(5.63) to obtain,

$$\rho_{23}\overline{R}_{21} = -\tau_3(\hat{Q}_{21}^*\hat{R}_{23}\hat{Q}_{31})\overline{R}_{31} \tag{5.68}$$

$$\rho_{32}\overline{R}_{31} = -\tau_2(\hat{Q}_{31}^*\hat{R}_{32}\hat{Q}_{21})\overline{R}_{21} \tag{5.69}$$

$$\rho_{31}(\hat{Q}_{31}^*\hat{R}_{32}\hat{Q}_{21}) = \tau_1\overline{R}_{31}\overline{R}_{21}^*. \tag{5.70}$$

Since $\overline{R}_{21}$ and $\overline{R}_{31}$ are upper-triangular, from (5.68) and (5.69) it follows that $\hat{Q}_{21}^*\hat{R}_{23}\hat{Q}_{31}$ and $\hat{Q}_{31}^*\hat{R}_{32}\hat{Q}_{21}$ should also be upper-triangular. However $\hat{Q}_{21}^*\hat{R}_{23}\hat{Q}_{31} = (\hat{Q}_{31}^*\hat{R}_{32}\hat{Q}_{21})^*$ and therefore $\hat{Q}_{21}^*\hat{R}_{23}\hat{Q}_{31}$ should be diagonal. Next from (5.63) it follows that $\overline{R}_{31}\overline{R}_{21}^*$ should be diagonal. However since $\overline{R}_{31}$ and $\overline{R}_{21}$ are both full rank and upper-triangular, this can be satisfied only if they are both diagonal as well. Therefore

if we define U as,

$$U = \begin{pmatrix} I_{\hat{T}} & 0 & 0 & 0 & 0 & 0 \\ 0 & I_{T-\hat{T}} & 0 & 0 & 0 & 0 \\ 0 & 0 & \hat{Q}_{21} & 0 & 0 & 0 \\ 0 & 0 & 0 & I_{T-\hat{T}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \hat{Q}_{31} & 0 \\ 0 & 0 & 0 & 0 & 0 & I_{T-\hat{T}} \end{pmatrix} \tag{5.71}$$

we can write the following,

$$R = U^* \bar{R} U \tag{5.72}$$

where,

$$\bar{R} = \begin{pmatrix} \alpha_{11} I_{\hat{T}} & 0 & \overline{R}_{21}^* & 0 & \overline{R}_{31}^* & 0 \\ 0 & \alpha_{11} I_{T-\hat{T}} & 0 & 0 & 0 & 0 \\ \overline{R}_{21} & 0 & \alpha_{22} I_{\hat{T}} & 0 & \hat{Q}_{21}^* R_{23} \hat{Q}_{31} & 0 \\ 0 & 0 & 0 & \alpha_{22} I_{T-\hat{T}} & 0 & 0 \\ \overline{R}_{31} & 0 & \hat{Q}_{31}^* R_{32} \hat{Q}_{21} & 0 & \alpha_{33} I_{\hat{T}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \alpha_{33} I_{T-\hat{T}} \end{pmatrix} \tag{5.73}$$

where $\overline{R}_{21}$ and $\overline{R}_{31}$ are diagonal. In fact, since all blocks of $\bar{R}$ are diagonal, it can be viewed as a timeshare of scalar random variables. Moreover, since $\bar{R}$ is obtained from R in such a way that it has the same minors as R, therefore $\bar{R}$ is an optimal solution of optimization (5.60) as well. □

In order to prove Theorem 5.3.3 we further need the following lemma:

Lemma 5.3.7 *Consider the function*

$$f(\theta) = \left(\max \left[0, -2 + \sum_{l=1}^{3} x_l^{\frac{1}{\theta}} + 2\sqrt{\prod_{l=1}^{3}(1 - x_l^{\frac{1}{\theta}})} \right] \right)^{\theta}, \tag{5.74}$$

where $0 < x_l \leq 1$, *for* $l = 1, 2, 3$. f *has either a single maximum or supremum given by:*

$$\max_{\theta} f(\theta) = \frac{\prod_l x_l}{\max_l(x_l)}. \tag{5.75}$$

Moreover, if we let $\tilde{y} = \frac{\prod_l x_l}{\max_l x_l} + 2\max_l x_l - \sum_l x_l$,

$$\max_{0 \leq \theta \leq 1} f(\theta) = \begin{cases} \frac{\prod_l x_l}{\max_l(x_l)} & \text{If } \tilde{y} \leq 0 \\ f(1) & \text{If } \tilde{y} > 0 \end{cases}. \tag{5.76}$$

Proof: We will first show that $\forall i, j,\ f(\theta) \leq x_i x_j$. Let,

$$e(\theta) = -2 + \sum_{l=1}^{3} x_l^{\frac{1}{\theta}} + 2\sqrt{\prod_{l=1}^{3}(1 - x_l^{\frac{1}{\theta}})}. \tag{5.77}$$

For distinct $i, j, k \subseteq \{1, 2, 3\}$, this can also be written as,

$$\begin{aligned} e(\theta) &= (x_i x_j)^{\frac{1}{\theta}} - \left((1 - x_i^{\frac{1}{\theta}})(1 - x_j^{1}\theta) + (1 - x_k^{\frac{1}{\theta}}) - 2\sqrt{(1 - x_i^{\frac{1}{\theta}})(1 - x_j^{\frac{1}{\theta}})(1 - x_k^{\frac{1}{\theta}})} \right) \\ &= (x_i x_j)^{\frac{1}{\theta}} - \left(\sqrt{(1 - x_i^{\frac{1}{\theta}})(1 - x_j^{\frac{1}{\theta}})} - \sqrt{1 - x_k^{\frac{1}{\theta}}} \right)^2 \end{aligned} \tag{5.78}$$

which shows $e(\theta) \leq (x_i x_j)^{\frac{1}{\theta}}$, and therefore for all $\theta \geq 0$, $f(\theta) \leq x_i x_j$ with equality, if and only if, $(1 - x_i^{\frac{1}{\theta}})(1 - x_j^{\frac{1}{\theta}}) = 1 - x_k^{1\theta}$, or equivalently,

$$(x_i x_j)^{\frac{1}{\theta}} + x_k^{\frac{1}{\theta}} - x_i^{\frac{1}{\theta}} - x_j^{\frac{1}{\theta}} = 0. \tag{5.79}$$

Note that this is only possible when $x_i x_j = \frac{x_1 x_2 x_3}{\max_l x_l}$. Without loss of generality, assume $x_1 \leq x_2 \leq x_3$ and define,

$$y(\theta) = (x_1 x_2)^{\frac{1}{\theta}} + x_3^{\frac{1}{\theta}} - x_1^{\frac{1}{\theta}} - x_2^{\frac{1}{\theta}}. \tag{5.80}$$

Clearly zeros of $y(\theta)$ determine the global maximums of $f(\theta)$. We analyze the behavior of $y(\theta)$ in the following scenarios (based on initial assumption, $x_1 \leq x_2 \leq x_3$):

- x_1, x_2, and x_3 are distinct, and $x_1, x_2, x_3 \neq 1$
- $x_1 = x_2 < x_3 \neq 1$
- $x_1 < x_2 = x_3 \neq 1$
- $x_1 = x_2 = x_3 = x < 1$
- $x_1 < x_2 \neq 1$, and $x_3 = 1$
- $x_1 = x_2 = x < 1$, and $x_3 = 1$
- $x_1 \leq x_2 = x_3 = 1$

In all of the above cases, we find that $y(\theta)$ has at most one zero, say θ^*, which is not at 0, or ∞. Moreover, $y(0)$ may or may not be zero, and $y(\infty)$ is also always (asymptotically) zero. Therefore, the global maximums of $f(\theta)$, may occur at 0, ∞, or at the unique horizontal-axis crossing of $y(\theta)$ (if it exists). Analyzing the behavior of $f(\theta)$ at these 3 points in the above cases, reveals that $f(\theta)$ has a unique global maximum, or a supremum. Therefore, $f(\theta)$ always achieves $\frac{x_1 x_2 x_3}{\max_l x_l}$ either at some specific θ^*, at 0, or asymptotically. Next it can be shown that if for some $\theta', y(\theta') > 0$ then $f(\theta')$ has not reached its global supremum for $\theta < \theta'$. Combining this with the quasi-concavity property of f, yields the desired result. □

Now we can proceed to the proof of Theorem 5.3.3.

Proof of Theorem 5.3.3: To find the boundary entropies of the region we use Lemma (5.3.5) to time-share a set of independent random variables with covariance matrix of block size $T-\hat{T}$ and another set of random variables with orthogonal covariance matrix of block size $\hat{T}$ (5.50). Calculating the determinant of this matrix and using Lemma (5.3.6) we obtain:

$$\begin{aligned}\max_{\Phi} \det R &= (\alpha_{11}\alpha_{22}\alpha_{33})^{T-\hat{T}} \\ &\quad \times\Big(\alpha_{11}\alpha_{22}\alpha_{33}-\alpha_{11}\alpha_{23}^2-\alpha_{22}\alpha_{13}^2-\alpha_{33}\alpha_{12}^2+2|\alpha_{12}\alpha_{13}\alpha_{23}|\Big)^{\hat{T}} \quad (5.81)\end{aligned}$$

Let m be the vector of block principal minors of the above matrix and let $p=m^{\frac{1}{T}}$ where the exponential acts componentwise. Then if we assume p_i and p_{ij} are fixed, it is easy to see that $\alpha_{ii}=p_i\geq 0$ and $\alpha_{ij}=\pm\sqrt{p_ip_j(1-(\frac{p_{ij}}{p_ip_j})^{\frac{T}{\hat{T}}})}$. This imposes the constraint:

$$p_{ij}\leq p_ip_j. \quad (5.82)$$

Assuming $\theta=\frac{\hat{T}}{T}$ and substituting these in (5.81) results in:

$$\begin{aligned}\max_{\Phi} p_{123} = p_1p_2p_3\Bigg(&-2+\left(\frac{p_{12}}{p_1p_2}\right)^{\frac{1}{\theta}}+\left(\frac{p_{13}}{p_1p_3}\right)^{\frac{1}{\theta}}+\left(\frac{p_{23}}{p_2p_3}\right)^{\frac{1}{\theta}} \\ &+2\sqrt{\left(1-\left(\frac{p_{12}}{p_1p_2}\right)^{\frac{1}{\theta}}\right)\left(1-\left(\frac{p_{13}}{p_1p_3}\right)^{\frac{1}{\theta}}\right)\left(1-\left(\frac{p_{23}}{p_2p_3}\right)^{\frac{1}{\theta}}\right)}\Bigg)^{\theta} \quad (5.83)\end{aligned}$$

Of course this corresponds to the determinant of a covariance matrix of some Gaussian random variables only if the term inside the braces in (5.83) is positive. Therefore assuming $x_1=\frac{p_{12}}{p_1p_2},\ x_2=\frac{p_{23}}{p_2p_3},\ x_3=\frac{p_{13}}{p_3p_1}$, and using (5.74) in Lemma 5.3.7:

$$\sup_{\Phi,\ \theta} p_{123} = p_1p_2p_3 \sup_{0<\theta\leq 1} f(\theta). \quad (5.84)$$

Note that since we have fixed p_i and p_{ij}, and that θ represents the timesharing of 2 sets of random variables, $\theta = 0$ is not generally allowed (otherwise we enforce the random variables to be independent). Therefore we have used sup instead of max in (5.84). Now what remains, is to find $\sup f(\theta)$ with respect to θ over $0 < \theta \leq 1$. Using Lemma 5.3.7 we obtain,

$$\sup_{\Phi,\ \theta} p_{123} = \begin{cases} \min_j \frac{p_{ij}p_{jk}}{p_j} & \text{If } \tilde{y} \leq 0 \\ p_1p_2p_3 \max\left(0,\ -2 + \sum_{k=1}^{3} x_k + 2\sqrt{\prod_{k=1}^{3}(1-x_k)}\right) & \text{If } \tilde{y} > 0 \end{cases} \tag{5.85}$$

Replacing p with the corresponding entropies ($p = e^g$) in (5.85) and also (5.82) gives (5.11) and (5.12). Finally, since $\sup p_{123}$ is found, (5.85) characterizes the closure of the region. □

Proof of Theorem 5.3.4: Let g be an arbitrary entropy vector for which $\tilde{y} > 0$, and therefore it does not fall in the Gaussian region. If $\max_k x_k < 1$, let $\theta^* = \text{argmax}_\theta f(\theta)$. Now for any $\theta' \geq \theta^*$ define the normalized entropy vector $g' = \frac{1}{\theta'}g$, and the corresponding $x'_k = e^{g'_{ij}-g'_i-g'_j}$. Clearly $x'_k = x_k^{\frac{1}{\theta'}}$ and $y'(\theta) = y(\theta\theta')$. Therefore $\tilde{y}' = y(\theta')$. From Lemma (5.3.7) it follows that when $\max_k x_k < 1$ and $\tilde{y} > 0$, the function $y(\theta)$ has a single zero which coincides with the maximizing point of $f(\theta)$, namely θ^*. As a result for all $\theta' \geq \theta^*$, $y(\theta') < 0$ which immediately translates to $\tilde{y}' < 0$, meaning that the maximum of the corresponding function f' will happen for some $0 \leq \theta \leq 1$ and therefore by Theorem (5.3.3), $g' = \frac{1}{\theta'}g$ can be generated by Gaussians. On the other hand if $\max_k x_k = 1$, $\forall\theta,\ y(\theta) \geq 0$ and $\lim_{\theta\to\infty} y(\theta) = 0$. In terms of the function f we have, $\lim_{\theta\to\infty} f(\theta) = \min_{i,j} x_i x_j$. Nonetheless since $f(\theta)$ achieves its supremum in an asymptotic manner, it means that a small perturbation of g' will put it in the Gaussian region and hence g will be in the closure of the convex cone of Gaussian entropy region. □

5.4 Cayley's Hyperdeterminant

Recall from (5.4) that the entropy of a collection of Gaussian random variables is simply the "log-determinant" of their covariance matrix. Similarly, the entropy of any subset of variables from a collection of Gaussian random variables is simply the "log" of the principal minor of the covariance matrix corresponding to this subset. Therefore one approach to characterizing the entropy region of Gaussians is to study the determinantal relations of a symmetric positive semi-definite matrix.

For example, consider 3 Gaussian random variables. While the entropy vector of 3 random variables is a 7-dimensional object, there are only 6 free parameters in a symmetric positive semi-definite matrix. Therefore the minors should satisfy a relation which is essentially implied by (5.53) when the matrix entries are expressed in terms of the principal minors. It has very recently been shown that this relation is given by the Cayley's so-called $2 \times 2 \times 2$ "hyperdeterminant" [HS07b]. The hyperdeterminant is a generalization of the determinant concept for matrices to tensors and it was first introduced by Cayley in 1845 [Cay45].

There are a couple of equivalent definitions for the hyperdeterminant among which we choose the definition through the degeneracy of a multilinear form. Consider the following multilinear form of the format $(k_1+1) \times (k_2+1) \times \ldots \times (k_n+1)$ in variables $X_1, \ldots, X_n$ where each variable X_j is a vector of length (k_j+1) with elements in $\mathbb{C}$:

$$f(X_1, X_2, \ldots, X_n) = \sum_{i_1=0}^{k_1} \sum_{i_2=0}^{k_2} \cdots \sum_{i_n=0}^{k_n} a_{i_1,i_2,\ldots,i_n} x_{1,i_1} x_{2,i_2}, \ldots, x_{n,i_n}. \tag{5.86}$$

The multilinear form f is said to be degenerate if and only if there is a non-trivial solution $(X_1, X_2, \ldots, X_n)$ to the following system of partial derivative equations [GKZ94]:

$$\frac{\partial f}{\partial x_{j,i}} = 0 \quad \text{for all} \quad j = 1, \ldots, n \quad \text{and} \quad i = 1, \ldots, k_j. \tag{5.87}$$

The unique (up to a scale) irreducible polynomial with integral coefficients in the entries $a_{i_1,i_2,\ldots,i_n}$ of a tensor A that vanishes when f is degenerate is called the hyperdeterminant.

Example (2×2 hyperdeterminant): For the 2×2 hyperdeterminant, consider $f(X_1, X_2) = \sum_{i,j=0}^{1} a_{i,j} x_i y_j$. The multilinear form f is degenerate if there is a non-trivial solution for X_1, X_2,

$$\frac{\partial f}{\partial x_0} = a_{00}y_0 + a_{01}y_1 = 0 \tag{5.88}$$

$$\frac{\partial f}{\partial y_0} = a_{00}x_0 + a_{10}x_1 = 0 \tag{5.89}$$

$$\frac{\partial f}{\partial x_1} = a_{10}y_0 + a_{11}y_1 = 0 \tag{5.90}$$

$$\frac{\partial f}{\partial y_1} = a_{01}x_0 + a_{11}x_1 = 0. \tag{5.91}$$

Trying to solve this system of equations, we obtain that,

$$\frac{y_0}{y_1} = \frac{-a_{01}}{a_{00}} = \frac{-a_{11}}{a_{10}} \tag{5.92}$$

$$\frac{x_0}{x_1} = \frac{-a_{10}}{a_{00}} = \frac{-a_{11}}{a_{01}}. \tag{5.93}$$

We see that a non-trivial solution exists if and only if, $a_{00}a_{11} - a_{10}a_{01} = 0$, i.e., the hyperdeterminant is simply the determinant in this case.

The hyperdeterminant of a $2 \times 2 \times 2$ multilinear form was first computed by Cayley [Cay45] and is as follows:

$$\begin{aligned}
&-a_{000}^2a_{111}^2 - a_{100}^2a_{011}^2 - a_{010}^2a_{101}^2 - a_{001}^2a_{110}^2 \\
&-4a_{000}a_{110}a_{101}a_{011} - 4a_{100}a_{010}a_{001}a_{111} \\
&+2a_{000}a_{100}a_{011}a_{111} + 2a_{000}a_{010}a_{101}a_{111} \\
&+2a_{000}a_{001}a_{110}a_{111} + 2a_{100}a_{010}a_{101}a_{011} \\
&+2a_{100}a_{001}a_{110}a_{011} + 2a_{010}a_{001}a_{110}a_{101} = 0.
\end{aligned} \tag{5.94}$$

In [HS07b] it is further shown that the principal minors of an $n \times n$ symmetric matrix satisfy the $\underbrace{2 \times 2 \times \ldots \times 2}_{n \text{ times}}$ hyperdeterminant. It is thus clear that determining the entropy region of Gaussian random variables is intimately related to Cayley's hyperdeterminant.

It is with this viewpoint in mind that we study the hyperdeterminant in this section. In the next 2 subsections, first we present a new determinant form for the $2 \times 2 \times 2$ hyperdeterminant, which may be of interest since computing the hyperdeterminant of higher formats is extremely difficult and our formula may suggest a way of attacking more complicated hyperdeterminants. Next we give a novel proof of one of the a main results of [HS07b], that the principal minors of any $n \times n$ symmetric matrix satisfy the $\underbrace{2 \times 2 \times \ldots \times 2}_{n \text{ times}}$ hyperdeterminant. Our proof hinges on identifying a determinant formula for the multilinear form from which the hyperdeterminant arises.

5.4.1 A Formula for the $2 \times 2 \times 2$ Hyperdeterminant

Obtaining an explicit formula for the hyperdeterminant is not an easy task. The first nontrivial hyperdeterminant which is the $2 \times 2 \times 2$, was obtained by Cayley in 1845 [Cay45]. However, surprisingly, calculating the next hyperdeterminant which is the $2 \times 2 \times 2 \times 2$ proves to be very difficult. Until recently the only method for computing the $2 \times 2 \times 2 \times 2$ was the nested formula of Schläfli, which he obtained in 1852 [Shl52, GKZ94] and although after 150 years Luque and Thibon [LT03] expressed it in terms of the fundamental tensor invariants, the monomial expansion of this hyperdeterminant remained as a challenge. It was finally solved recently in [HSYY08] where they show that the $2 \times 2 \times 2 \times 2$ hyperdeterminant consists of 2,894,276 terms. It is interesting to mention that Cayley had a 340-term expression for the $2 \times 2 \times 2 \times 2$ hyperdeterminant which satisfies many invariance properties of the hyperdeterminant and only fails to satisfy a few extra conditions [TW09]. Therefore, as mentioned

previously, computing hyperdeterminants of different formats is generally non-trivial. In fact even Schläfli's method only works for some special hyperdeterminant formats. Moreover according to [GKZ94] it is not easy to prove directly that (5.94) vanishes if and only if (5.87) has a non-trivial solution. Here we propose a new formula for (and a method to obtain) the $2 \times 2 \times 2$ hyperdeterminant which shows this if and only if connection directly. Moreover this method might be extendable to hyperdeterminants of larger format.

Theorem 5.4.1 *(Determinant formula for $2 \times 2 \times 2$ hyperdeterminant) Define*

$$B_0 \triangleq \begin{bmatrix} a_{000} & a_{100} \\ a_{001} & a_{101} \end{bmatrix}, \; B_1 \triangleq \begin{bmatrix} a_{010} & a_{110} \\ a_{011} & a_{111} \end{bmatrix}, \; J \triangleq \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. \tag{5.95}$$

Then the $2 \times 2 \times 2$ hyperdeterminant is given by

$$det(B_0 J B_1^T - B_1 J B_0^T). \tag{5.96}$$

Proof: Let f be a multilinear form of the format $2 \times 2 \times 2$,

$$f(X, Y, Z) = \sum_{i,j,k=0}^{1} a_{ijk} x_i y_j z_k. \tag{5.97}$$

Then by the change of variables, $w_0 = x_0y_0$, $w_1 = x_1y_0$, $w_2 = x_0y_1$, $w_3 = x_1y_1$, the function f can be written as,

$$f(X,Y,Z) = \begin{pmatrix} z_0 & z_1 \end{pmatrix} \begin{pmatrix} a_{000} & a_{100} & a_{010} & a_{110} \\ a_{001} & a_{101} & a_{011} & a_{111} \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \end{pmatrix}$$

$$\triangleq Z^T \begin{pmatrix} B_0 & B_1 \end{pmatrix} W. \tag{5.98}$$

To proceed, recall from (5.87) that the hyperdeterminant of the multilinear form of the format $2 \times 2 \times 2$, vanishes if and only if there is a non-trivial solution (X, Y, Z) to the system of partial derivative equations:

$$\frac{\partial f}{\partial x_i} = 0 \qquad \frac{\partial f}{\partial y_j} = 0 \qquad \frac{\partial f}{\partial z_k} = 0 \qquad i,j,k = 0,1. \tag{5.99}$$

(a) First we show that if there is a non-trivial solution to the equations (5.99), then (5.96) vanishes. By the chain rule $\frac{\partial f}{\partial x_i} = \sum_k \frac{\partial w_k}{\partial x_i} \frac{\partial f}{\partial w_k}$, we can write $\frac{\partial f}{\partial (X,Y)} = \left(\frac{\partial W}{\partial (X,Y)}\right)^T \frac{\partial f}{\partial W}$. Also from (5.98), $\frac{\partial f}{\partial Z} = \begin{pmatrix} B_0 & B_1 \end{pmatrix} W$. Therefore the degeneracy conditions equivalent with (5.99) become:

$$\left(\frac{\partial W}{\partial (X,Y)}\right)^T \frac{\partial f}{\partial W} = 0 \tag{5.100}$$

$$\begin{pmatrix} B_0 & B_1 \end{pmatrix} W = 0. \tag{5.101}$$

Condition (5.100) implies that the vector $\frac{\partial f}{\partial W}$ should belong to the null space of $\left(\frac{\partial W}{\partial (X,Y)}\right)^T$.

The following Lemma gives the structure of this null space.

Lemma 5.4.2 *Null space of the matrix* $\left(\frac{\partial W}{\partial (X,Y)}\right)^T$ *is characterized by vectors of the form,* $\begin{pmatrix} w_3 & -w_2 & -w_1 & w_0 \end{pmatrix}^T$.

Proof: Let V be a 4×1 vector. Noting that for $j = \{1, 2\}$, $\left(\frac{\partial W}{\partial (X,Y)}\right)_{ij} = \frac{\partial w_i}{\partial x_{j-1}}$ and for $j = \{3, 4\}$, $\left(\frac{\partial W}{\partial (X,Y)}\right)_{ij} = \frac{\partial w_i}{\partial y_{j-3}}$,

$$\left(\frac{\partial W}{\partial (X,Y)}\right)^T V = \begin{pmatrix} y_0 & 0 & y_1 & 0 \\ 0 & y_0 & 0 & y_1 \\ x_0 & x_1 & 0 & 0 \\ 0 & 0 & x_0 & x_1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{pmatrix} = 0. \tag{5.102}$$

Solving for V in the above, yields the equations:

$$\frac{v_1}{v_3} = \frac{v_2}{v_4} = -\frac{y_1}{y_0} \tag{5.103}$$

$$\frac{v_1}{v_2} = \frac{v_3}{v_4} = -\frac{x_1}{x_0}. \tag{5.104}$$

Letting $v_4 = x_0 y_0$ characterizes the vectors in the null space up to a scale:

$$\begin{aligned} V^T &= \begin{pmatrix} x_1 y_1 & -x_0 y_1 & -x_1 y_0 & x_0 y_0 \end{pmatrix}^T \\ &= \begin{pmatrix} w_3 & -w_2 & -w_1 & w_0 \end{pmatrix}^T. \end{aligned} \tag{5.105}$$

Going back to the proof of Theorem 5.4.1, using Lemma 5.4.2 we conclude that we should have $\frac{\partial f}{\partial Z} = \begin{pmatrix} B_0 & B_1 \end{pmatrix} W = 0$, and for an arbitrary non-zero scalar α, $\frac{\partial f}{\partial W} = \begin{pmatrix} B_0 & B_1 \end{pmatrix}^T Z = \alpha \begin{pmatrix} w_3 & -w_2 & -w_1 & w_0 \end{pmatrix}^T$. Putting these two equations

into matrix form we can further write the following:

$$\begin{pmatrix} 0 & 0 & {B_0}^T \\ 0 & 0 & {B_1}^T \\ B_0 & B_1 & 0 \end{pmatrix} \begin{pmatrix} W \\ Z \end{pmatrix} = \alpha \begin{pmatrix} w_3 \\ -w_2 \\ -w_1 \\ w_0 \\ 0 \\ 0 \end{pmatrix} \tag{5.106}$$

or in other form:

$$\begin{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} & \alpha \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} & {B_0}^T \\ \alpha \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} & {B_1}^T \\ B_0 & B_1 & 0 \end{pmatrix} \begin{pmatrix} W \\ Z \end{pmatrix} = 0. \tag{5.107}$$

A non-trivial solution for X, Y, Z and hence for W, Z requires the matrix to be low rank. Therefore using the fact that $J^{-1} = -J$ we can write the following,

$$\det \left(\begin{pmatrix} B_0 & B_1 \end{pmatrix} \begin{pmatrix} 0 & J \\ -J & 0 \end{pmatrix} \begin{pmatrix} {B_0}^T \\ {B_1}^T \end{pmatrix} \right) = \det(B_0 J {B_1}^T - B_1 J {B_0}^T) = 0. \tag{5.108}$$

Note that the explicit calculation of (5.108) gives,

$$\det \begin{pmatrix} 2(a_{100}a_{010} - a_{000}a_{110}) & a_{100}a_{011} + a_{101}a_{010} \\ & -a_{000}a_{111} - a_{001}a_{110} \\ a_{100}a_{011} + a_{101}a_{010} & 2(a_{101}a_{011} - a_{001}a_{111}) \\ -a_{000}a_{111} - a_{001}a_{110} & \end{pmatrix} = 0 \tag{5.109}$$

which when expanded gives the $2\times2\times2$ hyperdeterminant formula stated in equation (5.94), as expected.

(b) Conversely suppose that (5.108) vanishes and therefore there is a non-trivial solution for W and Z in (5.107). To prove that there is also a non-trivial solution to (5.99), we need to show that such $X,\ Y$, and Z exist so that (5.100) and (5.101) hold. By definition of $w_0,\ldots,w_3$, it is not hard to see that a valid x_0, x_1, y_0, and y_1 can be found from w_i only if $W = (\ w_0 \quad w_1 \quad w_2 \quad w_3\ )^T$ in (5.107) has the property,

$$\frac{w_0}{w_2} = \frac{w_1}{w_3}. \tag{5.110}$$

In the following we show that the solution of (5.107) in fact satisfies relation (5.110). Let $p = \begin{pmatrix} w_0 & w_1 \end{pmatrix}^T$ and $q = \begin{pmatrix} w_2 & w_3 \end{pmatrix}^T$. Then from (5.107) we obtain:

$$\alpha Jq \ + \ {B_0}^T Z = 0 \tag{5.111}$$

$$-\alpha Jp \ + \ {B_1}^T Z = 0 \tag{5.112}$$

$$B_0 p \ + \ B_1 q = 0. \tag{5.113}$$

Multiplying the first equation by p^T and the second one by q^T and adding them together we obtain,

$$\alpha(p^T Jq - q^T Jp) + (p^T {B_0}^T + q^T {B_1}^T)Z = 0 \tag{5.114}$$

which by the use of (5.113) simplifies to:

$$p^T Jq = q^T Jp. \tag{5.115}$$

Noting that $p^T Jq = (p^T Jq)^T = -q^T Jp$ gives,

$$p^T Jq = q^T Jp = 0 \tag{5.116}$$

(5.110) then follows immediately from (5.116) by substituting for p and q.

5.4.2 Minors of a Symmetric Matrix Satisfy the Hyperdeterminant

It has recently been shown in [HS07b] that the principal minors of a symmetric matrix satisfy the hyperdeterminant relations. There this was found by either checking or explicitly computing the determinant of a 3×3 matrix in terms of the other minors and noticing that it satisfied the $2 \times 2 \times 2$ hyperdeterminant. In this section we give an explanation of why this relation holds for the principal minors of a symmetric matrix. The key ingredient is by identifying a simple determinant formula for the multilinear form (5.86) when the coefficients $a_{i_1,i_2,\ldots,i_n}$ are the minors of an $n \times n$ symmetric matrix.

Lemma 5.4.3 *Let the elements of the tensor* $A = [a_{i_1,i_2,\ldots,i_n}]$, $i_k = \{0,1\}$ *be the principal minors of an* $n \times n$ *matrix* $\tilde{A}$ *such that* $a_{i_1,i_2,\ldots,i_n}$, $i_k = \{0,1\}$ *denotes the principal minor obtained by choosing the rows and columns of* $\tilde{A}$ *indexed by the set* $\alpha = \{k | i_k = 1\}$ *(by convention when all indices are zero* $a_{00\ldots0} = 1$*). Then the following multilinear form of the format* $2 \times 2 \times \ldots \times 2$ *(n times),*

$$f(X_1, X_2, \ldots, X_n) = \sum_{i_1,i_2,\ldots,i_n=0}^{1} a_{i_1,i_2,\ldots,i_n} x_{1,i_1} x_{2,i_2} \ldots x_{n,i_n} \tag{5.117}$$

can be rewritten as the determinant of the matrix M*, i.e.,* $f(X_1, X_2, \ldots, X_n) =$

$\det(M)$ *where* M *is the following matrix:*

$$M = \begin{pmatrix} x_{1,0} & 0 & \dots & 0 \\ 0 & x_{2,0} & \dots & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \dots & x_{n,0} \end{pmatrix} + \begin{pmatrix} x_{1,1} & 0 & \dots & 0 \\ 0 & x_{2,1} & \dots & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \dots & x_{n,1} \end{pmatrix} \tilde{A} \triangleq N_1 + N_2\tilde{A}. \quad (5.118)$$

Proof: First note that determinant of M has the form,

$$\det(M) = \sum_{i_1,i_2,\dots,i_n=0}^{1} b_{i_1,i_2,\dots,i_n} x_{1,i_1} x_{2,i_2} \dots x_{n,i_n} \quad (5.119)$$

for some $\tilde{A}$-dependent coefficients $b_{i_1,i_2,\dots,i_n}$. To prove that $\det(M)$ is in fact equal to (5.117), we need to show that $b_{i_1,i_2,\dots,i_n} = a_{i_1,i_2,\dots,i_n}$, $\forall i_1, \dots i_n$, or in other words $b_{i_1,i_2,\dots,i_n}$ are the corresponding minors of $\tilde{A}$.

Let $(p_1 \dots p_n)$ be a realization of $\{0,1\}^n$. For $j = 1, \dots, n$, let the variables $x_{j,p_j} = 1$ and the rest of the variables be zero. This choice of values makes $\det(M) = b_{p_1,p_2,\dots,p_n}$ and $f(X_1, X_2, \dots, X_n) = a_{p_1,p_2,\dots,p_n}$. Moreover it can be easily seen that in this case $\det(M)$ in (5.118) will simply be equal to the minor of the matrix $\tilde{A}$ obtained by choosing the set of rows and columns $\alpha \subseteq \{1, \dots, n\}$ such that $p_j = 1$ for all $j \in \alpha$. By assumption this is nothing but the coefficient $a_{p_1,p_2,\dots,p_n}$ in (5.117) and therefore the lemma is proved. *Remark:* Note that Lemma 5.4.3 does not require the matrix $\tilde{A}$ to be symmetric.

Lemma 5.4.4 (Partial derivatives of det M) *Let* $\alpha = \{1, \dots, n\} \setminus j$. *If* $\tilde{A}$ *is nonsingular, computing the partial derivatives of the* $\det M$ *gives*

$$\frac{\partial \det M}{\partial x_{j,0}} = \det M_{\alpha,\alpha} \quad (5.120)$$

$$\frac{\partial \det M}{\partial x_{j,1}} = \det \tilde{A} \det M'_{\alpha,\alpha} \quad (5.121)$$

where $M' = N_1\tilde{A}^{-1} + N_2$.

Proof: To prove (5.120) note that we can write,

$$\frac{\partial \det M}{\partial x_{j,0}} = \sum_{k,l} \frac{\partial \det M}{\partial M_{kl}} \frac{dM_{kl}}{dx_{j,0}} = \mathrm{tr}\left(\frac{d(\det M)}{dM} \left(\frac{dM}{dx_{j,0}} \right)^T \right). \tag{5.122}$$

Since $\frac{d(\det M)}{dM} = M^{-T} \det M$ and $\frac{dM}{dx_{j,0}} = e_j$ where e_j is an $n \times n$ matrix whose jth diagonal entry is 1 and all of its other entries are 0, we can further write,

$$\frac{\partial \det M}{\partial x_{j,0}} = \det M \cdot \left(M^{-T}\right)_{jj} = \det M \cdot \left(M^{-1}\right)_{jj}. \tag{5.123}$$

Inverse of M is calculated by, $M^{-1} = \frac{\mathrm{adj} M}{\det M}$ and therefore $M^{-1}_{jj} = \frac{\det M_{\alpha,\alpha}}{\det M}$ and (5.120) follows immediately. For (5.121) note that:

$$\begin{aligned} \frac{\partial \det M}{\partial x_{j,1}} &= \frac{\partial}{\partial x_{j,1}} \det[(N_1\tilde{A}^{-1} + N_2)\tilde{A}] \\ &= \det \tilde{A} \frac{\partial \det M'}{\partial x_{j,1}}. \end{aligned} \tag{5.124}$$

Using (5.120), and the above equation, (5.121) follows immediately. Now we can write the condition for the minors of $\tilde{A}$ to satisfy the hyperdeterminant:

Lemma 5.4.5 (rank of M) *The minors of the non-singular matrix $\tilde{A}$ satisfy the hyperdeterminant equation if there exists a set of solutions $x_{j,0}$ and $x_{j,1}$ for which rank of M in (5.118) is at most $n-2$.*

Proof: To satisfy the hyperdeterminant, we require (5.120) and (5.121) to be equal to zero simultaneously for all j. If there is a non-trivial set of solutions $x_{j,0}$ and $x_{j,1}$ for which M has rank of at most $n-2$, then clearly (5.120) vanishes. Moreover if we assume that $\tilde{A}$ is non-singular then rank M' = rank $M\tilde{A}^{-1}$ = rank M, and (5.121) vanishes as well. Therefore the multilinear form (5.117) becomes degenerate which

means the coefficients $a_{i_1,i_2,\ldots,i_n}$, i.e., the principal minors of the matrix, will satisfy the hyperdeterminant.

Theorem 5.4.6 (hyperdeterminant and the principal minors) *The principal minors of an $n \times n$ symmetric matrix $\tilde{A}$ satisfy the hyperdeterminants of the format $2 \times 2 \ldots \times 2$ (k times) for all $k \leq n$.*

Proof: First we show that the minors satisfy the $2 \times 2 \ldots \times 2$ (n times) hyperdeterminant. Recall that for the tensor of coefficients $a_{i_1,i_2,\ldots,i_n}$ in the multilinear form (5.86) to satisfy the hyperdeterminant relation, there must exist a non-trivial solution to make all the partial derivatives of f with respect to its variables zero. Lemma (5.4.5) suggests that a set of non-trivial $x_{j,0}$ and $x_{j,1}$ for which rank of M is at most $n-2$ would be sufficient. To use this lemma we first assume that $\tilde{A}$ is non-singular. In the following we will show that one can always find a solution to make rank $M \leq n-2$. First we find a non-trivial solution in the case of 3 variables and then extend it to the the case where there are n variables. For 3 variables, the matrix M which is of the following form,

$$M = \begin{pmatrix} x_{1,0} + x_{1,1}a_{11} & x_{1,1}a_{12} & x_{1,1}a_{13} \\ x_{2,1}a_{12} & x_{2,0} + x_{2,1}a_{22} & x_{2,1}a_{23} \\ x_{3,1}a_{13} & x_{3,1}a_{23} & x_{3,0} + x_{3,1}a_{33} \end{pmatrix} \tag{5.125}$$

should be rank 1, or equivalently all the columns be multiples of one another. Enforcing this condition results in 3 equations for 6 unknowns. Therefore without loss of generality we let $x_{j,1} = 1$. Making the columns of M proportional, gives:

$$\frac{x_{1,0} + a_{11}}{a_{12}} = \frac{a_{12}}{x_{2,0} + a_{22}} = \frac{a_{13}}{a_{23}} \tag{5.126}$$

$$\frac{x_{3,0} + a_{33}}{a_{23}} = \frac{a_{13}}{a_{12}}. \tag{5.127}$$

If $\overline{x}_j = (x_{j,0}, x_{j,1})$, then the solution to the above equations is clearly as follows:

$$\begin{aligned}
\overline{x}_1 &= (\frac{a_{12}a_{13} - a_{11}a_{23}}{a_{23}}, 1) \\
\overline{x}_2 &= (\frac{a_{23}a_{12} - a_{13}a_{22}}{a_{13}}, 1) \\
\overline{x}_3 &= (\frac{a_{13}a_{23} - a_{12}a_{33}}{a_{12}}, 1).
\end{aligned} \tag{5.128}$$

Now for the general case of n variables, let $\overline{x}_1, \overline{x}_2, \overline{x}_3$ be as (5.128) and for $j > 3$, $\overline{x}_j = (1, 0)$. It can be easily checked that this solution makes the matrix M of rank $n-2$ and therefore the principal minors satisfy the $2 \times 2 \times \ldots \times 2$ (n times) hyperdeterminant. Note that these solutions also appear in [HS07b] in an alternative proof of principal minors satisfying the hyperdeterminant relation. Now we can easily show that the principal minors also satisfy all hyperdeterminants of format $2 \times 2 \times \ldots \times 2$ (k times) for all $3 \leq k \leq n$. In order to consider the $2 \times 2 \times \ldots \times 2$ (k times) hyperdeterminant, let $x_{j,0} = 1$ and $x_{j,1} = 0$ for all $k + 1 \leq j \leq n$ such that the multilinear form (5.117) will be in terms of only k variables. In terms of the matrix M in (5.118) one can only consider the first k rows and therefore the problem reduces to the existence of a non-trivial solution to make M of rank $k - 2$, and, as previously shown, this is always possible, and hence the principal minors satisfy any $2 \times 2 \times \ldots \times 2$ (k times) hyperdeterminant for $3 \leq k \leq n$. Finally, note that any singular matrix $\tilde{A}$ can be considered as the limit of a sequence of non-singular matrices whose principal minors satisfy the hyperdeterminant relations and therefore the principal minors of the singular matrix will do so as well. Rewriting the hyperdeterminant relation (5.94) in terms of the principal minors by adopting the notation of Lemma 5.4.3 for a 3×3

matrix gives,

$$A_{\emptyset}^2 A_{123}^2 + A_1^2 A_{23}^2 + A_2^2 A_{13}^2 + A_3^2 A_{12}^2 + 4A_{\emptyset} A_{12} A_{13} A_{23} + 4A_1 A_2 A_3 A_{123}$$
$$-2A_{\emptyset} A_1 A_{23} A_{123} - 2A_{\emptyset} A_2 A_{13} A_{123} - 2A_{\emptyset} A_3 A_{12} A_{123}$$
$$-2A_1 A_2 A_{13} A_{23} - 2A_1 A_3 A_{12} A_{23} - 2A_2 A_3 A_{12} A_{13} = 0. \quad (5.129)$$

Letting $A_{\emptyset} = 1$ this can also be written as,

$$(A_{123} - A_3 A_{12} - A_2 A_{13} - A_1 A_{23} + 2A_1 A_2 A_3)^2 \;=\; 4(A_1 A_2 - A_{12})(A_1 A_3 - A_3)(A_2 A_3 - A_{23}). \quad (5.130)$$

5.5 Minimal Conditions for Realizing a Vector with Gaussian Entropies

In order to determine whether a $2^n - 1$ dimensional vector g corresponds to the entropy of n scalar jointly Gaussian random variables, one needs to check whether e^g, i.e., the supposed vector of principal minors, corresponds, to all the principal minors of a symmetric positive semi-definite matrix. Define $A \triangleq e^g$ and let the elements of the vector $A \in \mathbb{R}^{2^n-1}$ be denoted by $A_\alpha,\ \alpha \subseteq \{1, \ldots, n\}$. An interesting problem is to find the minimal set of conditions under which the vector A can be considered as the vector of all principal minors of a symmetric $n \times n$ matrix. This problem is known as the "principal minor assignment" problem and has been addressed before in [HS07b, GT06b]. In fact in a recent remarkable work, [HS07b] gives the set of necessary and sufficient conditions for this problem. Nonetheless it does not point out the minimal set of such necessary and sufficient equations. Instead [HS07b] is mainly interested in the generators of the prime ideal of all homogenous polynomial relations among the principal minors of an $n \times n$ symmetric matrix. Here we propose

the minimal set of such conditions for $n \geq 4$.

Roughly speaking there are $2^n - 1$ variables in the vector A and only $\frac{n(n+1)}{2}$ parameters in a symmetric $n \times n$ matrix. Therefore if the elements of A can be considered as the minors of a $n \times n$ symmetric matrix, one suspects that there should be $2^n - 1 - \frac{n(n+1)}{2}$ constraints on the elements of A. These constraints, which can be translated to relations between the elements of the entropy vector arising from n scalar Gaussian random variables, can be used as the starting point to determine the entropy region of $n \geq 4$ jointly Gaussian scalar random variables.

We start this section by studying the entropy region of 4 jointly Gaussian random variables using the results of the hyperdeterminant already mentioned in the previous section, and we shall explicitly state the sufficiency of 5 constraints among all the constraints given in [HS07b] by using a similar proof to [HS07b]: that for a given vector A and under such constraints, one can construct the symmetric matrix $\tilde{A} = [a_{ij}]$ with the desired principle minors. Later in this section we state such minimal number of conditions for a $2^n - 1$ dimensional vector for $n \geq 4$.

Let

$$g_{ijk} = A_{ijk} - A_i A_{jk} - A_j A_{ik} - A_k A_{ij} + 2A_i A_j A_k. \tag{5.131}$$

Theorem 5.5.1 *The minimal set of necessary and sufficient conditions for the elements of the vector A to be the principal minors of a symmetric 4×4 matrix consists of three hyperdeterminant equations, one consistency of the signs of g_{ijk}, and the*

determinant identity of the 4×4 *matrix:*

$$g_{123}^2 = 4(A_1A_2 - A_{12})(A_2A_3 - A_{23})(A_1A_3 - A_{13}) \tag{5.132}$$

$$g_{124}^2 = 4(A_1A_2 - A_{12})(A_2A_4 - A_{24})(A_1A_4 - A_{14}) \tag{5.133}$$

$$g_{134}^2 = 4(A_1A_3 - A_{13})(A_3A_4 - A_{34})(A_1A_4 - A_{14}) \tag{5.134}$$

$$g_{123}g_{124}g_{134} = 4(A_1A_2 - A_{12})(A_1A_3 - A_{13})(A_1A_4 - A_{14})g_{234} \tag{5.135}$$

$$\begin{aligned} A_{1234} = -\frac{1}{2} \sum_{\substack{i',j' \in \{1,2,3\} \\ k',l' \in \{1,2,3,4\} \setminus \{i',j'\}}} \frac{g_{i'j'k'}g_{i'j'l'}}{A'_iA'_j - A_{i'j'}} + A_1g_{234} + A_2g_{134} + A_3g_{124} \\ + A_4g_{123} - 2A_1A_2A_3A_4 + A_{12}A_{34} + A_{13}A_{24} + A_{14}A_{23}. \end{aligned} \tag{5.136}$$

Proof: If elements of the vector A are the principal minors of a symmetric matrix they satisfy the hyperdeterminant relations. In particular we will have,

$$g_{ijk}^2 = 4(A_iA_j - A_{ij})(A_iA_k - A_{ik})(A_jA_k - A_{jk}) \tag{5.137}$$

and therefore the necessity of equations (5.132)–(5.136) is straightforward to show. By using a similar method to [HS07b] one can show the sufficiency of equations (5.132)–(5.136). To make the chapter self-contained we explain the steps in more detail. First note that all the elements of $\tilde{A}$ can be determined up to a sign from the A_i and A_{ij} elements of the vector A.

$$a_{ii} = A_i \tag{5.138}$$

$$a_{ij}^2 = a_{ii}a_{jj} - A_{ij} = A_iA_j - A_{ij} \tag{5.139}$$

It remains to choose the signs of all the off-diagonals in such a way that the 3×3 and 4×4 minors of $\tilde{A}$ will correspond to A_{ijk} and A_{1234}. First let's consider the 3×3 minors. Assuming $\tilde{A}$ to be the symmetric matrix with minors corresponding to elements of A, a direct calculation of a 3×3 principal minor with rows and columns

indexed by $\{i, j, k\}$ gives:

$$\begin{aligned} A_{ijk} =& a_{ii}a_{jj}a_{kk} - a_{ii}a_{jk}^2 - a_{jj}a_{ik}^2 - a_{kk}a_{ij}^2 + 2a_{ij}a_{jk}a_{ik} \\ =& -2A_iA_jA_k + A_iA_{jk} + A_jA_{ik} + A_kA_{ij} \\ & \pm 2\sqrt{(A_iA_j - A_{ij})(A_iA_k - A_{ik})(A_jA_k - A_{jk})} \end{aligned} \tag{5.140}$$

which can be written as:

$$\begin{aligned} g_{ijk} =& 2a_{ij}a_{jk}a_{ik} \\ =& \pm 2\sqrt{(A_iA_j - A_{ij})(A_iA_k - A_{ik})(A_jA_k - A_{jk})}. \end{aligned} \tag{5.141}$$

Note that although the sign ambiguities of the 3 off-diagonal elements in a 3×3 minor imply 8 possible matrices, the determinant of a 3×3 matrix depends only on the sign of the product of the off-diagonal terms or in other words the parity of g_{ijk}. Squaring both sides yields the hyperdeterminant relation (5.130). There are four such hyperdeterminants for a 4×4 matrix, each corresponding to a 3×3 minor,

$$g_{ijk}^2 = 4{a_{ij}}^2{a_{ik}}^2{a_{jk}}^2 \qquad i, j, k \in \{1, 2, 3, 4\}. \tag{5.142}$$

(5.142) for all permutations of $\{i, j, k\}$ assures that there is a sign choice for the four g_{ijk} such that all the A_{ijk} will correspond to the 3×3 minors of $\tilde{A}$. However what we require next is the consistency of the signs. In other words there should exist at least one sign assignment of the off-diagonal terms that results in the assumed signs

of g_{ijk}. To be more specific we have,

$$g_{123} = 2a_{12}a_{13}a_{23} \tag{5.143}$$

$$g_{124} = 2a_{12}a_{14}a_{24} \tag{5.144}$$

$$g_{134} = 2a_{13}a_{14}a_{34} \tag{5.145}$$

$$g_{234} = 2a_{23}a_{24}a_{34}. \tag{5.146}$$

Considering the first 3 equations, it is clear that one can freely choose any signs for g_{123}, g_{124}, and g_{234} by assigning signs to a_{ij}. However once these signs are fixed, the sign of g_{234} should comply with the rest. In fact multiplication of the three of g_{ijk} gives:

$$g_{ijk}g_{ijl}g_{ikl} = 4a_{ij}^2a_{ik}^2a_{il}^2g_{jkl} \tag{5.147}$$

which means, once the signs of the three out of four g_{ijk} are determined, the last one should be consistent with them through (5.147). Considering one of these equations, i.e., a particular permutation of $\{i, j, k\}$, is sufficient for our purpose,

$$g_{123}g_{124}g_{134} = 4a_{12}^2a_{13}^2a_{14}^2g_{234}. \tag{5.148}$$

It only remains to insist that the whole determinant of the constructed matrix be equal to A_{1234}. This is guaranteed through (5.136), which is obtained by direct calculation of the 4×4 determinant. In (5.136) note that since the $g_{i'j'k'}g_{i'j'l'}$ in the numerator has $A'_iA'_j - A_{i'j'}$ in it, vanishing of $A'_iA'_j - A_{i'j'}$ in the denominator will not cause any problems. Finally noting that, one hyperdeterminant equation, for example,

$$g_{234}^2 = 4(A_2A_3 - A_{23})(A_3A_4 - A_{34})(A_2A_4 - A_{24}) \tag{5.149}$$

can be obtained from the other three hyperdeterminants, i.e., (5.132), (5.133), and (5.134), and the parity consistency condition (5.148) leaves 5 equations of (5.132) to

(5.136) through which we can construct the matrix $\tilde{A}$.

Using a similar approach, which closely follows the proof methods of [HS07b], we can write the set of minimal necessary and sufficient conditions for a 2^n-1 dimensional vector to be the principal minors of a symmetric matrix.

Theorem 5.5.2 *The necessary and sufficient conditions for a 2^n-1 dimensional vector to be the principal minors of a symmetric $n\times n$ matrix consists of $2^n-1-\frac{n(n+1)}{2}$ equations, and are as follows:*

$$\forall j,k\in\{2,\ldots,n\}, g_{1jk}^2=4(A_1A_j-A_{1j})(A_1A_k-A_{1k})(A_jA_k-A_{jk}) \tag{5.150}$$

$$\forall i,j,k\in\{2,\ldots,n\}, g_{1ij}g_{1ik}g_{1jk}=4(A_1A_i-A_{1i})(A_1A_j-A_{1j})(A_1A_k-A_{1k})g_{ijk} \tag{5.151}$$

Also $\forall\beta\subseteq\{1,\ldots,n\}, |\beta|\geq 4$, choose one set of $\{i,j,k,l\}\subseteq\beta$, s.t., $i<j<k<l$, and let $\alpha=\beta\backslash\{i,j,k,l\}$,

$$D^{\alpha}_{ijkl}=0 \tag{5.152}$$

where D^{α}_{ijkl} is obtained from the following by replacing every $A_S, S\subseteq\{i,j,k,l\}$ by $\frac{A_{S\cup\alpha}}{A_\alpha}$.

$$\begin{aligned} D_{ijkl}=A_{ijkl}+\frac{1}{2}\sum_{\substack{i',j'\in\{i,j,k\}\\ k',l'\in\{i,j,k,l\}\backslash\{i',j'\}}}\frac{g_{i'j'k'}g_{i'j'l'}}{A'_iA'_j-A_{i'j'}}-A_ig_{jkl}-A_jg_{ikl}-A_kg_{ijl}\\ -A_lg_{ijk}+2A_iA_jA_kA_l+A_{ij}A_{kl}-A_{ik}A_{jl}-A_{il}A_{jk}=0 \end{aligned} \tag{5.153}$$

Proof: The proof is essentially the same as the proof technique of [HS07b], and is a generalization of Theorem 5.5.1 for a 15-dimensional vector. However, we would like to highlight why this set is the minimal set of necessary, and sufficient conditions among all conditions given in [HS07b]. As mentioned in Theorem 5.5.1 one can obtain all the off diagonal entries up to a sign. Moreover, by a similarity transformation by

a diagonal ± 1 matrix, one can make all the entries of the first row positive. Then the signs of the rest of the off-diagonal entries can be fixed by using the fact that $g_{1jk} = 2a_{1j}a_{1k}a_{jk}$, provided that the conditions (5.150) are met. Now we need to enforce that the hyperdeterminants g_{ijk} for which $i, j, k \neq 1$ also hold. However since all the g_{1jk} are already determined, g_{ijk} is also essentially determined and must obey (5.151), which guarantees $g_{ijk} = 2a_{ij}a_{ik}a_{jk}$. Up to now all the minors of up to size 3 are considered. Note that $D_{ijkl} = 0$ is simply obtained from the determinant of the 4×4 submatrix with rows and columns indexed by $\{i, j, k, l\}$ (compare with (5.136)). For any submatrix with rows and columns indexed by β, say $\tilde{A}_\beta$, we can write this determinant formula for the Schur complement of the $\tilde{A}_\alpha$ in $\tilde{A}_\beta$, which essentially gives (5.152). Note that all we need is that the minors of size greater than or equal to 4 be consistent with the already defined matrix entries, and (5.152) takes care of this since all these minors appear linearly. Moreover, for each β only 1 equation of type (5.152) is required. Finally, note that there are $\binom{n-1}{2}$ number of equations of type (5.150), $\binom{n-1}{3}$ of type (5.151), and $\sum_{m=4}^{n}\binom{n}{m}$ of type (5.152), which sums up to $2^n - 1 - \frac{n(n+1)}{2}$. This is the number that we expect, noting that there are only $\frac{n(n+1)}{2}$ free parameters in a symmetric matrix while the given vector of principal minors is of size $2^n - 1$. $\square$

Note that if we insist that for all $\alpha \subseteq \{1, \ldots, n\}$, $A_\alpha \geq 0$ and substitute each A_α by e^{g_α} in (5.150)–(5.152), then (5.150)–(5.152) give the necessary and sufficient conditions for a $2^n - 1$ dimensional vector to correspond to the entropies of n scalar jointly Gaussian random variables. Nonetheless in order to characterize the entropy region of scalar Gaussian random variables, what one really needs is the convex hull of all such entropy vectors.

In an algebraic geometry language, one can define the "amoeba" of a polynomial f where $f(x_1, \ldots, x_k) = \sum_i q_i x_1^{p_{1i}} \ldots x_k^{p_{ki}}$ as the image of $f = 0$ in $\mathbb{R}^k$ under the mapping that acts on $(x_1, \ldots, x_k)$ as $(x_1, \ldots, x_k) \mapsto (\log|x_1|, \ldots, \log|x_k|)$ [GKZ94]. It turns

out that many properties of amoebas can be deduced from the Newton polytope of f, which is defined as the convex hull of the exponent vectors $(p_{1i}, \ldots, p_{ki})$ in $\mathbb{R}^k$ (see, e.g., [PR04]). In terms of our problem of interest, the scalar Gaussian entropy points are the intersection of the amoebas associated to polynomials (5.150)–(5.152) and one should look for the convex hull of the locus of these intersection points. If we allow the notion of amoeba to be defined as the log mapping for any function (not just polynomials), then one could also formulate our problem of interest as the convex hull of the amoeba of the algebraic variety obtained from the intersection of (5.150)–(5.152).

Finally we mention that in general, to characterize the entropy region of Gaussian random variables, one should consider vector-valued random variables, which are probably more complex than the case of scalars. In Section 5.3 we showed that for $n = 3$ the vector-valued random variables do not result in a bigger region than the convex hull of scalar ones. However in general it is not known whether the entropy region of n vector-valued jointly Gaussian random variables is greater than the convex hull of the entropy region of scalar valued Gaussians.

5.6 Entropy Region and Wireless Networks

Studying the entropy region of continuous random variables is especially interesting in the context of wireless networks. However as was explained in Chapter 2, due to the broadcast and interference nature of wireless channels, one needs to determine the channel-constrained entropic region. Since in the event of interference it is usually the sum of the incoming signals, possibly plus noise, that is received, studying the information inequalities which involve sums of random variables is particularly important.

As the simplest case in this section, we consider three continuous random variables

x, y, and $z = x + y$, where x and y are independent. For such random variables it is well known that the differential entropy of z is lower bounded in terms of the entropies of x and y through the entropy-power inequality, a.k.a., *EPI*. There exist several proofs of the EPI, e.g., based on the de Bruijn identity [Sta59, Bla65], or on the Brunn-Minkowski theorem [DCT91, CC84, CT91], or via MMSE [VG06, GSV06].

To make all this more precise let $x, y \in \mathbb{R}^m$ be two independent vector-valued continuous random variables and let $z = x + y$. The entropy power inequality states that the entropy of the sum, i.e., $H(z)$ has a lower bound given by,

$$e^{\frac{2}{m}H(z)} \geq e^{\frac{2}{m}H(x)} + e^{\frac{2}{m}H(y)}. \tag{5.154}$$

However, based on arguments of Chapter 2, the quantities $h_x = \frac{1}{m}H(x)$, $h_y = \frac{1}{m}H(y)$, and $h_z = \frac{1}{m}H(z)$ are simply the *normalized* entropies. Recalling from Chapter 2, that the definition of normalized entropy is more natural for network information theory (since it represents the entropy "per channel use"), it can be seen that the EPI is also more naturally expressed in terms of normalized entropy, since we can simply write

$$e^{2h_z} \geq e^{2h_x} + e^{2h_y}. \tag{5.155}$$

The EPI has found many applications in information theory, e.g., channels with non-Gaussian noise [Sha48], scalar broadcast channels [Ber74], MIMO broadcast channles [WSS06], the Gaussian wiretap channel [LYCH78], and many others. The EPI was originally stated in Shannon's seminal 1948 paper and a variational "proof", based on minimizing $H(x+y)$, subject to fixed $H(x)$ and $H(y)$, was presented. However, Shannon's proof was incomplete and only considered sufficiency.[1] The first

[1]Shannon's idea was to find the first order, i.e., KKT, conditions for the optimal distributions minimizing the constrained optimization problem. He then showed that Gaussian distributions satisfy the first-order condition. However, since the original problem of minimizing $H(x+y)$, subject to fixed $H(x)$ and $H(y)$, is nonconvex over the underlying distributions, Shannon would have further needed to show that either the KKT conditions have no other solution, or that all other solutions

complete proof was given by Stam in 1959 [Sta59] and used a different approach (de Bruijn's identity and Fisher information) [CT91]. In general, it is an interesting question to determine the relations between the entropy-powers of sums of a collection of independent random variables [MT10].

Now we consider the issue of determining the entropy region of three random variables x, y, and $z = x + y$, where x and y are independent.

In particular, we show that the seven-dimensional vector of normalized entropies and joint entropies $[h_x\ h_y\ h_z\ h_{xy}\ h_{yz}\ h_{zx}\ h_{xyz}]$ satisfies

$$\left[\ h_x,\quad h_y,\quad h_z \geq \tfrac{1}{2}\log(e^{2h_x}+e^{2h_y}),\quad h_x+h_y,\quad h_x+h_y,\quad h_x+h_y,\quad -\infty\ \right]. \tag{5.156}$$

In other words, all entropies h_x, h_y, and h_z, satisfying the EPI, are achievable.

5.6.1 The Entropy Region of x, y, and $z = x + y$

Let $x, y, z \in \mathbb{R}^m$ be 3 vector valued continuously distributed random variables such that x and y are independent and $z = x + y$. Furthermore let h represent their corresponding normalized differential entropy vector. An interesting question is to characterize the entropy region of these 3 variables, i.e., to characterize all 7-dimensional vectors that can arise as the entropy vector of such 3 variables.

Clearly h belongs to the entropy region of 3 arbitrary distributed continuous random variables which assume we denote by Γ_3^c. Thus, $h \in \Gamma_3^c$, and therefore satisfies all the Shannon inequalities. Moreover from the entropy-power inequality(EPI)[CT91], we know that the EPI for normalized entropy is,

$$e^{2h_z} \geq e^{2h_x} + e^{2h_y}. \tag{5.157}$$

are not minimizers of the cost.

Therefore if we denote the entropy region of $x,\ y$, and z by Ψ_3, then,

$$\Psi_3 \subseteq \Gamma_3^c \cap \Upsilon \cap \Xi \tag{5.158}$$

where,

$$\Upsilon = \{h | h_{xy} = h_{xz} = h_{yz} = h_x + h_y,\ h_{xyz} = -\infty\}, \tag{5.159}$$

$$\Xi = \{h | h_z \geq \frac{1}{2}\log(e^{2h_x} + e^{2h_y})\}. \tag{5.160}$$

An interesting observation is the following:

Lemma 5.6.1 (Convexity of the set defined by the EPI) *The set of entropy vectors* $\Xi = \{h | h_z \geq \frac{1}{2}\log(e^{2h_x} + e^{2h_y})\}$ *is convex.*

Proof: Convexity in h_z is obvious. Convexity in h_x and h_y follows from the (readily-verified) fact that the function $\log(e^{2h_x} + e^{2h_y})$ is convex in these variables. □

In what follows we will show that Ψ_3 can be completely characterized.

Theorem 5.6.2 (Entropy region of x, y, and $x+y$) *If* $x, y \in \mathbb{R}^m$ *are two independent, vector-valued continuous random variables and* $z = x + y$*, the entropy region of* $x,\ y,$ *and* z*, i.e.,* Ψ_3*, is*

$$\Psi_3 = \Gamma_3^c \cap \Upsilon \cap \Xi. \tag{5.161}$$

Proof: From (5.158), we know that $\Gamma_3^c \cap \Upsilon \cap \Xi$ is an outer bound. Therefore in order to prove the tightness of this outer bound we need to show that the points in the right-hand side of (5.161) are all achievable. To do this, we shall show that, for any fixed h_x and h_y, the value of h_z can grow unbounded. Since the lower bound on the EPI can be achieved by Gaussians with proportional covariance matrix, the convexity

of Ξ (established in Lemma 5.6.1) implies that all points in the set defined by the EPI are achievable.

We therefore focus on showing that for any fixed finite h_x and h_y, h_z can grow unbounded. Let x and y be two independent Gaussian random variables, $\mathcal{N}(0, \epsilon I_m + \sigma_x^2 U_x U_x^T)$ and $\mathcal{N}(0, \epsilon I_m + \sigma_y^2 U_y U_y^T)$, respectively, where U_x, U_y are $m \times m/2$ unitary matrices orthogonal to each other, i.e., $U_x^T U_x = U_y^T U_y = I_{m/2}$ and $U_x^T U_y = 0$.

Calculating the normalized entropy of a Gaussian gives,

$$h_x = \frac{1}{2}\log 2\pi e + \frac{1}{4}\log \epsilon(\epsilon + \sigma_x^2), \tag{5.162}$$

$$h_y = \frac{1}{2}\log 2\pi e + \frac{1}{4}\log \epsilon(\epsilon + \sigma_y^2). \tag{5.163}$$

On the other hand z is also a Gaussian, $\mathcal{N}(0, 2\epsilon I_m + \sigma_x^2 U_x U_x^T + \sigma_y^2 U_y U_y^T)$, for which calculating the normalized entropy gives,

$$h_z = \frac{1}{2}\log 2\pi e + \frac{1}{4}\log(2\epsilon + \sigma_x^2)(2\epsilon + \sigma_y^2). \tag{5.164}$$

The orthogonality of U_x and U_y is critical in the above calculation.

Now note that if, in particular, we choose $\sigma_x^2 = \frac{e^{4c_x}}{(2\pi e)^2\epsilon} - \epsilon$ and $\sigma_y^2 = \frac{e^{4c_y}}{(2\pi e)^2\epsilon} - \epsilon$ for some finite positive c_x and c_y and let $\epsilon \to 0$, we obtain,

$$h_x = c_x, \quad h_y = c_y, \quad h_z \to \infty. \tag{5.165}$$

Therefore while x and y have finite entropy, the entropy of their sum can become arbitrarily large. □

We should remark that while aligned covariances for x and y result in tightness of the EPI, the above proof shows that the orthogonality structure of U_x and U_y helped in making h_z arbitrary big. This resonates with the Brunn-Minkowski viewpoint of the EPI [CT91].

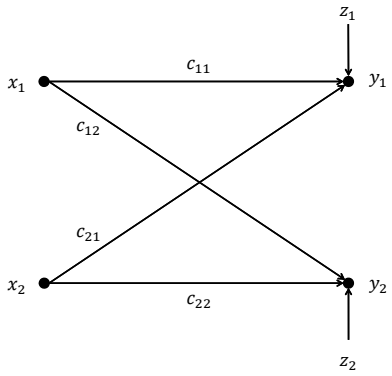


Figure 5.2: Interference channel

5.6.2 A Case Study: The Interference Channel

In this section, we show how by performing an optimization over the polymatroid region, one can obtain outer bounds for certain networks. In particular, we consider the Gaussian interference channel and obtain the outer bound of [ETW08] through a duality argument.

Consider the Gaussian interference channel of Fig. 5.2, where we are interested in optimizing the sum rate $I(x_1; y_1) + I(x_2; y_2)$. The received signals y_1 and y_2 can be described by the following equations:

$$y_1 = c_{11}x_1 + c_{21}x_2 + z_1 \tag{5.166}$$

$$y_2 = c_{12}x_1 + c_{22}x_2 + z_2, \tag{5.167}$$

where z_1 and z_2 are independent, zero-mean, complex Gaussian random variables $\mathcal{CN}(0, N_0)$ and x_1 and x_2 are power constrained by P_1 and P_2. To maximize the sum rate, based on discussions of chapter 2, we should solve an optimization problem over the entropy region of $x_1, x_2, y_1, y_2, z_1, z_2$. However, if we are interested in an outer bound, we can perform the optimization over the polymatroid region (which is known), and use some auxiliary random variables as well. Define the following

auxiliary random variables

$$s_1 = y_2|x_2 \tag{5.168}$$

$$s_2 = y_1|x_1, \tag{5.169}$$

which are the parallels of the side-information defined in [ETW08]. Through these definitions, and using the properties of z_1 and z_2, we can write the following constraints for joint entropies,

$$h_{s_1,x_1} - h_{x_1} = h_{x_1,x_2,y_2} - h_{x_1,x_2} = h_{z_2} = \log(\pi e N_0) \tag{5.170}$$

$$h_{s_2,x_2} - h_{x_2} = h_{x_1,x_2,y_1} - h_{x_1,x_2} = h_{z_1} = \log(\pi e N_0) \tag{5.171}$$

These conditions simplify to

$$h_{x_1,x_2,y_2} - h_{x_1,x_2} - \log(\pi e N_0) = 0 \tag{5.172}$$

$$h_{x_1,x_2,y_1} - h_{x_1,x_2} - \log(\pi e N_0) = 0 \tag{5.173}$$

Furthermore, for $y_1|(s_1, x_1)$, we have $h_{y_1|(s_1,x_1)} = h_{y_1|(z_2,x_1)} = h_{y_1|x_1}$. Therefore,

$$h_{y_1|(s_1,x_1)} = h_{y_1,s_1,x_1} - h_{s_1,x_1} = h_{x_1,y_1} + h_{x_1x_2y_1y_2} - h_{x_1x_2y_1} - h_{x_1} - h_{y_2x_1x_2} + h_{x_1x_2} = h_{x_1y_1} - h_{x_1} \tag{5.174}$$

which also simplifies to

$$h_{x_1x_2y_1y_2} - h_{x_1x_2y_1} - h_{y_2x_1x_2} + h_{x_1x_2} = 0 \tag{5.175}$$

Likewise for $y_2|(s_2, x_2)$ we can write

$$h_{x_1x_2y_1y_2} - h_{x_1x_2y_2} - h_{y_1x_1x_2} + h_{x_1x_2} = 0 \tag{5.176}$$

Now, for $h_{y_1|s_1}$, we have

$$h_{y_1|s_1} = h_{y_1,s_1} - h_{s_1} = h_{x_2y_1y_2} + h_{y_1} + h_{x_2} - h_{x_2y_1} - h_{x_2y_2} \tag{5.177}$$

Assume $u_1 = (y_1|s_1)$, then

$$E\left[\mathrm{Var}(u_1)\right] = E|y_1|^2 - \frac{E(y_1s_1^*)E(s_1y_1^*)}{E|s_1|^2} = N_0\left(1 + \frac{|c_{21}|^2P_2}{N_0} + \frac{|c_{11}|^2P_1/N_0}{1+|c_{12}|^2P_1/N_0}\right) \tag{5.178}$$

If we let $K_1 \triangleq 1 + \frac{|c_{21}|^2P_2}{N_0} + \frac{|c_{11}|^2P_1/N_0}{1+|c_{12}|^2P_1/N_0}$, then

$$h_{y_1|s_1} = h_{x_2y_1y_2} + h_{y_1} + h_{x_2} - h_{x_2y_1} - h_{x_2y_2} \leq \log(\pi e N_0 K_1) \tag{5.179}$$

Repeating the same process for $y_2|s_2$ and defining $K_2 \triangleq 1 + \frac{|c_{12}|^2P_1}{N_0} + \frac{|c_{22}|^2P_2/N_0}{1+|c_{21}|^2P_2/N_0}$, we obtain

$$h_{y_2|s_2} = h_{x_1y_1y_2} + h_{y_2} + h_{x_1} - h_{x_1y_2} - h_{x_1y_1} \leq \log(\pi e N_0 K_2) \tag{5.180}$$

Now we should optimize the following objective function:

$$\max I(x_1; y_1) + I(x_2; y_2) \tag{5.181}$$

subject to $h \in \Gamma_8$, and the constraints (5.172)–(5.173), (5.175)–(5.176), and (5.179)–(5.180), where h is the entropy vector of the random variables $x_1, x_2, y_1, y_2, z_1, z_2, s_1, s_2$, and Γ_8 is the polymatroid region of 8 variables.

Assume we denote the equality constraints of equations (5.172),(5.173),(5.175), and (5.176) by $g_1 = 0, \ldots, g_4 = 0$ respectively, and the inequality constraints (5.179)–(5.180) by $f_1 \leq 0$ and $f_2 \leq 0$, correspondingly. If we denote the rate of transmitter i

by R_i then we can use the duality argument of convex optimization to write

$$\begin{aligned} R_1 + R_2 &\leq \max_{h\in\Gamma_8, f_i\leq 0, g_i=0} I(x_1;y_1) + I(x_2;y_2) \\ &= \max_{h\in\Gamma_8} \min_{\lambda_i\geq 0,\mu_i} I(x_1;y_1) + I(x_2;y_2) - \sum_i \lambda_i f_i + \sum_i \mu_i g_i \\ &= \min_{\lambda_i\geq 0,\mu_i} \max_{h\in\Gamma_8} I(x_1;y_1) + I(x_2;y_2) - \sum_i \lambda_i f_i + \sum_i \mu_i g_i \\ &\leq \max_{h\in\Gamma_8} I(x_1;y_1) + I(x_2;y_2) - \sum_i \tilde{\lambda}_i f_i + \sum_i \tilde{\mu}_i g_i \end{aligned} \tag{5.182}$$

where in the last inequality, $\tilde{\lambda}_i$ and $\tilde{\mu}_i$, denote a particular choice for $\lambda_i \geq 0$ and μ_i. In fact, if we choose $\tilde{\lambda}_1 = \tilde{\lambda}_2 = 1$, and $\tilde{\mu}_i = 1, i = 1, \ldots, 4$, we obtain

$$\begin{aligned} R_1 + R_2 \leq \max_{h\in\Gamma_8} &(\log K_1 + \log K_2 - (h_{x_2y_1y_2} + h_{x_1x_2y_1} - h_{x_2y_1} - h_{x_1x_2y_1y_2}) \\ &-(h_{x_1y_1y_2} + h_{x_1x_2y_2} - h_{x_1y_2} - h_{x_1x_2y_1y_2})) \end{aligned} \tag{5.183}$$

However,

$$h_{x_2y_1y_2} + h_{x_1x_2y_1} - h_{x_2y_1} - h_{x_1x_2y_1y_2} \geq 0 \tag{5.184}$$

$$h_{x_1y_1y_2} + h_{x_1x_2y_2} - h_{x_1y_2} - h_{x_1x_2y_1y_2} \geq 0, \tag{5.185}$$

as they are polymatroid inequalities. Hence,

$$R_1 + R_2 \leq \log K_1 + \log K_2, \tag{5.186}$$

which if we replace for K_1 and K_2 yields the upper bound of [ETW08].

5.7 Conclusions

In this chapter, we studied the entropy region of jointly Gaussian random variables as an interesting subclass of continuous random variables. In particular we characterized

the region for $n \leq 3$, and for $n \geq 4$ we explicitly stated the set of $2^n - 1 - \frac{n(n+1)}{2}$ constraints that an entropy vector (equivalently the vector of principal minors) should satisfy in order to correspond to the entropy vector of n scalar jointly Gaussian random variables. These relations are intimately related to the Cayley's hyperdeterminant formula. Therefore with this viewpoint we also examined the hyperdeterminant relations. In particular, by giving a determinant formula for a multilinear form, we gave a transparent proof that the hyperdeterminant relation is satisfied by the principal minors of an $n \times n$ symmetric matrix. Moreover we also obtained a determinant form for the $2 \times 2 \times 2$ hyperdeterminant which might be extendible to higher-order formats and is an interesting problem even on its own.

We also considered the entropy region of continuous random variables in the context of wireless networks and argued that in such cases the information inequalities involving sums of random variables, such as the well-known entropy power inequality (EPI), are important. We then studied the entropy region of x, y, and $z = x + y$, where x and y are independent, as the simplest case and showed that all the entropy vectors of 3 random variables which satisfy EPI are achievable by such x, y, and z. Finally, as a particular example of a wireless network, we considered the interference channel, and obtained the capacity outer bound of [ETW08] through the entropy optimization framework.

Chapter 6

Entropy Optimization and Nonlinear Network Coding via MCMC

6.1 Introduction

Although determining the space of entropic vectors for n random variables, denoted by Γ_n^*, is crucial for solving a large class of network information theory problems, there has been scant progress in explicitly characterizing Γ_n^* for $n \geq 4$. Since the goal is most often to perform optimization over $\bar{\Gamma}_n^*$ (to solve a network information theory problem, say), in the absence of an explicit characterization of the entropy region, the next best thing is to present a method to *numerically* perform optimization over this region. Presenting such a numerical framework is the goal of the current chapter.

The approach we shall take is via a design of a random walk over probability distributions, and, in particular, over the class of quasi-uniform distributions. It is well known that the class of quasi-uniform distributions is sufficient to approximate the entropic region to any fidelity. The random walk over this characterization of distributions, when coupled with a suitable Monte Carlo Markov Chain (MCMC) method, allows for optimization of any function of the entropy vector. As an example, we apply this method to maximize the Ingleton violation for entropy vectors where the results are very encouraging. Furthermore, we show how the MCMC method can

be used as a framework to design optimal nonlinear network codes in a distributed fashion via performing a random walk over certain truth tables. Moreover we show how this method may also be used to find linear representations for matroids. We demonstrate the efficacy of the method by looking at many different examples: maximizing capacity of the Vamos, Fano, non-Fano, and M networks, and the exact repair problem in (4,2) and (5,3) settings. We also apply the method to the non-Pappus and the U_{24} matroids and show how (multi-)linear representations can be easily found for them.

6.2 Entropy Vectors and Quasi-Uniform Distributions

At first sight, the difficulty in characterizing $\bar{\Gamma}_n^*$ appears to be that one must consider *all* possible joint distributions of n random variables for *all* alphabet sizes. However, recall from Chapter 3 that the class of quasi-uniform distributions are sufficient for characterizing the whole entropy region. A distribution is called *quasi-uniform* [Cha01] if its probability mass function, as well as the probability mass function of all its marginals, takes on a constant or zero value on all points in the sample space. An example of a quasi-uniform distribution in two variables is given in Fig. 6.1, where each "x" means that a constant nonzero probability of $\frac{1}{24}$ is assigned to that point in the sample space. As can be seen, one marginal is uniform with probability $\frac{1}{8}$ and the other is quasi-uniform with probabilities 0 and $\frac{1}{6}$.

Let Λ_n denote the space of entropy vectors generated by quasi-uniform distributions. We already saw the following result in Chapter 3, Theorem 3.2.3. Here we state it again as it is relevant:

Theorem 6.2.1 (Quasi-Uniform Distributions) *[Cha01] The closure of the cone of Λ_n is the closure of Γ_n^*.*

In Chapter 3 we saw a sketch of the proof of this theorem via the concept of finite groups. However this result can also be motivated by recourse to the concept of "strong typical sequences". To this end, make T independent copies of each of our random variables, to get vector-valued sequences of length T. As $T \to \infty$, with probability approaching one, we will only encounter typical sequences. If we assign a constant probability to all typical sequences, and zero probability to non-typical ones, it is straightforward to see that we end up with a quasi-uniform distribution with the same entropy vector. The entropy is simply the log of the number of typical sequences divided by T, thus for the joint entropy of a set of random variables $(X_i, i \in \alpha)$, $\alpha \subseteq \{1, \ldots, n\}$, and alphabet size N, we have

$$h_\alpha \simeq \frac{1}{T} \log \frac{T!}{\prod_{x_\alpha} T_{x_\alpha}!}, \qquad T_{x_\alpha} = T \cdot p(X_\alpha = x_\alpha), \qquad \sum_{x_\alpha} T_{x_\alpha} = T. \tag{6.1}$$

In fact, this is essentially the statistical physics interpretation of entropy. However, (6.1) can also be interpreted in terms of subgroups of the permutation group on T elements. $T!$ is simply the size of the permutation group, whereas if we partition the T elements into $N^{|\alpha|}$ disjoint sets of sizes T_{x_α}, respectively, then $\prod_{x_\alpha} T_{x_\alpha}!$ is simply the size of the subgroup of permutations that respects this partition. Therefore as was seen in Chapter 3, Theorem 3.2.3, this again leads to a connection between entropy and groups.

Although, based on Theorem 6.2.1, determining all the quasi-uniform distributions

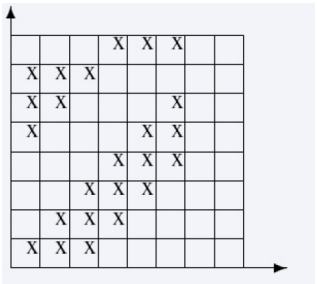


Figure 6.1: An example of a quasi-uniform distribution

is equivalent to characterizing Γ_n^*, it appears that determining all quasi-uniforms is a hard combinatorial problem. In the next section, we shall use (6.1) to characterize all possible entropy vectors of quasi-uniform distributions and to propose a random walk over them.

6.3 Entropy Optimization

6.3.1 A Characterization of Quasi-Uniform Distributions

As mentioned in the previous section, determining all quasi-uniform distributions seems to be a hard combinatorial problem. An idea to tackle this problem is to be able to sample from the space of such distributions by designing a random walk on them. However in order to do so, we need a method that 1) determines how to move from any quasi-uniform to any other such distribution, therefore defining an irreducible Markov chain, and 2) exhausts all quasi-uniforms. Working with distribution tables, like the one in Fig. 6.1, quickly reveals that devising a method to move from one quasi-uniform distribution to another is highly non-trivial. On the other hand, given that any entropy can be approximated by (6.1), one can characterize the entropies of quasi-uniform distributions (by characterizing all possible partitions and joint partitions of T elements), and then perform a random walk on the entropy vectors. The idea is as follows:

Let n be the number of random variables. Choose values T and N and construct a $T \times n$ table with entries drawn from the set $\{0, 1, \ldots, N-1\}$. Each column of the table corresponds to one of the random variables and induces a partition of T elements into at most N disjoint sets, if we let the entries with the same value belong to the same partition. The entropy of the corresponding random variable is simply computed from (6.1) using this induced partition. Similarly for $\alpha \subseteq \{1, \ldots, n\}$, any $|\alpha|$-tuple of columns defines a partition of the T elements into at most $N^{|\alpha|}$ disjoint sets (identical

rows belong to the same partition). Again, the joint entropy of the corresponding collection of random variables is computed from (6.1) using this induced partition [CY02].

Example: Consider the following table of size $T = 5$ by $n = 2$ with $N = 2$:

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

The partitions for the first column will be $T_0 = |\{0, 0\}| = 2$ and $T_1 = |\{1, 1, 1\}| = 3$, whose corresponding quasi-uniform entropy will be $h_1 = \log_2 \frac{5!}{2!3!} = \log_2 10$. For the second column the partitions are similarly $T_0 = |\{0, 0\}| = 2$ and $T_1 = |\{1, 1, 1\}| = 3$, giving $h_2 = \log_2 \frac{5!}{2!3!} = \log_2 10$. And finally for both columns the partitions are $T_{00} = |\{(0, 1), (0, 1)\}| = 2$, $T_{10} = |\{(1, 0), (1, 0)\}| = 2$, and $T_{11} = |\{(1, 1)\}| = 1$, and clearly $T_{00} = 0$, resulting in $h_{12} = \log_2 \frac{5!}{2!2!1!} = \log_2 30$.

Lemma 6.3.1 *Every such $T \times n$ table corresponds to a quasi-uniform distribution. Furthermore, as T and N grow, we encounter the set of all quasi-uniform distributions over n variables that are sufficient to characterize Γ_n^*.*

Proof: Assume that the given table corresponds to T independent copies of n random variables X_i with alphabet size N. Then from the above definition of partition on T elements and also (6.1), it is clear that we can assign a permutation group G on T elements and define its subgroups G_i as the ones that permute within each partition. It is then straightforward to generate quasi-uniform distributions from groups. In fact for a group G and its subgroups $G_1, \ldots, G_n$, define new random variables $\tilde{X}_i,\ i = 1, \ldots .n$ with alphabet sizes $\frac{|G|}{|G_i|}$ each, i.e., the number of cosets

induced by each G_i. For each element $g \in G$, obtain an n-dimensional vector v whose i-th component is the index of the coset induced by G_i that g belongs to. Assign a constant probability to $P_{\tilde{X}_1,\ldots,\tilde{X}_n}(v)$ for every vector v in the sample space encountered in this fashion, and assign zero probability to all other vectors in the sample space. It is not too difficult to see that the resulting distribution on $\tilde{X}_i$ is quasi-uniform, whose joint entropy $h_{\tilde{X}_\alpha} = h(\tilde{X}_i, i \in \alpha)$ for $\alpha \subseteq \{1,\ldots,n\}$ is obtained from $\log \frac{|G|}{|\cap_{i\in\alpha} G_i|} = \log \frac{T!}{\prod_{x_\alpha} T_{x_\alpha}!}$. Therefore to every $T \times n$ table, we can assign a quasi-uniform distribution. Moreover as N and T grow, we allow for all alphabet sizes of distributions and also make the approximation (6.1) more precise, which means that we will asymptotically encounter the set of all the quasi-uniform distributions over n random variables that are sufficient for characterizing Γ_n^*. □

Remark: Note that an alternative way of obtaining entropy vectors from a generated $T \times n$ table is to view it as the empirical distribution of the variables $X_1,\ldots,X_n$, in which case we simply have $h_{X_\alpha} = -\sum_{x_\alpha} \frac{T_{x_\alpha}}{T} \log \frac{T_{x_\alpha}}{T}$. Therefore from every table we can obtain two entropy vectors; h_X from the empirical distribution on X_i and $h_{\tilde{X}}$ from the associated quasi-uniform distributions of $\tilde{X}_i$. Note that in the limit when $T \to \infty$, approximation (6.1) becomes exact and, as described in Lemma 6.3.1, we will have $h_X = \frac{1}{T} h_{\tilde{X}}$.

For fixed N and T, the space of such obtained quasi-uniform or empirical distributions is connected. In other words one can move from a quasi-uniform/empirical distribution, corresponding to a table A, to another quasi-uniform/empirical distribution, corresponding to a table B, by a chain of changes in the entries of the table that transforms table A to table B. We can thus perform a random walk over the distributions by randomly choosing an entry of the $T \times n$ table and randomly changing its value. In this manner we can numerically stake out the entropic region.

Of course, to numerically stake out the entropy region with higher and higher fidelity requires one to increase the values of T and N. This results in an increase

in the size of the search space and slows down the MCMC methods we describe next. Thus, there is a trade-off between the quality of the results and the speed of the optimization program. Choosing the right T and N may therefore be of critical importance.

6.3.2 Entropy Optimization via Markov Chain Monte Carlo

Assume that we have the following optimization problem,

$$\max_{h \in \Gamma_n^*} f(h), \tag{6.2}$$

where $f(\cdot)$ is some function of the entropy vector. As mentioned earlier, an idea to perform this optimization numerically is to use Monte Carlo methods to sample the entropy region, or, equivalently, the space of distribution tables. Assuming each distribution table as a state S, this means that one needs to sample from this state space according to some probability distribution π. Markov Chain Monte Carlo methods are usually used for this purpose in which by designing a Markov Chain with a stationary distribution π on the system's state space, and then simulating the Markov Chain for a long time (such that the chain has converged), one can sample from π. To design a Markov Chain one needs to define a local move in the state space and the probability of moving (transition) from one state to another. Following the arguments of the last section, we can easily define a local move on the distribution tables (and hence on the entropy space) of n random variables for fixed T and N. To do so, we first generate a $T \times n$ table[1] either randomly, or by initializing it to some desired value. Then the local move would simply consist of choosing an entry of the table at random and changing its value to any other of the $N-1$ possible choices randomly. If we accept each new move with probability $\frac{1}{2}$, then this would

[1]In general we can assume that the random variables are vector valued of size l, in which case we should replace n with nl everywhere.

amount to a random walk on the space of distribution tables which is equivalent to sampling from a uniform distribution π. However a pure random walk explores the state space very slowly and is not an efficient method for performing the optimization (6.2). Denoting the cost associated with entropy vector h_S of the distribution table S by c_S, i.e., $c_S = f(h_S)$, then a standard technique to do the optimization is to set the target distribution π as,

$$\pi(S) = \frac{e^{\theta c_S}}{\sum_{S'} e^{\theta c_{S'}}} \tag{6.3}$$

where θ is a parameter usually called the temperature. Note that by tuning θ one can somewhat control the highs and lows of the distribution. In particular a distribution with large θ would favor states with higher costs and a small θ would make the distribution close to uniform. To sample from this distribution we choose a variant of the Metropolis algorithm [1]. One can consider different choices of transition probabilities. Here we accept each move with probability,

$$a = \frac{\pi(S')}{\pi(S) + \pi(S')} = \frac{e^{\theta c_{S'}}}{e^{\theta c_S} + e^{\theta c_{S'}}}. \tag{6.4}$$

Taking into account the probability of choosing an entry of the table as the result of the local move yields the transition probability $p_{S'S} = \frac{1}{nT(N-1)}a$. Note that the choice of acceptance ratio (6.4) is not as common as $\min\left(1, \frac{\pi(S')}{\pi(S)}\right)$, which is the usual acceptance ratio of a move in the Metropolis algorithm [JS98].[2] However this transition probability (likewise the traditional transition probability) also renders the Markov chain irreducible and aperiodic; irreducible because simply there is a path

[1] Note that in the Metropolis algorithm there is usually a proposal distribution $Q_{cc'}$ involved [Mac03] which, at each step, is the distribution (different from the target distribution) that determines how to make a local move from a current state S to the next S'. Of course accepting the move is another matter. In this case our proposal distribution is nothing but the symmetric $Q_{S'S} = Q_{SS'} = \frac{1}{nT(N-1)}$.

[2] Note that if the alphabet size $N = 2$, then (6.4) may be viewed as $\text{Prob}(S'|S)$ and therefore this method will be equivalent as well to the Gibbs sampler (a.k.a. heat bath or Glauber dynamics).

between any two states, and aperiodic because the probability of returning to a state in any number of steps is positive. Moreover note that the Markov chain will be reversible, i.e., $\pi_S p_{S'S} = \pi_{S'} p_{SS'}$, and hence π will be a stationary distribution for the chain. Furthermore since the chain is irreducible and aperiodic, if we run the chain for a long time it will converge to the stationary distribution (6.3).

Note that this method can be considered as a simulated annealing with fixed temperature. While in simulated annealing the parameter θ is changed during the simulation based on a cooling schedule, here we choose a fixed θ at the beginning. However care should be given to the choice of θ. When θ is large, the stationary distribution will have a large peak at the optimal cost and therefore the chances of encountering it (once we are in steady state) is high. However, a large θ often means that convergence to the steady-state distribution can be slow (we may frequently get stuck in local maxima), as (6.4) heavily favors transitions to higher costs. On the other hand, for small values of θ, convergence to the steady state is much faster (as (6.4) is more amenable to escape from local maxima). However, the peak in (6.3) is not very pronounced at the optimal cost and so it might take a very long time until we encounter it. Therefore there is a trade-off between speed of convergence to the stationary distribution and the probability of encountering the optimal cost once the Markov chain has converged. And so choice of the correct value of θ is critical and may require trial and error.

Henceforth we will refer to the method just described as the MCMC method for simplicity. In the next section we show how using this algorithm yields interesting new results for maximization of a function of an entropy vector.

6.3.3 Ingleton Violation via MCMC

As an application of the MCMC method just described, we consider maximizing the violation of the Ingleton inequality. The Ingleton inequality holds for entropy vectors

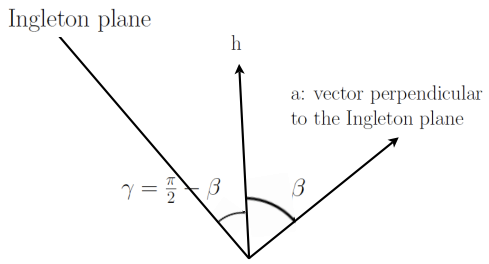


Figure 6.2: The violation index $\frac{\Delta_{ij}}{\|h\|}$ is proportional to $\cos(\beta)$.

of the random variables involved in a linear network code (and more generally entropy vectors obtained from Abelian groups [Cha07b]) and is given by [Ing71],

$$\Delta_{ij} \triangleq h_i + h_j + h_{ijk} + h_{ijl} + h_{kl} - h_{ij} - h_{ik} - h_{il} - h_{jk} - h_{jl} \leq 0. \tag{6.5}$$

However, the Ingleton inequality is not a bound on the entropy region, and there exist entropy vectors that violate it. We define the "violation index" as $\frac{\Delta_{ij}}{\|h\|}$. Note that the normalization is critical as entropy is a cone. Furthermore, the violation index is proportional to the cosine of the angle between the vector h and the vector orthogonal to the Ingleton plane (see Fig. 6.2).

To maximize the violation index using the Monte Carlo method, first we have generated a distribution table of size $T \times 4l$, i.e., for 4 vector-valued random variables of size l in general. As stated earlier, to each distribution table we can associate two entropy vectors; one obtained by considering the table as the empirical distribution of the random variables, and the other by recognizing the partitions induced on the random variables through the table and computing the entropies based on the quasi-uniform argument. We have computed violation indices through both methods for each table. Interestingly, by using the MCMC method for this problem, with parameters $T = 1000$, $N = 2$, $l = 1$, and $\theta = 6 \times 10^6$, we have found violation indices that are much bigger than the indices of the known Ingleton violating examples in the literature. When computing the entropies based on quasi-uniforms, we have found

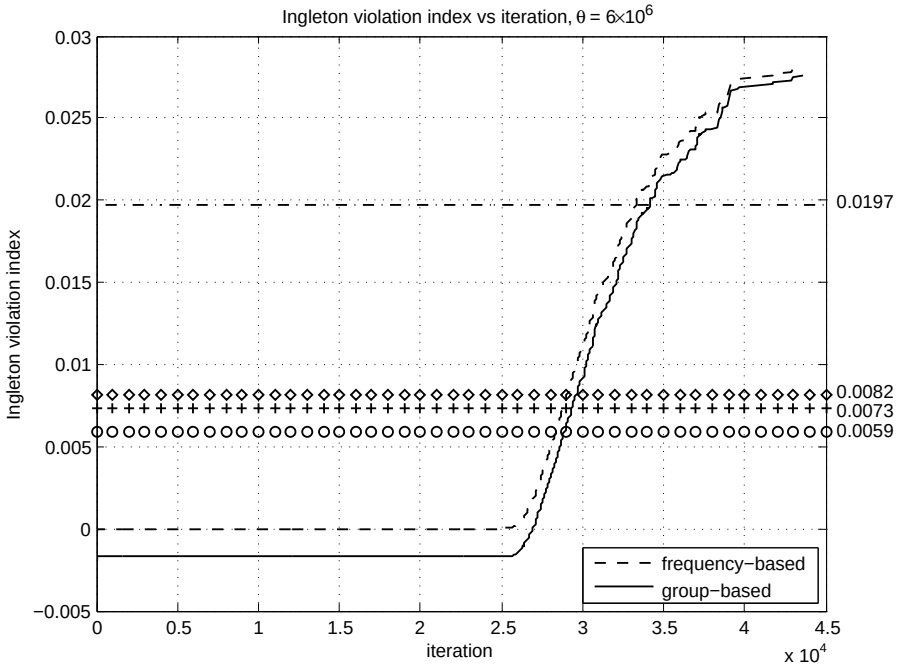


Figure 6.3: A sample run of MCMC for Ingleton violation. Entropy computed based on empirical distribution (frequency-based) and also partitions (group-based). Horizontal lines show the violations of known examples in the literature.

a maximum violation of 0.02761, and when computing the entropies directly from the distribution tables (empirical), we have found a maximum value of 0.02812. The corresponding simulation is depicted in Fig. 6.3.

Let X_i $i = 1, 2, 3, 4$ be the 4 random variables that we are considering, and let $f_\alpha(x_\alpha), \alpha \subseteq \{1, 2, 3, 4\}$ denote the frequency of appearance of $\{X_i = x_i, i \in \alpha\}$ in the T rows of the distribution table, and f_α denote the vector of $f_\alpha(x_\alpha)$ for all values of x_α. Then f_{1234} which is proportional to the joint distribution corresponding to the optimized violation indices is as follows,

$$f_{1234}(x_1x_2x_3x_4) = \begin{cases} 334 & (x_1x_2x_3x_4) = (0111) \\ 351 & (x_1x_2x_3x_4) = (1000) \\ 158 & (x_1x_2x_3x_4) = (1101) \\ 157 & (x_1x_2x_3x_4) = (1110) \\ 0 & \text{otherwise} \end{cases} . \tag{6.6}$$

Note that to compute the entropy based on the empirical distribution of the random variables, one can obtain joint entropy through $h_\alpha = -\sum \frac{f_\alpha(x_\alpha)}{T} \log \frac{f_\alpha(x_\alpha)}{T}$. On the other hand if one wishes to compute the entropies based on partitions, then each value of $f_\alpha(x_\alpha)$ gives the size of one of the segments of the partition of X_α, and therefore the group-derived entropy will be obtained via $h_\alpha = \frac{1}{T} \log \frac{T!}{\prod_{x_\alpha} f_\alpha(x_\alpha)!}$.

Note that although it is well known that the entropies do not satisfy the Ingleton inequality in general, there are only a handful of examples known that violate this bound. Understanding which entropy vectors lie outside of the Ingleton bound or how far one can go beyond this bound while staying in the entropy region are interesting questions whose answers will help us in better understanding the entropy region.

For the sake of comparison, note that the violation index of the non-quasi-uniform examples of [HRSV00] are 0.01974 and 0.00590. The violation index of the quasi-uniform example of [ZY98] which is obtained by defining a certain distribution based on projective planes is 0.0073, and the maximum violation index value of Ingleton violating example $PGL(2, p)$ of [MH09] is 0.0082, which occurs for $p = 13$. These values are marked in Fig. 6.3.

6.4 Nonlinear Network Coding

The idea of a random walk over distributions in Section 6.3 can be extended to a biased random walk (Markov Chain Monte Carlo) over all (possibly nonlinear) operations in a network.

Assume that vector-valued signals of size l over alphabet size N are transmitted across edges of the network.[1] In such a setup, if the in-degree of a particular node in the network is D, then the node must map each of its possible $N^{l \times D}$ inputs to its corresponding outputs. For each output, this mapping can be represented by a truth table with $D + 1$ block columns of size l, the last block column representing

[1] In general, source variables and middle variables of the network can be of different vector sizes.

the output, and $N^{l\times D}$ rows (one for each input combination). There are a total of $N^{l\times N^{l\times D}}$ possible truth tables, and thus a total of $N^{l\times N^{l\times D}}$ possible nonlinear network operations for this particular output of the internal node. On the other hand, note that, if we restrict ourselves to linear mappings, there will only be $N^{l^2\times D}$ possible mappings (i.e., the coefficient matrix) for this output of the node. The total number of nonlinear network operations is obtained from the conjunction of the possible operations for each internal node and can be computed to be:

$$N^{\sum_{j\in\nu, j\notin\mathcal{S}_{source}\cup\mathcal{S}_{sink}} l\times|Out(j)|N^{l\times|In(j)|}}, \tag{6.7}$$

where $|Out(j)|$ and $|In(j)|$ are the out- and in-degree of the node j. On the other hand for linear coding, the total number of possible codes are,

$$N^{\sum_{j\in\nu, j\notin\mathcal{S}_{source}\cup\mathcal{S}_{sink}} l^2D|out(j)|}. \tag{6.8}$$

Considering truth tables in the case of nonlinear mappings, and coefficient matrices in the case of linear codes, as the states of the system, we can define a Markov chain on these state spaces similar to what we did in Section 6.3. However we can perform the local moves in two manners. The first way is to choose an entry of the truth table or the coefficient matrix uniformly and changing its value to any other $N-1$ possible values. We call this method the "uniform flip" and this will be similar to what we did in Section 6.3. Another way is to first choose an internal node randomly, choose an output of this node at random, and then select one of the entries of its truth table or the local coefficient matrix randomly, and then flip its value to any other $N-1$ possible values. We call this method the "node-wise flip". Assuming that we want to maximize some cost function of the network that can be written in terms of entropy (such as the weighted sum rate), we can define the stationary distribution and the transition probabilities as (6.3) and (6.4) in Section 6.3. Both

the uniform and node-wise flip render the Markov chain aperiodic and irreducible, and assuming (6.3) and (6.4) the chain will also be reversible. However the node-wise flip will allow us to adjust the method for distributed operation over the networks, as will be discussed in Section 6.6. To summarize, the same MCMC technique can be applied on the network operations to bias them toward large costs.

In what follows we have applied this method to maximize the sum-rate of different networks. In particular we show how solutions (linear or nonlinear) can emerge from this technique.

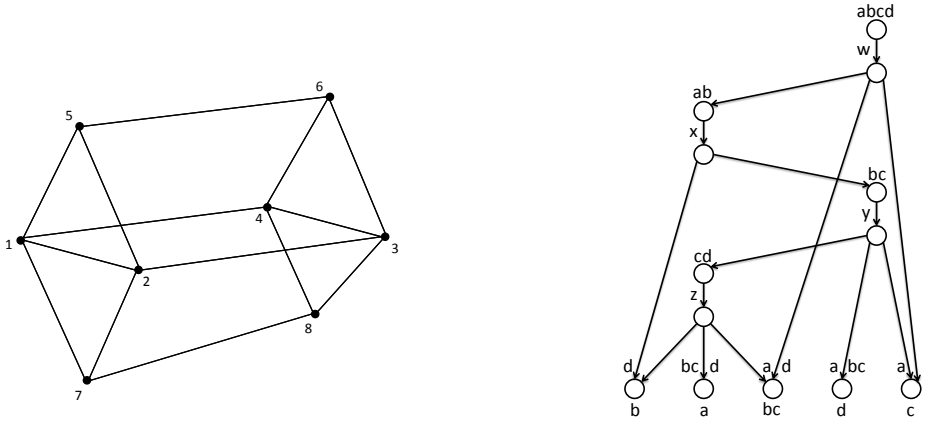
Remark: In the networks that we analyze, we normalize the sum-rate such that when all the sinks successfully recover their demands, normalized sum-rate becomes equal to 1. In such cases the MCMC method gives a valid linear or nonlinear code that solves the network. However, if the optimum sum-rate turns out to be less than 1, then it essentially means that one or more sinks do not fully reconstruct their demands, in which case a coding is required to deliver the found optimum rate to each sink.

6.4.1 Random Walk on Truth Tables

We have applied the MCMC method described in the last section to some networks of interest, and in what follows we give the simulation results and their comparison with the existing results.

6.4.1.1 Vamos Network

The Vamos network (Fig. 6.4b) is obtained from the well-known Vamos matroid (Fig. 6.4a) and was first introduced in [DFZ06b], where the authors showed that the network is not solvable and proved the insufficiency of Shannon-type information inequalities for determining the capacity of general networks, reaffirming the importance of the full characterization of Γ_n^*. However using a non-Shannon type in-



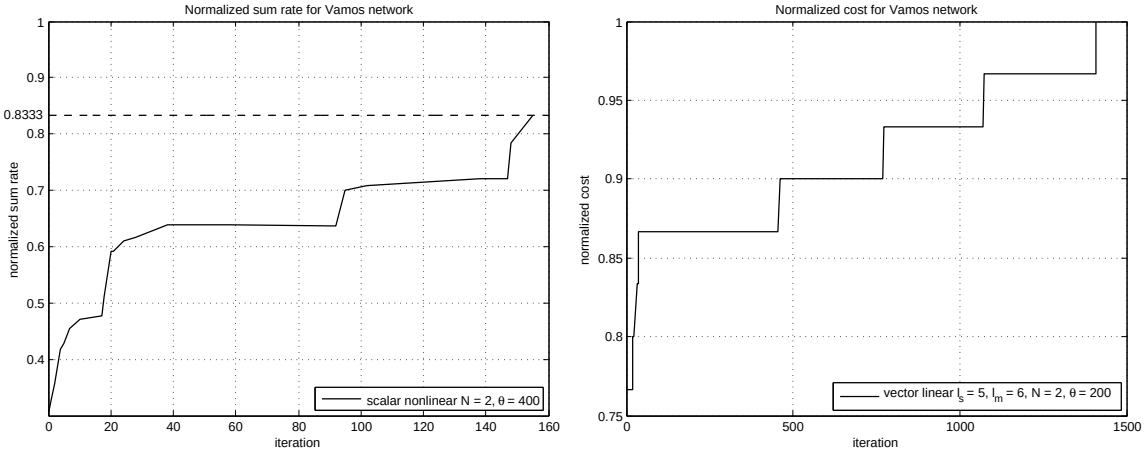
(a) The Vamos matroid

(b) Vamos network constructed from the Vamos matroid [DFZ06b]

Figure 6.4: The Vamos matroid and network

formation inequality, they provided an upper bound of 10/11 for the network coding capacity.

They also found the linear coding capacity of the Vamos network to be $\frac{5}{6}$ over every finite field and gave a $(5, 6)$ vector-valued solution, i.e., a linear solution with vector size of 5 for sources and vector size of 6 for the rest of the network variables. In this network a, b, c, d are sources and x, y, z, w are internal messages. There are 5 sinks whose demands are shown below them in Fig. 6.4b.



(a) Nonlinear coding, alphabet size 2

(b) Vector linear code of rate $\frac{5}{6}$

Figure 6.5: Monte Carlo simulation to optimize the sum rate of the Vamos network

Table 6.1: Truth tables for the Vamos network yielding the normalized sum rate of $\frac{5}{6}$

a	b	c	d	w
0	0	0	0	1
0	0	0	1	0
0	0	1	0	0
0	0	1	1	1
0	1	0	0	0
0	1	0	1	1
0	1	1	0	1
0	1	1	1	0
1	0	0	0	1
1	0	0	1	0
1	0	1	0	0
1	0	1	1	1
1	1	0	0	0
1	1	0	1	1
1	1	1	0	0
1	1	1	1	0

a	b	w	x
0	0	0	0
0	0	1	1
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	0
1	1	1	1

b	c	x	y
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	0
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	0

c	d	y	z
0	0	0	0
0	0	1	1
0	1	0	0
0	1	1	0
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1

We consider the following normalized sum rate as the cost function,

$$\frac{1}{6 \times l \times \log(N)} \Big(I(b; dz) + I(a; bcdz) + I(bc; adzw) + I(d; abcy) + I(c; awy) \Big)$$

where 6 is the total number of demands, l accounts for the vector length of source random variables, and N is the alphabet size. To maximize this cost, we employ the Monte Carlo method as stated in the previous section to do a random walk on the truth tables of this network. Since there are 4 message variables in the network, there will be a total of 4 truth tables—for x, y, z, and w, respectively.

Assuming the simplest case, we considered $N = 2$, i.e., binary alphabet-size, and scalar valued random variables for all the source and message variables and searched for nonlinear codes. A sample run of this Monte Carlo maximization can be seen in Fig. 6.5a where the normalized sum rate has quickly reached the point $\frac{5}{6} = 0.8333$, i.e., the linear coding capacity of the network. The truth tables that correspond to this maximized sum rate can be seen in Table 6.1. Note that each run of the network with similar parameters potentially finds new truth tables for the network,

and the one that we state here is only one of the many potential solutions. A little examination of the table reveals the following nonlinear coding found for the variables of the network,

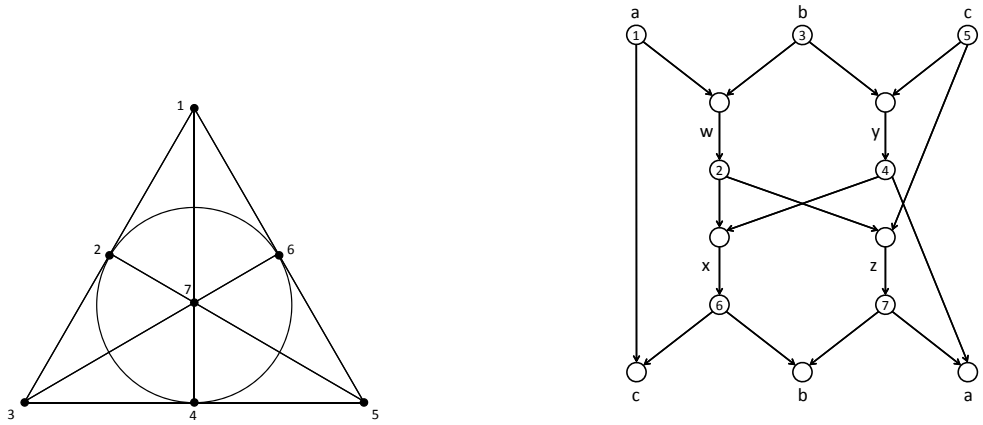
$$w = \overline{b + c + d}, \quad x = w, \quad y = c + x, \quad z = \overline{d}y + cd \tag{6.9}$$

where all operations are over GF(2) and $\overline{(.)}$ refers to the NOT operation over binaries. One can easily see that all the demands of the network can be fully recovered except the demand of the second sink which wants a. This is clearly due to the fact that in this coding message a is not carried into the network. Therefore the normalized sum rate of the network for this code becomes $\frac{5}{6}$.

Although there is the $\frac{10}{11}$ upper bound for this network, running the simulation for many more iterations did not result in a better sum rate than $\frac{5}{6}$, raising the possibility that this is the best achievable rate among all nonlinear scalar binary codes. Nonetheless, we have simply found a rate $\frac{5}{6}$ linear code over binaries for this network—in which sources are vectors of size 5 and the rest of the variables are of size 6—that achieves the normalized sum rate of 1. Note that a $(5, 6)$ linear solution for this network has been previously reported in [DFZ07]. Here we want to emphasize the ability of this method to find similar solutions. The corresponding simulation can be seen in Fig. 6.5b, and the actual found solution with the encoding and decoding mappings are stated in the appendix.

6.4.1.2 Fano and Non-Fano Networks

The Fano network is constructed from the Fano matroid (Fig. 6.6) and was first introduced in [DFZ07]. The Fano network was shown in [DFZ06c] to be solvable if and only if the alphabet size is a power of 2. In particular for the case of linear codes, it was proved in [DFZ05] that this network has a scalar linear solution over any ring with characteristic 2, and does not have any vector linear solution over a finite field



(a) Fano matroid

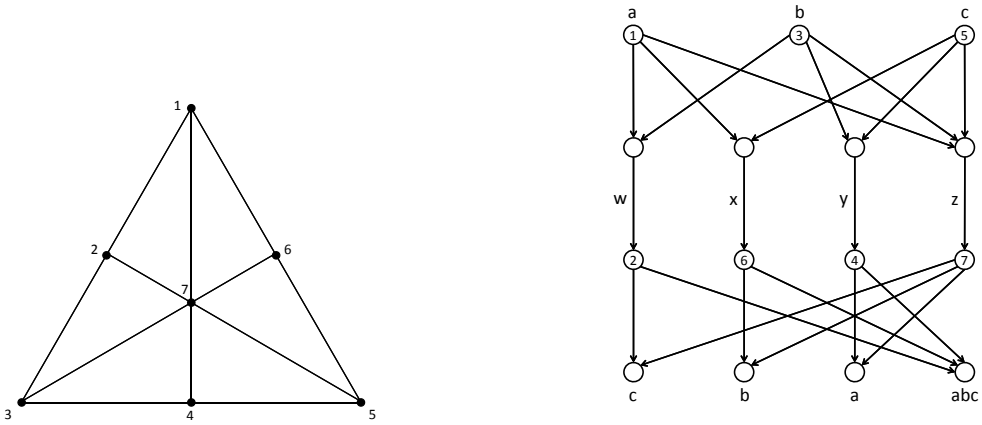
(b) Fano network constructed from the Fano matroid. The numbers inside nodes show the corresponding matroid element.

Figure 6.6: Fano matroid and network

with odd characteristic, irrespective of the vector dimension.

This is similar to the property of the Fano matroid which is known to be representable if and only if the field characteristic is 2. Note that since any scalar/multilinear representation of the Fano matroid immediately induces a linear solution to the Fano network, the fact that the Fano network does not admit any vector linear solution over fields of odd characteristic implies that the Fano matroid does not have a multi-linear representation over fields of odd characteristic either. The Monte Carlo method has shown to be promising in this case as well. As can be seen from Fig. 6.8a, it has quickly found scalar linear codes for the network over even characteristic fields GF(2) and GF(4), and also a nonlinear code over alphabet size 2.

The non-Fano network is also similarly constructed from the non-Fano matroid (Fig. 6.7) [DFZ07]. The non-Fano matroid (Fig. 6.7a) is very similar to the Fano matroid except that, as opposed to the Fano where the elements $\{2, 4, 6\}$ formed a circuit, this set is now an independent set of the non-Fano matroid. Interestingly the non-Fano network was shown in [DFZ06c] to be solvable if and only if the alphabet size is odd. In other words this network is solvable only over alphabet sizes where the

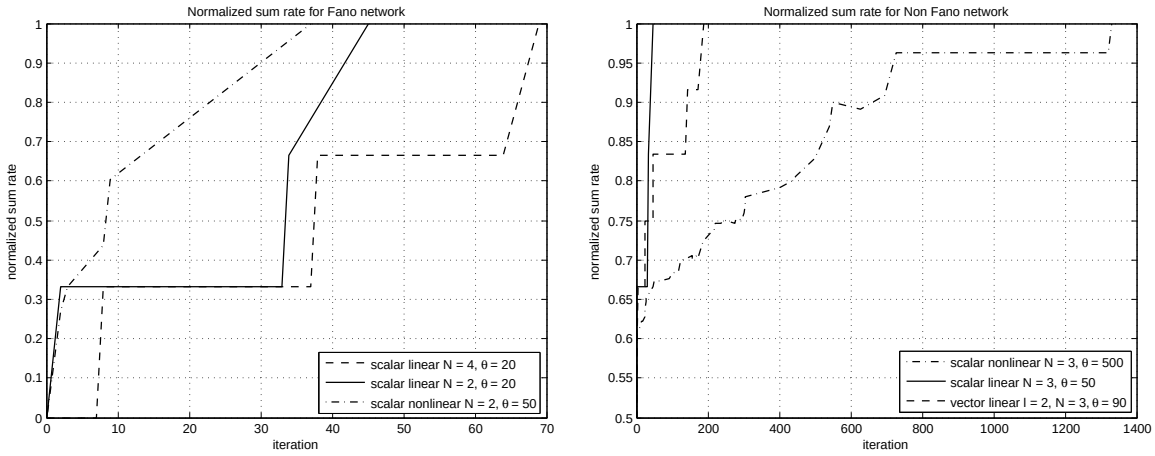


(a) Non-Fano matroid

(b) Non-Fano network constructed from the non-Fano matroid. Numbers inside nodes show the corresponding matroid element.

Figure 6.7: The non-Fano matroid and network

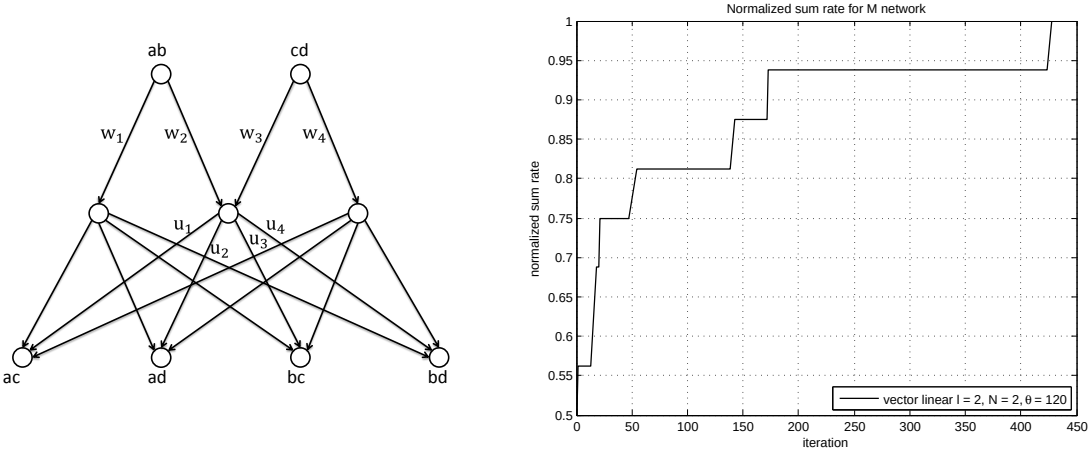
Fano network is not solvable. Moreover in [DFZ05] it is proved that while this network admits a scalar linear solution over any ring in which 2 is an invertible element, it does not have vector linear solution over any field with characteristic 2 for any vector dimension. Again this is similar to the property of the non-Fano matroid, which is known to be representable only over fields with characteristic other than 2. Moreover from the property of the non-Fano network, we deduce that the non-Fano matroid



(a) Sample runs of MCMC for Fano network

(b) Sample runs of MCMC for non-Fano network

Figure 6.8: MCMC for Fano and non-Fano networks



(a) M network

(b) A vector linear solution for the M network $l = 2, N = 2$

Figure 6.9: M network and the corresponding MCMC simulation

does not have a multi-linear representation over fields with characteristic 2. Some of the Monte Carlo simulations for the non-Fano matroid can be seen in Fig. 6.8b, where it has successfully found scalar and vector linear solutions over GF(3) and also a scalar nonlinear code over GF(3). The combination of the Fano and non-Fano network was used in [DFZ05] to construct a network that is not linearly solvable.

6.4.1.3 M Network

This network was first introduced in [MEKH03] as an example of a network which does not have any scalar linear solution, however it has a simple routing solution on a vector space of dimension 2. Later [DFZ07] showed that this solution can be easily extended to any vector linear solution of even dimension and in fact this network does not admit any linear solution over vector spaces of odd dimensions. This network is depicted in Fig. 6.9a, from which it can be seen that the network gets its name from its shape. The Monte Carlo has successfully found a vector linear solution of length 2 over binaries, as expected (Fig. 6.9b).

6.4.1.4 Repair Problem in a Storage System

In distributed storage systems where there is a possibility of failure for storing devices, some form of redundancy needs to be introduced in order to maintain a reliable system. While the simplest form of redundancy is replication, it has been proved that coding is more advantageous than replication [DRWS11]. In general the data that needs to be stored is assumed to be of size $\mathcal{M}$ and is encoded into n packets of the same size. Each encoded piece is assumed to be stored at a "storage node". Since the ultimate goal is to recover the original data, one should be able to recover the source messages by merely accessing the n encoded data packets. This can be achieved using different coding schemes, such as the erasure codes. However if the storage nodes themselves also fail or leave the network over the time, then the reliability of the storage system will diminish. Therefore there should be a mechanism in place that allows the network to repair itself, meaning that whenever a storage node fails, the network can construct new data packets and store them at a new storage node, such that the new data—along with the information of the surviving storage nodes—again forms a desired code that allows the recovery of the original source data. The network should construct data for the substitute storage node solely by accessing the surviving (working) storage nodes. This is called the *repair problem* [DGYWR] and is studied in three different scenarios which are recalled as the *functional repair*, *exact repair*, and *exact repair of systematic parts* [DRWS11]. While the "functional repair" requires that the newly constructed encoded packet for the substitute storage node forms a desired code with the other surviving storage nodes such that the original data can still be recovered, the "exact repair" requires that the newly constructed data be exactly the same as the lost encoded data of the failed storage node. The "exact repair of systematic parts" is a combination of the previous two methods, where it is assumed that the code is systematic, meaning that an uncoded copy of the source messages exists in the n encoded pieces and, when the systematic part of the

encoded nodes fail, exact repair, and for the failure of the rest of the storage nodes, functional repair is required. Since storage nodes may fail every once in a while, in all these scenarios it is important to construct the new storage nodes by minimally downloading data from the survivor nodes so as to prevent a large network traffic due to repair. Determining the minimum required (download) bandwidth is a main question in the repair problems.

The repair problem has been fully solved in some special cases; namely the functional repair [DGYWR] and some regimes of the exact repair problem [DRWS11]. In all those cases, it has happened that the cutset lower bound for optimal bandwidth is tight.

As there has been an increasing interest in the repair problem, in this section we consider two special cases of the "exact" repair problem and show that our MCMC method is able to find explicit codes for these networks, even though these are larger networks (i.e., involve greater number of random variables) compared to our previous examples, and our MCMC search is performed over larger finite fields. Before getting into details of the cases that we have considered, we explain the general setting for the "exact" repair problem.

The graph model (Fig. 6.10) for this problem consists of a source and n storage nodes which are directly connected to the source with infinite capacity [DGYWR]. Each storage node is formed from two sub-nodes and a directed edge between the sub-nodes, which is assumed to carry the encoded data packet. The capacity of this edge is the capacity of the storage node and is denoted by α. To ensure the recovery of the original data one should be able to reconstruct the source from a subset of the n storage nodes. Usually the symmetric case is considered where one assumes that the source can be recovered from every subset of size k of the storage nodes. Moreover to make the code "self-healing" in the exact repair sense, one needs to guarantee the construction of each storage packet by accessing a subset of the rest of

the other storage packets. Therefore considering symmetry, the assumption is that every storage node can be recovered by accessing *any* d number of the remaining storage nodes and downloading β bits of information from each of them. As a result, the repair bandwidth will be equal to $\gamma = d\beta$. By analyzing the cutset bound for this graph, [DGYWR] has shown that there is a trade-off curve between optimal α and γ, giving rise to two particular points of interest; the *minimum bandwidth* and the *minimum storage* points. The codes that achieve these points are called *minimum bandwidth regenerating (MBR) codes* and the *minimum storage regenerating (MSR) codes*, respectively.

For a given original file size of $\mathcal{M}$, parameters d and k, the (α, γ) of the MSR point is characterized in [DGYWR] by $(\alpha_{\text{MSR}}, \gamma_{\text{MSR}}) = \left(\frac{\mathcal{M}}{k}, \frac{\mathcal{M}d}{k(d-k+1)}\right)$. Interestingly it is observed in [DGYWR] that γ_{MSR} is a decreasing function of d, and therefore to achieve the smallest repair bandwidth one should set $d = n-1$, i.e., once a storage node fails, if all the remaining $n-1$ nodes are employed for recovery of the lost encoded packet, the total repair bandwidth can be reduced. Under the assumption of $d = n-1$, the (α, γ) of the MSR point becomes $(\alpha_{\text{MSR}}, \gamma^{\min}_{\text{MSR}}) = \left(\frac{\mathcal{M}}{k}, \frac{\mathcal{M}(n-1)}{k(n-k)}\right)$. Nonetheless, note that since the MSR point corresponds to the optimal point obtained from the

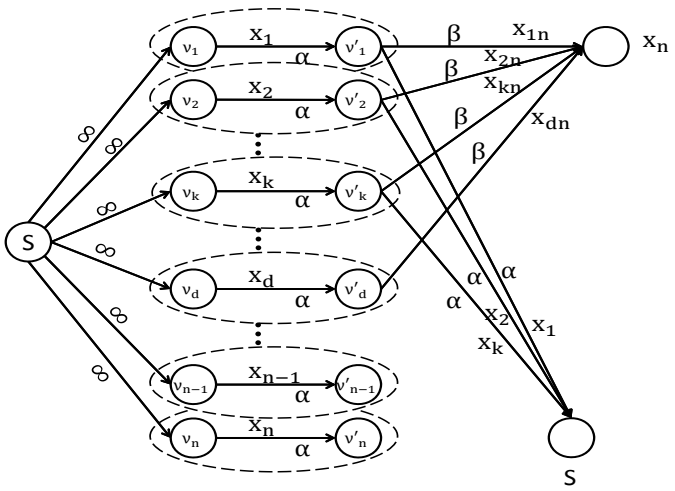
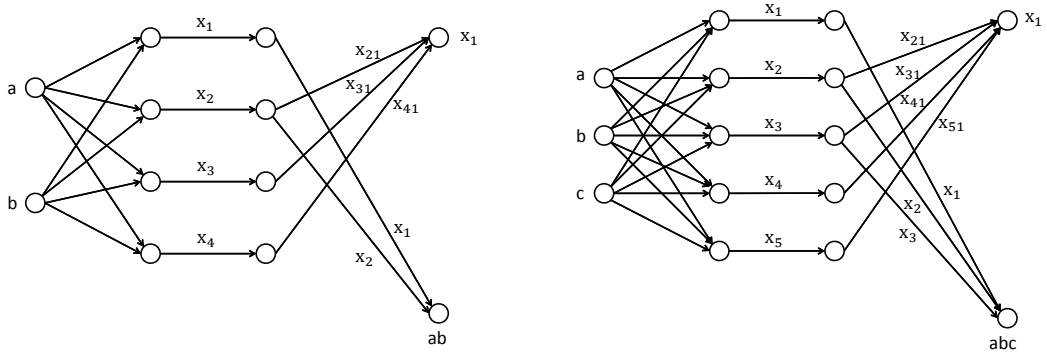


Figure 6.10: Exact repair model

cutset bound, it may not be achievable for the exact repair problem (in contrast to the functional repair for which the cutset bound is tight, as it can be reduced to a multicast network coding scenario [DGYWR]). Assuming such point is achievable one can make two assumptions about a storage network operating at the MSR point. First, since the file is of size $\mathcal{M}$ and $\frac{1}{k}$ of it is stored at each storage node, and since every k of the storage nodes recovers the original file (similar to the property of *maximum distance separable* (MDS) codes), one can assume that there are also k sources each of size $\frac{\mathcal{M}}{k}$ that are being encoded into n nodes. Furthermore, since $\beta = \frac{1}{n-k} \cdot \frac{\mathcal{M}}{k}$, one can assume that each encoded packet is further split into $n-k$ sub-packets. Note that this is equivalent to considering vector codes of length $n-k$ for the repair problem at the MSR point. To summarize, this setting is equivalent to a network coding problem where there are k sources with messages of length $n-k$, n storage nodes with messages of also length $n-k$ directly connected to the sources, and two types of sinks (sinks that connect to any k subset of storage nodes with edges of capacity $n-k$ and demand the k sources, and sinks that connect to $n-1$ of the n storage nodes via edges of capacity 1 and demand the data of the nth storage node). Note that if the MSR point is achievable for the exact repair, this network will be solvable.

Here we consider two examples of the exact repair problem at the MSR point, namely the $(n,k) = (4,2)$ and the $(n,k) = (5,3)$ cases. We denote the source variables by $a, b,$ etc., the encoded messages at the storage nodes by X_i, and the outgoing signal of storage node i that is used for the recovery of storage packet j by X_{ij}. As explained previously, while source messages and also the X_i variables are vectors of size $n-k = 2$, X_{ij} variables are scalar. These settings are partly shown in Fig. 6.11. Note that not all the sinks are drawn. In both cases the cost that we intend to optimize via the



(a) Exact repair $(4, 2)$; partial setting (b) Exact repair (5,3); partial setting

Figure 6.11: Exact repair settings of (4,2) and (5,3)

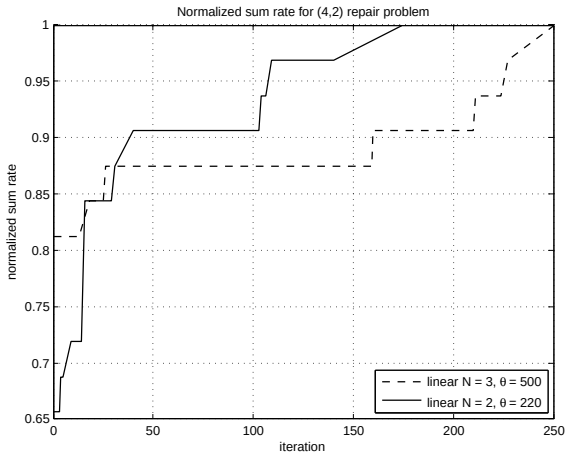
MCMC method is the following normalized sum rate,

$$\frac{1}{2m\log(N)}\left(\sum_{p=1}^{k}\sum_{\substack{\alpha\subseteq\{1,\ldots,n\}\\|\alpha|=k}} I(X_{ps};\{X_i\ i\in\alpha\})+\sum_{i=1}^{n} I(X_i;\{X_{ji},j\neq i\})\right) \tag{6.10}$$

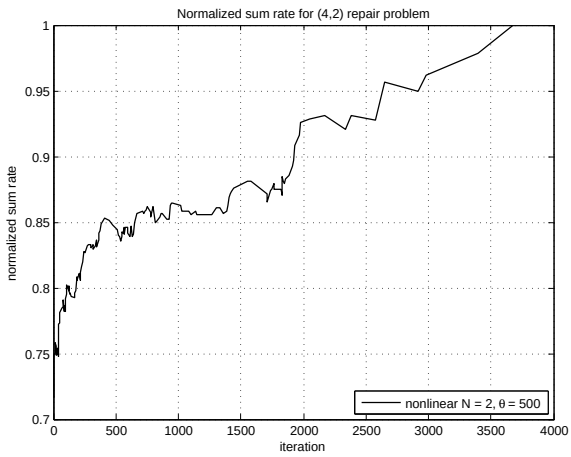
where X_{ps} denotes the source messages (i.e., $X_{1s}=a$, $X_{2s}=b$, etc., division by 2 accounts for normalization by the length of the vectors $n-k=2$, division by $m=k\binom{n}{k}+n$ accounts for the number of terms in the parentheses, and division by $\log(N)$ accounts for normalization by log of the alphabet-size).

Achievability of MSR for exact repair $(n,k)=(4,2)$*:*

This network is shown in Fig. 6.11a. To find codes over alphabet size N that achieve the MSR point, we need to solve this network, which we will do by maximizing the cost (6.10) through a MCMC method. Linear codes that achieve the MSR point have been previously reported [DRWS11]. Our Monte Carlo method has also successfully found optimal cutset achieving code (corresponding to regenerating MSR codes). Interestingly for the $(4,2)$ problem we have found both a linear and also a non-linear solution. The simulations can be seen in Fig. 6.12.



(a) Simulation for (4,2) repair problem (linear codes)



(b) Simulation for (4,2) repair problem (nonlinear code)

Figure 6.12: Simulations for the (4,2) exact repair

Achievability of MSR for exact repair $(n, k) = (5, 3)$*:*

The structure of this network is shown in Fig. 6.11b. Similar to the $(4, 2)$ case, we have employed the Monte Carlo method to solve this network over different alphabet sizes N by maximizing the normalized sum rate (6.10). We have studied the network over linear operations. Simulation results show (Fig. 6.13) that the MCMC method finds linear codes for the MSR point over alphabet sizes of $3, 4$, and 7.

6.5 Matroid Representation Via Linear Network Coding

In the previous section we saw how the MCMC method can be used to yield linear or non-linear codes for a given network. In this section we use the same ideas to find linear representations for matroids. One may review Chapter 4 for more details about matroids and linear representability.

Definition 6.5.1 (Matroid Representability) *Let M be a matroid with n elements and rank r. Moreover let A be an $rk \times nk$ matrix with entries over finite field F. Partition the columns of A into n sets of equal size k and call each parti-*

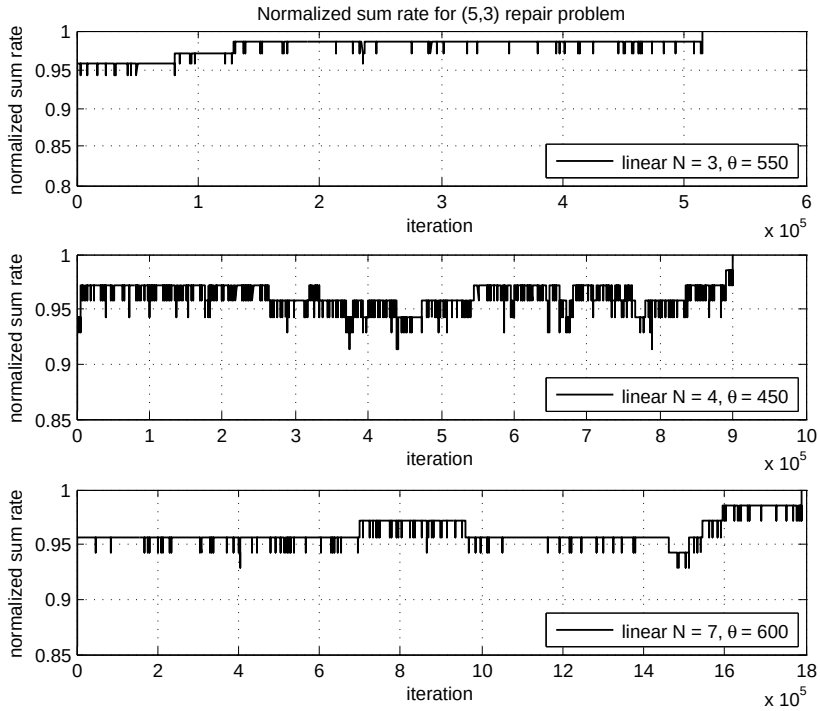


Figure 6.13: Simulation for the (5,3) repair problem

tion a supercolumn. The matroid M is said to be representable if there is a bijection between elements of M and supercolumns of such matrix A, such that a subset of elements of M is an independent set in M if and only if the set of corresponding supercolumns of A are linearly independent. In other words, if we normalize the rank of A by k, then the rank of any subset of matroid elements is equal to the normalized rank of the corresponding supercolumns. If $k = 1$ we say that M has a (scalar) linear representation, while if $k > 1$ we say that M has a multilinear (or k-linear) representation. Finally matroid M is said to be representable if it is representable over some finite field F.

As was discussed in Chapter 4, determining if a matroid is representable over a particular finite field, or in general representable, is an interesting question, and there has been a lot of research in this area. Most of the attention though, has been towards scalar representability of matroids. Even in the scalar case, although the representability problem over finite fields GF(2),GF(3),GF(4) is completely solved

[Oxl04], in general it remains an open problem.

As an application of the MCMC method, and using the ideas of the previous section, one can potentially find scalar or multilinear representations for matroids. The idea is to construct a network from the matroid of interest (using the method of [DFZ07]) and try to find a scalar/vector linear code for the network. However in order for the found solution of the network to be a valid representation of the matroid, one needs to incorporate all the dependency and independencies of the matroid in the network construction. In the rest of the section we consider the non-Pappus and U_{24} matroids, and find multilinear representations for them. We show that in these two cases one need not include all the dependency relations of the matroid in the network construction to obtain a valid representation of the matroid through linear solution of the network. Before explaining these two cases however, we briefly explain the method of Dougherty et al. [DFZ07] for constructing networks from matroids to describe constructions of non-Pappus and U_{24} networks.

6.5.1 Network Model of a Matroid

One can construct a network from a matroid by using the method of [DFZ07] such that the dependency and independency relations of the matroid elements are reflected in the network topology, demands, and independency of its sources. We summarize the steps of this method as follows, where we assume that M is a given a matroid of rank r:

1. (Creating source nodes) Choose a base of the matroid $B = \{x_1, \ldots, x_r\}$ and create a source node containing a source message S_i for each element x_i of that base.

2. (Creating the rest of the network nodes and messages) Find a circuit of the matroid $C = \{x_1, x_2, \ldots, x_p\}$ whose all elements except one (say x_j) are assigned

to a node in the network. Create a new node with input edges from all the nodes of $C \setminus x_j$ and with a single output edge e. Then create a second node with a single input from e and assign the element x_j to this node. Moreover assign a new message m_j to the edge e. Repeat this step until it is no longer possible.

3. (Creating sinks and defining demands) Repeat this step as many times as desired: Find a circuit of the matroid $C = \{x_1, \ldots, x_p\}$ such that the dependency of at least one of its elements (say x_k) on the other variables in the circuit is not enforced in the network. Create a sink with input edges from $C \setminus x_k$ that demands m_k.

4. Repeat this step as many times as desired: Find a base $B' = \{x'_1, \ldots, x'_r\}$ and create a sink node with input edges from the corresponding nodes of $x'_1, \ldots, x'_r$ which demands all the network messages $S_1, \ldots, S_r$.

In order to find a linear representation for the matroid, the constructed network needs to adopt the matroid properties as much as possible. In particular, in the worst case all dependency and independency conditions need to be enforced in the network (i.e., repeating steps 3 and 4 until it is no longer possible to do so). Note that since we want to have the option of enforcing all dependency conditions, we have tweaked step 3 of the algorithm compared to the method of [DFZ07], where they only define demands that correspond to source messages (i.e., x_k corresponds to S_k). Although it is more sensible to demand the recovery of sources at the sinks, for our purpose it is important to be able to incorporate all the matroid relations in the network. Depending on which steps are taken and how many times they are repeated in the above construction, many different networks may be obtained from a single matroid.

6.5.2 Multilinear Representation for Matroids

In this part we describe the network construction for the matroids non-Pappus and U_{24}, and show how using the biased MCMC results in multilinear representations of these matroids. While scalar linear representability over fields has been extensively studied, there has not been much research on multi-linear representability. In fact there are networks which are not scalar representable over some field F, but admit a multilinear representation over the same field. As an example, although the non-Pappus matroid is not linear representable over any field [Oxl06], it has been shown that it has a multilinear representation over $GF(3)$ [SA98, Mat99]. Moreover as it is well known that the uniform matroid on 4 elements (U_{24}) is not representable over binaries (and, in fact, it is a forbidden minor for linear representability over binaries [Oxl06]), we show that it has a multi-linear representation over $GF(2)$.

6.5.2.1 Non-Pappus Network

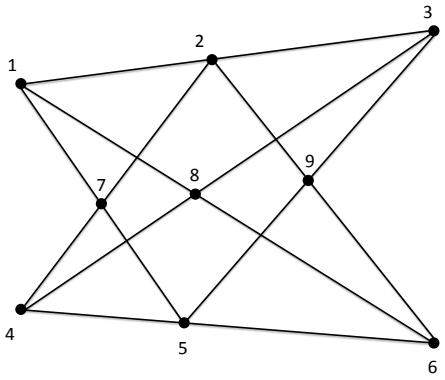
This network is constructed from the non-Pappus matroid [Oxl06] (see Fig. 6.14a).

Definition 6.5.2 (Non-Pappus Matroid) *Let $E = \{1, \ldots, 9\}$ be a set of 9 elements and let $S = \{\{1,2,3\}, \{1,5,7\}, \{1,6,8\}, \{2,4,7\}, \{2,6,9\}, \{3,4,8\}, \{3,5,9\}, \{4,5,6\}\}$. Then define the function $r : 2^M \to \mathbb{Z}^+$ as,*

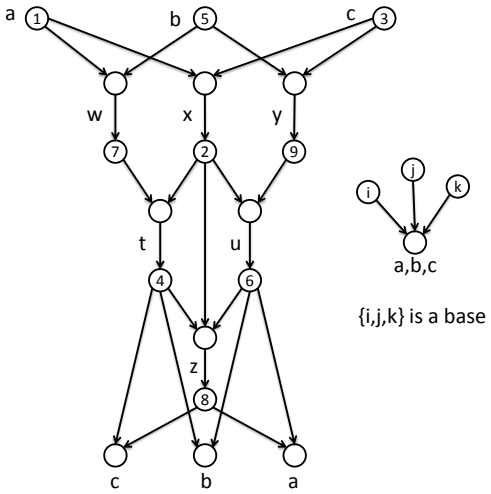
$$r(\alpha) = \begin{cases} \min\{|\alpha|, 3\} & \alpha \notin S, \alpha \subseteq M \\ 2 & \alpha \in S \end{cases} . \tag{6.11}$$

Then the non-Pappus matroid is the matroid M with ground set E and rank function r.

As the rank function implies, the circuits (i.e., the minimal dependent sets) of size 3 of this matroid are given by the set S, and the bases (i.e., the maximal independent sets of the matroid) consist of all the 3-element subsets of E which are not in S.



(a) Non-Pappus matroid



(b) A construction of the non-Pappus network. The numbers inside the nodes show the corresponding matroid elements.

Figure 6.14: Non-Pappus matroid and the constructed network

The non-Pappus matroid is interesting in that it has been proven that it is not representable over any field [Oxl06]. Nonetheless [SA98] and [Mat99] have shown that this matroid has a multi-linear representation over $GF(3)$. Using the method of [DFZ07], briefly described previously, many possible networks can be constructed from the non-Pappus matroid. However if one wishes that the resulting network inherits the properties of the matroid, one must incorporate the dependency and independency relations of the matroid in the network construction as much as possible. Our construction is as follows:

We start off with the base $B = \{1, 3, 5\}$ of the matroid to create the sources, and construct the rest of the network based on the steps mentioned before. The circuits

and the relations that we have used are as follows:

$$\text{circuit } \{1,5,7\} : 1,5 \to 7, \quad \text{circuit } \{1,2,3\} : 1,3 \to 2,$$

$$\text{circuit } \{3,5,9\} : 3,5 \to 9, \quad \text{circuit } \{2,4,7\} : 2,7 \to 4,$$

$$\text{circuit } \{2,6,9\} : 2,9 \to 6, \quad \text{circuit } \{2,4,6,8\} : 2,4,6 \to 8$$

$$\text{circuit } \{3,4,8\} : 4,8 \to 3 \text{ demand},$$

$$\text{circuit } \{4,5,6\} : 4,6 \to 5 \text{ demand},$$

$$\text{circuit } \{1,6,8\} : 6,8 \to 1 \text{ demand}.$$

The core of the network is shown in Fig. 6.14b. Note that the numbers inside the nodes show the corresponding matroid element assigned to that node. We further add all the sinks that can be obtained via step 4 of the construction (i.e., for each of the matroid bases we add a sink with input edges connected to the nodes corresponding to the elements of that base, and which demands all the source messages a, b, c). There are 76 such sinks.

Authors of [RSG] have a similar construction for the non-Pappus network in the context of index coding. Note that while we have enforced all the independency conditions through the extra 76 sinks in the network of Fig. 6.14, we have not imposed all the dependency relations of the matroid on the network. Nonetheless, in the following theorem we show that we do not need to impose any other relation, as this network already reflects the matroid properties for the matter of linear representation. A weaker version of this theorem is presented for an alternative construction of non-Pappus network in [RSG].

Theorem 6.5.3 *The constructed non-Pappus network admits a scalar/vector linear solution over $GF(N)$ if and only if the non-Pappus matroid has a linear/multilinear representation over $GF(N)$.*

Proof: First note that if the non-Pappus matroid has a linear (scalar or multilinear) representation, then clearly that representation can be written in such a way that all the elements of the matroid are obtained in terms of the elements of the base $\{1, 3, 5\}$, similar to Fig. 6.16. This representation immediately gives a linear solution for the non-Pappus network. Conversely, assume that there is a vector linear solution of size m for the non-Pappus network. This means that each variable of the network X_i can be written as a linear combination of the source messages,

$$X_i = A_i \begin{pmatrix} a \\ b \\ c \end{pmatrix} \tag{6.12}$$

where a, b, c and all variables of the network (i.e., $X_i,\ \forall i$) are m-dimensional. Thus A_i is a $m \times 3m$ matrix. Assume that variable X_i corresponds to the matroid node i. Since all the base independency conditions are enforced in the network, for any i, j, k that represents a base we should have $\text{rank}[A_i^T, A_j^T, A_k^T] = 3m$, which, when normalized by the dimensionality, means $\text{rank}(X_i, X_j, X_k) = 3$. Moreover this implies that any set of variables with less than 3 elements should also be independent (as for any set with less than 3 elements in this matroid, there is an independent set which contains it). On the other hand, since all the 3-element circuits of the matroid are also enforced in the network, for any i, j, k that represents a circuit, one of the variables can be written in terms of the other two (e.g., $X_k = B_1 X_i + B_2 X_j$, or equivalently $A_k = B_1 A_i + B_2 A_j$). Therefore, $\text{rank}[A_i^T, A_j^T, A_k^T] = 2m$ (i.e., the normalized rank of the variables gives $\text{rank}(X_i, X_j, X_k) = 2$). Furthermore, for any set of more than 3 elements $\{i_1, \ldots, i_n\}, 4 \leq n \leq 9$, we have $\text{rank}[A_{i_1}^T, A_{i_2}^T, \ldots, A_{i_n}^T] \leq 3m$, however since all the bases are ensured to have full rank, and since in this network every set with at least 4 elements includes a base, inevitably it should satisfy $\text{rank}[A_{i_1}^T, A_{i_2}^T, \ldots, A_{i_n}^T] = 3m$. We conclude that if the set of A_i constitutes a linear code for the non-Pappus

network, their normalized rank should satisfy,

$$\text{rank}[A_\alpha^T] = \begin{cases} min(|\alpha|, 3) & \alpha \text{ is not a 3-element circuit} \\ 2 & \alpha \text{ is a 3-element circuit} \end{cases} . \tag{6.13}$$

This coincides with the properties of the rank function of the non-Pappus matroid, and therefore the set of A_i composes a multilinear representation for the non-Pappus matroid. □

Therefore, based on Theorem 6.5.3, if we find a linear solution for the non-Pappus network, it will immediately give us a linear representation for the non-Pappus matroid. Note that since it is already known that the non-Pappus network is not scalar representable over any field, the non-Pappus network will not admit a scalar linear solution. However, it was found in [SA98, Mat99] that this matroid has a 2-linear representation over GF(3). Their representation is as follows,

$$\begin{pmatrix} 10 & 10 & 00 & 10 & 00 & 10 & 10 & 10 & 00 \\ 01 & 01 & 00 & 01 & 00 & 01 & 01 & 01 & 00 \\ 00 & 00 & 00 & 10 & 10 & 21 & 01 & 10 & 10 \\ 00 & 00 & 00 & 02 & 01 & 20 & 12 & 02 & 01 \\ 00 & 10 & 10 & 01 & 00 & 01 & 00 & 11 & 10 \\ 00 & 01 & 01 & 21 & 00 & 21 & 00 & 10 & 01 \end{pmatrix} . \tag{6.14}$$

Note that this is in fact a 6×18 matrix where every two columns that represent one element of the matroid are clustered together forming a supercolumn as was stated in Definition 6.5.2. Thus the ith supercolumn represents the ith element of the matroid.

Using the Monte Carlo method we have successfully found a vector linear solution of size 2 over GF(3) for this network, which gives us an alternative 2-linear repre-

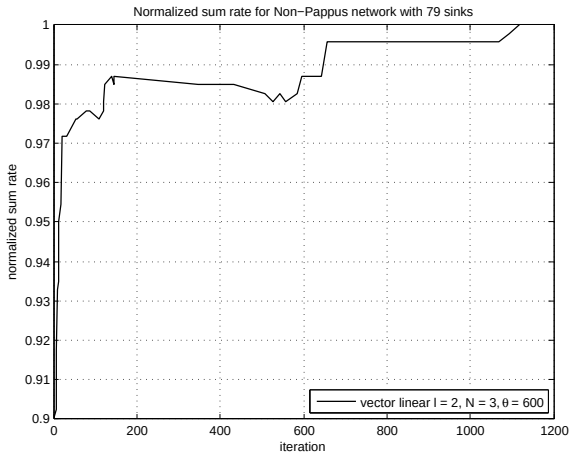


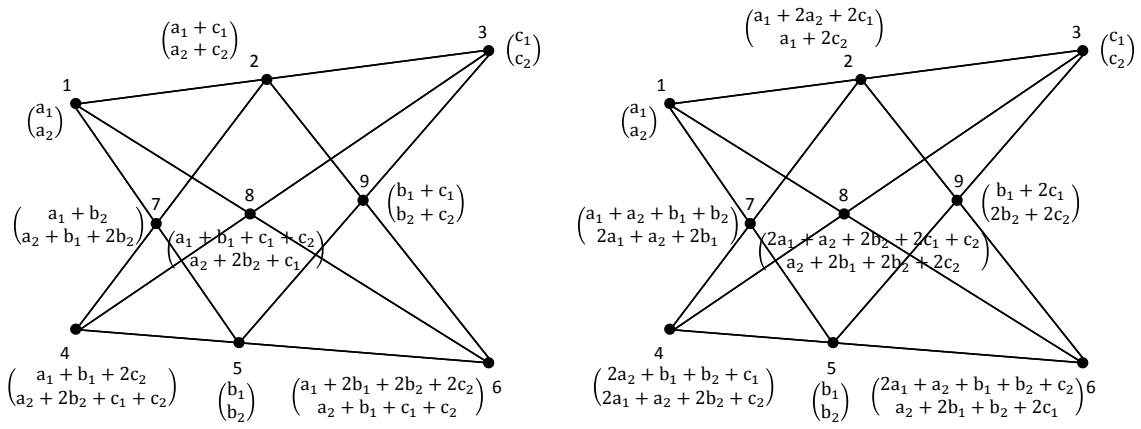
Figure 6.15: Vector linear solution for the non-Pappus network

sentation for the non-Pappus matroid.[1] The relevant simulation can be seen in Fig. 6.15. An example of a representation found for this matroid is stated in the following:

$$\begin{pmatrix} 10 & 11 & 00 & 02 & 00 & 20 & 12 & 20 & 00 \\ 01 & 20 & 00 & 21 & 00 & 11 & 11 & 11 & 00 \\ 00 & 00 & 00 & 10 & 10 & 12 & 12 & 02 & 10 \\ 00 & 00 & 00 & 12 & 01 & 11 & 10 & 22 & 02 \\ 00 & 20 & 10 & 10 & 00 & 02 & 00 & 20 & 20 \\ 00 & 02 & 01 & 01 & 00 & 10 & 00 & 12 & 02 \end{pmatrix}. \tag{6.15}$$

If we denote the elements 1, 5, and 3 of the matroid (see Fig. 6.14) by $\binom{a_1}{a_2}$, $\binom{b_1}{b_2}$, and $\binom{c_1}{c_2}$, correspondingly (since the vector solution is 2 dimensional), then by using the nice way of depiction of [Mat99], the representations (6.14) and (6.15) can be made more clear in Fig. 6.16.

[1]Note that here we want to emphasize the ability of this method to find linear representations, and we do not consider the problem of obtaining a genuinely different representation for the non-Pappus matroid. The issue of determining if two representations of a matroid are equivalent is beyond the scope of our problem (see [Oxl06] in this regard).



(a) 2-representation over GF(3) [SA98, Mat99]

(b) 2-representation over GF(3) via MCMC

Figure 6.16: Multilinear representations for the non-Pappus matroid

6.5.2.2 U_{24} Network

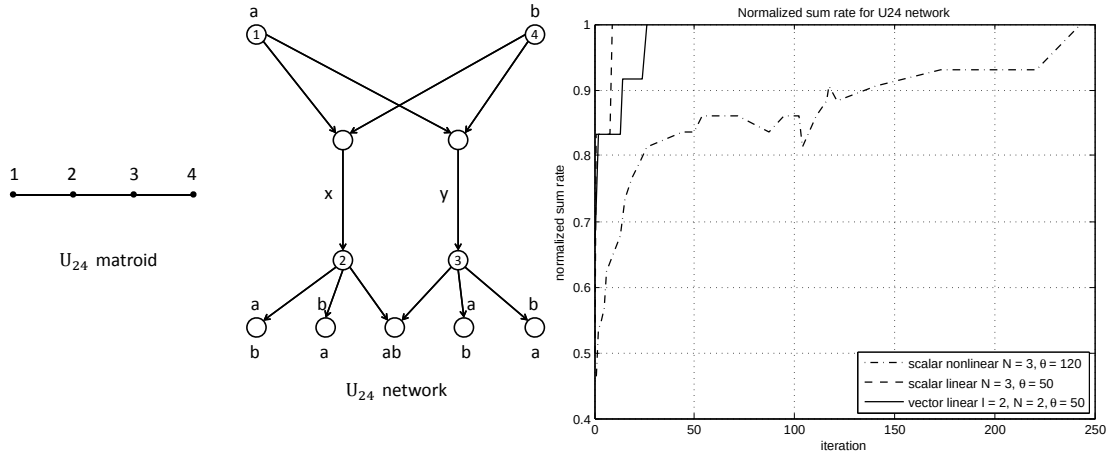
This is also a matroidal network, constructed from the U_{24} matroid, which is the rank 2 uniform matroid on 4 elements. The geometric representation of the matroid is shown in Fig. 6.17a.

Definition 6.5.4 (U_{24} matroid) *Let $E = \{1, 2, 3, 4\}$ be a set of 4 elements. Then define the function r as,*

$$r(\alpha) = \min(|\alpha|, 2). \tag{6.16}$$

Then the U_{24} matroid is the matroid M with ground set E and rank function r. Note that the bases are all the 2-element sets.

Recall from Chapter 4 that the U_{24} matroid plays an important role about the representability of the matroids. In fact it is known that U_{24} is representable over every field except $GF(2)$. Moreover it is the unique matroid that determines the representability of a general matroid over binaries. The following theorem due to Tutte (1958) states this fact [Oxl06]:



(a) U_{24} matroid and network. Numbers inside nodes show the corresponding matroid element

(b) Simulations for the U_{24} network

Figure 6.17: U_{24} matroid, network, and the MCMC simulation

Theorem 6.5.5 *A matroid is representable over binaries if and only if it has no* U_{24} *minor.*

Interestingly, although U_{24} is not representable over binaries, we show that it has a multilinear representation over $GF(2)$. Although it is not too hard to come up with such multilinear representation, we show how using MCMC can quickly give various multilinear representations of this matroid over $GF(2)$. Therefore, first we construct a U_{24} network from this matroid by following the steps of the network construction from a matroid which were previously stated. We have taken the set $B = \{1, 4\}$ as the base, and have used the rest of the circuits and bases of the matroid to build the rest of the network. The resulting network is shown in Fig. 6.17a. Note that all the circuits and bases are imposed in this construction.

Theorem 6.5.6 *The* U_{24} *network has a scalar/vector linear solution over* $GF(N)$ *if and only if the* U_{24} *network has a linear/multilinear representation over* $GF(N)$*.*

Proof: Bases of the U_{24} matroid are $\{\{1,2\},\{1,3\},\{1,4\},\{2,3\},\{2,4\},\{3,4\}\}$, and the set of the circuits of the matroid are $\{\{1,2,3\},\{1,2,4\},\{1,3,4\},\{2,3,4\}\}$. Note

that the demand of the sinks enforce all the base independency conditions. Moreover all the circuit relations are also used in the network. If we denote by X_i the variable corresponding to the matroid element i, and assume that there is linear code defined via the set of A_i's such that $X_i = A_i \begin{pmatrix} a \\ b \end{pmatrix}$, then for any $\alpha \subseteq \{1,2,3,4\}$ where $|\alpha| \leq 2$, we have $\text{rank}(X_\alpha) = 2$. Moreover, for $|\alpha| > 2$, we also get $\text{rank}(X_\alpha) = 2$. Therefore the set of A_i's give a representation for the matroid. It is rather obvious that any representation of the matroid immediately gives a linear solution for the network. $\square$

Based on Theorem 6.5.6, any linear solution that we find for the U_{24} network yields a linear representation for the matroid. Since it is already known that this matroid is not representable over binaries, we have used the MCMC method for alphabet size 3 in the scalar case. More interesting though is the solution of this network over binaries as a vector linear solution of size 2. The simulations can be seen in Fig. 6.17b. Here is a 2-linear representation for the U_{24} matroid

$$\begin{pmatrix} 10 & 00 & 11 & 10 \\ 01 & 00 & 01 & 01 \\ 00 & 10 & 01 & 10 \\ 00 & 01 & 11 & 11 \end{pmatrix}. \tag{6.17}$$

6.6 Distributed MCMC over Networks

We have seen that Monte Carlo methods can be used for entropy optimizations, or in networks to find the best sum rate under certain conditions, or even be used to find linear representations for matroids. In practice, especially for large networks, one would want to employ such methods in a distributed manner. In fact, this is easily done, as described in this section.

Algorithm 1 (Distributed Training "Proto"-Algorithm) *The network operations (truth table or local coefficient matrices) at each internal node are initially set*

to some fixed or random operations. Assume that there are q independent sources, the random variables of the network are vectors of size l, and operations are over GF(N). The algorithm consists of t training epochs. During each training epoch:

1. *Each source transmits a packet of length $N^{q\times l}$, representing one column of a q-input truth table. In conjunction, these $N^{q\times l}$ channel uses per training epoch represent all possible inputs to the network.*

2. *One (or more) internal nodes randomly choose themselves (a la Aloha). A chosen internal node performs a random step on its local truth table (as explained in Section 6.4) and implements the new truth table on the input signals it sees during the training epoch.*

3. *At the end of the training epoch, the sink nodes can compute their recovery rates by computing the mutual information between their received signals and their desired inputs (because they know the transmitted sequences of length $N^{q\times l}$, and have measured the corresponding outputs of the same size).*

4. *These recovery rates are fed back to the network so that every node can compute the new weighted sum rate.*

5. *The chosen internal node(s) compare the new weighted sum rate with the old one, and choose to keep their new truth table according to a "Metropolis" or "simulated annealing" step.*

During the training process all nodes store the largest weighted sum rate encountered, and their respective truth tables corresponding to it. At the end of the t training epochs, the internal nodes set their truth tables to these best-encountered ones. Data transmission at the best rates found can now commence.

When the number of sources in the network is not too large, say $q \leq 8$, and the alphabet size is binary, the packet lengths are not too long, and it is conceivable

that many thousand training epochs can be performed with ease. This will lead to distributed discovery of a network operation with high weighted sum rate.

6.7 Conclusions

In this chapter we proposed a method for numerical optimization of entropy functions over the entropy region. Such method is especially useful in the absence of an explicit characterization of the entropy region, which in fact has proven to be an extremely hard problem. By defining local moves on entropy functions (or rather on their respective distributions) and using Markov Chain Monte Carlo methods (in particular a variant of the Metropolis algorithm), we showed how an optimization may be performed for entropy functions or rate maximization in networks. Moreover we showed how this technique can be employed to study the multilinear representability of matroids as well, which is an interesting problem on its own. Last, but not least, we showed how this numerical framework can be performed in a distributive manner in networks; a scheme that is appealing for practical purposes.

Of course choice of parameters of the numerical algorithm and analysis of the convergence rate, etc., remain unresolved. Nonetheless the method proves to be promising, and we showed its capability through applying it to many different networks.

6.8 Appendix

The following gives a sample linear solution of rate 5/6 found for the Vamos network through the Monte Carlo method (i.e., source variables are of size 5, and the rest of the network variables are of size 6, and demands are fully reconstructed at the sinks). We have used the notation $\bigoplus a_{i_1,\ldots,i_k}$ as a shorthand for $a_{i_1}+\ldots+a_{i_k}$, and all operations are over GF(2). The local and global encoding for each variables is given.

$$w = \begin{pmatrix} \bigoplus a_{1,2,3,4,5} + \bigoplus b_{1,2} + \bigoplus c_{3,4,5} + \bigoplus d_{1,2,3,4,5} \\ a_5 + \bigoplus b_{1,2,4} + c_1 + \bigoplus d_{1,5} \\ a_1 + \bigoplus b_{3,5} + \bigoplus c_{2,3} + \bigoplus d_{1,2,3,4} \\ \bigoplus a_{1,4,5} + \bigoplus b_{1,3,4} + \bigoplus c_{3,5} + \bigoplus d_{1,2,3} \\ \bigoplus a_{1,3,4} + \bigoplus b_{1,2,3,5} + \bigoplus d_5 \\ \bigoplus a_{2,3,4,5} + \bigoplus b_{4,5} + c_5 + \bigoplus d_{1,2,4} \end{pmatrix} \tag{6.18}$$

$$\begin{aligned} x &= \begin{pmatrix} \bigoplus a_{1,4,5} + \bigoplus b_{1,3,4,5} + \bigoplus w_{1,4,5,6} \\ \bigoplus a_{1,3,5} + \bigoplus b_{1,3} + \bigoplus w_{1,\ldots,6} \\ \bigoplus a_{1,2,3,5} + \bigoplus b_{2,3,5} + \bigoplus w_{1,2,3,4,6} \\ \bigoplus a_{2,4} + \bigoplus w_{1,2,5,6} \\ \bigoplus a_{2,4,5} + \bigoplus b_{1,2} + \bigoplus w_{1,2,4} \\ \bigoplus a_{2,3,4,5} + \bigoplus b_{1,4,5} + \bigoplus w_{1,6} \end{pmatrix} \\ &= \begin{pmatrix} \bigoplus a_{3,4} + \bigoplus b_{3,4,5} + \bigoplus c_{4,5} + \bigoplus d_{1,2} \\ \bigoplus a_{1,5} + \bigoplus b_{1,2,4,5} + \bigoplus c_{1,2,3,4,5} + d_{1,3,4,5} \\ \bigoplus a_{2,3,4,5} + \bigoplus b_{1,2,3,4,5} + \bigoplus c_{1,2,3,4,5} + \bigoplus d_{1,3,4} \\ \bigoplus a_{2,3,5} + \bigoplus b_{1,2,3} + \bigoplus c_{1,3,4} + \bigoplus d_{1,3,5} \\ \bigoplus a_{3,4,} + \bigoplus b_{2,3} + \bigoplus c_{1,4} + \bigoplus d_{1,4} \\ \bigoplus a_{1,2,3,4,5} + b_2 + \bigoplus c_{3,4} + \bigoplus d_{3,5} \end{pmatrix} \end{aligned} \tag{6.19, 6.20}$$

$$y = \begin{pmatrix} \bigoplus b_{2,3,4,5} + \bigoplus c_{2,3,4,5} + \bigoplus x_{4,6} \\ b_3 + \bigoplus c_{4,5} + \bigoplus x_{1,2,3} \\ \bigoplus b_{1,2,3,4,5} + \bigoplus c_{1,3,4,5} + x_{1,3,4,6} \\ \bigoplus b_{2,3,4} + \bigoplus c_{2,5} + \bigoplus x_{1,6} \\ \bigoplus b_{1,2,3,4} + \bigoplus c_{2,4} + \bigoplus x_{1,2,5,6} \\ \bigoplus b_{1,2,3} + \bigoplus c_{1,3} + \bigoplus x_{1,4,5,6} \end{pmatrix} \tag{6.21}$$

$$= \begin{pmatrix} \bigoplus a_{1,4} + \bigoplus b_{1,2,4,5} + \bigoplus c_{1,2,3,4,5} + d_1 \\ \bigoplus a_{1,2} + \bigoplus b_{3,4,5} + \bigoplus d_{1,2,5} \\ \bigoplus a_{1,2,4,5} + \bigoplus b_{1,4,5} + \bigoplus c_{1,2,4,5} + \bigoplus d_{1,2,3,4} \\ \bigoplus a_{1,2,5} + b_5 + \bigoplus c_{2,3} + \bigoplus d_{1,2,3,5} \\ \bigoplus a_{2,3,4} + \bigoplus b_{3,4} + c_4 + \bigoplus d_{1,2} \\ \bigoplus a_{1,4} + \bigoplus b_{4,5} + \bigoplus c_{1,3,5} + \bigoplus d_{1,2,4} \end{pmatrix} \tag{6.22}$$

$$z = \begin{pmatrix} \bigoplus c_{1,2} + \bigoplus d_{1,2,4} + \bigoplus y_{2,3,5} \\ \bigoplus c_{2,4} + d_1 + \bigoplus y_{1,2,3} \\ \bigoplus c_{2,3} + \bigoplus d_{3,5} + \bigoplus y_{1,2} \\ \bigoplus c_{2,5} + \bigoplus d_{1,,3} + \bigoplus y_{2,3,4,6} \\ \bigoplus c_{1,2} + \bigoplus d_{2,3,4,5} + \bigoplus y_{1,2,6} \\ \bigoplus c_{1,5,} + \bigoplus d_{1,3,4,5} + \bigoplus y_{2,3,4,5} \end{pmatrix} \tag{6.23}$$

$$= \begin{pmatrix} \bigoplus a_{2,3,5} + \bigoplus b_{1,4} + c_5 + \bigoplus d_{3,5} \\ \bigoplus a_{1,5} + \bigoplus b_{2,3,4,5} + \bigoplus c_{2,3,4} + \bigoplus d_{3,4,5} \\ \bigoplus a_{2,4} + \bigoplus b_{1,2,3} + \bigoplus c_{1,4,5} + \bigoplus d_{2,3} \\ a_2 + \bigoplus b_{1,3,4} + \bigoplus c_{2,4,5} + \bigoplus d_{1,2,3} \\ \bigoplus a_{1,2} + \bigoplus b_{1,2,3,4,5} + \bigoplus c_{1,4} + \bigoplus d_{1,2,3} \\ \bigoplus a_{1,3} + \bigoplus b_{1,4,5} + c_3 + \bigoplus d_{1,3,5} \end{pmatrix}. \tag{6.24}$$

The decoding functions are stated in the following: To decode b from d, x, and z we have

$$b = \begin{pmatrix} \bigoplus d_{1,2,3,4} + \bigoplus x_{1,3} + \bigoplus z_{1,4,5,6} \\ \bigoplus d_{1,3,5} + \bigoplus x_{1,3} + \bigoplus z_{2,5} \\ d_5 + \bigoplus x_{1,4,5,6} + \bigoplus z_{1,2,3,4,5,6} \\ \bigoplus d_{1,2,5} + \bigoplus x_{1,3,4,5} + \bigoplus z_{3,4} \\ \bigoplus d_{2,5} + \bigoplus x_{1,2,3,5,6} + \bigoplus z_{2,4,5} \end{pmatrix}. \tag{6.25}$$

To decode a from b, c, d, and z we have

$$a = \begin{pmatrix} \bigoplus b_{2,5} + \bigoplus c_{1,2,5} + \bigoplus z_{4,5} \\ \bigoplus b_{1,3,4} + \bigoplus c_{2,4,5} + \bigoplus d_{1,2,3} + z_4 \\ b_4 + \bigoplus c_{1,2,3,5} + \bigoplus d_{1,2,3,4} + \bigoplus z_{1,2,5} \\ \bigoplus b_{2,4} + \bigoplus c_{1,2} + d_1 + \bigoplus z_{3,4} \\ \bigoplus b_{3,4} + \bigoplus c_{1,3,4,5} + \bigoplus d_{3,4,5} + \bigoplus z_{2,4,5} \end{pmatrix}. \tag{6.26}$$

For the middle node to decode b and c from a, d, w, and z we get

$$b = \begin{pmatrix} \bigoplus a_{1,3,4} + \bigoplus d_{1,3,4} + \bigoplus w_{1,2,5} + \bigoplus z_{3,6} \\ \bigoplus a_{1,2,5} + d_3 + \bigoplus w_{1,3,4,5,6} + \bigoplus z_{1,2} \\ \bigoplus a_{2,3,5} + \bigoplus d_{1,2,4} + \bigoplus w_{1,3,4} + z_2 \\ \bigoplus a_{3,5} + \bigoplus d_{2,5} + \bigoplus w_{1,3,6} + \bigoplus z_{2,6} \\ \bigoplus a_{1,3} + \bigoplus d_{2,5} + \bigoplus w_{1,2,5,6} + \bigoplus z_{1,3,6} \end{pmatrix} \tag{6.27}$$

$$c = \begin{pmatrix} \bigoplus a_{2,4,5} + \bigoplus d_{2,4} + \bigoplus w_{1,4} + \bigoplus z_{1,3} \\ \bigoplus a_{1,2,4} + \bigoplus d_{2,3,4} + \bigoplus w_{1,2,3,4,5,6} + \bigoplus z_{3,6} \\ \bigoplus a_{1,4,5} + \bigoplus d_{4,5} + \bigoplus w_{1,3} + \bigoplus z_{1,2} \\ \bigoplus a_{1,2,3,5} + \bigoplus d_{1,3} + \bigoplus w_{3,4,5} + \bigoplus z_{1,2,6} \\ \bigoplus a_{1,2,3,4} + \bigoplus d_{1,2,4} + \bigoplus w_{2,3,5,6} + \bigoplus z_{1,2,3} \end{pmatrix}. \tag{6.28}$$

To decode d from a, b, c, and y we have

$$d = \begin{pmatrix} \bigoplus a_{1,4} + \bigoplus b_{1,2,4,5} + \bigoplus c_{1,2,3,4,5} + y_1 \\ \bigoplus a_{1,2,3} + \bigoplus b_{1,2,3,5} + \bigoplus c_{1,2,3,5} + \bigoplus y_{1,5} \\ a_5 + \bigoplus b_{3,4} + \bigoplus c_{2,3} + \bigoplus y_{2,4} \\ \bigoplus a_{1,3} + \bigoplus b_{1,4,5} + \bigoplus c_{1,3,5} + \bigoplus y_{2,3,4,5} \\ \bigoplus a_{1,3,4} + b_5 + c_4 + \bigoplus y_{2,5} \end{pmatrix}. \tag{6.29}$$

And finally, to decode the demand of the last node (i.e., c from a, w, and y) we have

$$c = \begin{pmatrix} \bigoplus a_{2,3} + \bigoplus w_{2,4,6} + \bigoplus y_{1,2,3} \\ \bigoplus a_{2,4} + \bigoplus w_{3,4,6} + \bigoplus y_{2,3,4,5,6} \\ \bigoplus a_{1,5} + \bigoplus w_{2,4} + \bigoplus y_{1,2,3,6} \\ \bigoplus a_{2,3,5} + \bigoplus w_{1,3,4,5} + \bigoplus y_{3,5,6} \\ \bigoplus a_{1,2,3,4} + \bigoplus w_{3,6} + \bigoplus y_{2,4} \end{pmatrix}. \tag{6.30}$$

Chapter 7

Future Work

The field of network information theory still faces many challenges. The framework we presented for obtaining the capacity of network information theory problems in Chapter 2 is based on determining the entropy region. Although the complete characterization of this region for any number of random variables seems to be very ambitious, due to the central role of this region in multiuser information theory, even partial results about this region can have significant consequences for networks. In the following we discuss some of the questions that were raised throughout the research undertaken in this thesis, and some of the problems that need to be addressed in future investigations:

- **Realizing Entropy Vectors:** For the most part in this thesis we focused on determining if a given vector would belong to the entropy region in different scenarios. However an equally important question is how to realize an entropy vector. In other words: If we are given an entropy vector, how can we identify its underlying random variables and their joint probability distribution? While characterizing the entropy region would answer if a certain rate tuple is achievable in a network, realizing an entropy vector yields the coding scheme required to achieve that desired point in the capacity region.

- **Linear Network Coding and Matroid Representability:** As was discussed in Chapter 4, linear network codes are inherently related to the linear

representable entropy region, which is in turn equivalent to the matroid representability. In Chapter 4 we only used matroid representability results over binaries and rank-2 matroids. Leveraging matroid representability results for small finite fields can open up new avenues of research in the area of linear network codes. In particular, as was discussed in Chapter 4, the relevant and important object to study in this regard is the convex cone of matroids (over a certain finite field). Determining this convex cone allows for obtaining the optimal linear network codes over the desired finite field.

- **Region of a Subset of Entries of an Entropy Vector:** The entropy vector of n random variables is of dimension $2^n - 1$, and therefore the region grows exponentially in the number of random variables. Moreover even for a linear representable region, one probably needs exponential number of inequalities to define the region (the number of Shannon inequalities alone is exponential in the number of random variables). Therefore the complexity for all but small networks seems inhibiting. However, appealing to the whole $2^n - 1$ dimensional entropy region may not be necessary for solving a given network. In fact a close look at the topology of most networks reveals that only a small subset of the joint entropies of the network random variables appear in the capacity optimization constraints. As a result the important object to characterize in this case would be the projection of the entropy region onto those subsets of joint entropies. This is equivalent to determining if a given set of values for joint entropies can, in fact, be extended to an actual entropy vector.

- **Entropy Region of Gaussian Random Variables:** Further study of the entropy region of jointly Gaussian random variables would be very interesting, as we conjecture that they are sufficient for describing the whole entropy region of continuous random variables. The demanding task, though, would be to obtain

the convex cone of the minimal number of necessary and sufficient conditions that were pointed out in Chapter 5.

- **Analysis of Numerical Methods for Optimization Over Networks:** The Markov Chain Monte Carlo (MCMC) method that we presented in Chapter 6 seems to be very promising. However, as is the case with many other numerical methods, some parameters of the algorithm should be found heuristically. Moreover, in MCMC methods an important factor is knowing the rate of convergence of the Markov chain. It would be very interesting to study these issues for our algorithm analytically.

There are also some broader problems related to the research in this thesis that are also worth exploring. The following are few instances:

- **Group-Network Codes:** As was discussed in Chapter 3, every entropy vector is asymptotically constructible by a finite group and a set of its subgroups. This fact has been used in the literature to create network codes from groups [Cha07b]. Deploying specific finite groups can potentially yield interesting results [MTH10]. In particular it is conceivable that non-Abelian groups are more powerful than other types of groups, as any Abelian finite group is known to satisfy the Ingleton inequality (a necessary inequality for linear representability).

- **Study of Different Performance Measures:** There are many more performance measures in networks other than achievable rates. For example delay, fairness, and security are becoming more and more important for future systems, and an information theoretic approach toward these measures is still being developed.

- **Wireless Network Information Theory:** As was discussed in Chapter 5, some information inequalities, such as the entropy-power inequality, play an

important role in determining the capacity of wireless networks. Studying the implications of these inequalities in constraining the corresponding entropy region of wireless networks would be very interesting, and can give rise to new results for those networks.

Bibliography

[ACLY00] R. Ahlswede, N. Cai, S. Y. R. Li, and R. W. Yeung. Network information flow. *IEEE Transactions on Information Theory*, 46:1204–1216, July 2000.

[Ber74] P. Bergmans. A simple converse for broadcast channels with additive white Gaussian noise. *IEEE Tran. on Inf. Theory*, 20(3):279–280, March 1974.

[Bla65] N. M. Blachman. The convolution inequality for entropy powers. *IEEE Tran. on Inf. Theory*, 11(2):267–271, April 1965.

[Cay45] A. Cayley. On the theory of linear transformations. In *Cambridge Math. J. 4*, pages 1–16, 1845.

[CC84] M. H. M. Costa and T. M. Cover. On the similarity of the entropy power inequality and the Brunn-Minkowski inequality. *IEEE Tran. on Inf. Theory*, 30(6):837–839, Nov 1984.

[CDH] D. Cullina, A. Dimakis, and T. Ho. Searching for minimum storage regenerating codes. *arXiv:0910.2245v*.

[CGK10] T. Chan, A. Grant, and D. Kern. Existence of new inequalities for representable polymatroids. *IEEE Int. Symp. on Inf. Theory (ISIT)*, pages 1364–1368, 2010.

[Cha01] T. H. Chan. A combinatorial approach to information inequalities. *Communications in Information and Systems*, 1(3):241–254, 2001.

[Cha03] T. H. Chan. Balanced information inequalities. *IEEE Trans. on Information Theory*, 49(12):3261–3267, 2003.

[Cha07a] T. H. Chan. Group characterizable entropy functions. In *IEEE Int. Symp. on Inf. Theory (ISIT)*, pages 506–510, 2007.

[Cha07b] T. H. Chan. Capacity region for linear and Abelian network codes. In *Information Theory and Applications Workshop*, San Diego, CA, 2007.

[CT88] T. M. Cover and J. A. Thomas. Determinant inequalities via information theory. *SIAM J. Matrix Anal. Appl.*, 9(3):384–392, 1988.

[CT91] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., 1991.

[CV93] R. S. Cheng and S. Verdu. On limiting characterizations of memoryless multiuser capacity regions. *IEEE Transactions on Information Theory*, 39:609–612, March 1993.

[CY02] T. H. Chan and R. W. Yeung. On a relation between information inequalities and group theory. *IEEE Trans. on Information Theory*, 48(7):1992–1995, 2002.

[DCT91] A. Dembo, T. M. Cover, and J. A. Thomas. Information theoretic inequalities. *IEEE Tran. on Inf. Theory*, 37(6):1501–1518, Nov 1991.

[DFZ05] R. Dougherty, C. Freiling, and K. Zeger. Insufficiency of linear coding in network information flow. *IEEE Tran. on Inf. Theory*, 51(8):2745–2759, 2005.

[DFZ06a] R. Dougherty, C. Freiling, and K. Zeger. Six new non-Shannon information inequalities. In *IEEE International Symposium on Information Theory (ISIT)*, pages 233–236, 2006.

[DFZ06b] R. Dougherty, C. Freiling, and K. Zeger. The Vamos network. *NETCOD*, 2006.

[DFZ06c] R. Dougherty, C. Freiling, and K. Zeger. Unachievability of network coding capacity. *IEEE Tran. on Inf. Theory*, 52(6):2365–2372, June 2006.

[DFZ07] R. Dougherty, C. Freiling, and K. Zeger. Networks, matroids and non-Shannon information inequalities. *IEEE Trans. on Information Theory*, 53(6):1949–1969, June 2007.

[DFZ10] R. Dougherty, C. Freiling, and K. Zeger. Linear rank inequalities on five or more variables. *arXiv:0910.0284v3*, 2010.

[DGYWR] A. G. Dimakis, P. B. Godfrey, M. Wainwright Y. Wu, and K. Ramchandran. Network coding for distributed storage systems. *IEEE Trans. of Inf. Theory*, 56(9):4539–4551, Sep. 2010.

[DRWS11] A. G. Dimakis, K. Ramchandran, Y. Wu, and C. Suh. A survey on network codes for distributed storage. *Proceedings of IEEE*, 99(3), March 2011.

[EF09] E. Erez and M. Feder. Improving the multicommodity flow rates with networks codes for two sources. *IEEE Journal on Selected Areas in Comm.*, 27(5):814–824, 2009.

[EFS56] P. Elia, A. Feinstein, and C. E. Shannon. Note on maximum flow through a network. *IRE Transactions on Information Theory*, 2:117–119, 1956.

[ETW08] R. H. Etkin, D. N. C. Tse, and H. Wang. Gaussian interference channel capacity to whithin one bit. *IEEE Tran. on Inf. Theory*, 54(12):5534–5562, Dec. 2008.

[FF56] L. R. Ford and D. R. Fulkerson. Maximal flow through a network. *Canadian Journal of Mathematics*, 8:399–404, 1956.

[FJ00] S. M. Fallat and C. R. Johnson. Determinantal inequalities: Ancient history and recent advances. *Contemporary Mathematics*, 259:199–211, 2000.

[FSH08] D. Fong, S. Shadbakht, and B. Hassibi. On the entropy region and the Ingleton inequality. In *Intl Symp on Mathematical Theory of Networks and Systems (MTNS)*, 2008.

[Fuj78] S. Fujishije. Polymatroidal dependence structure of a set of random variables. *Information and Control*, 39:55–72, 1978.

[GK00] P. Gupta and P. R. Kumar. The capacity of wireless networks. *IEEE Transactions on Information Theory*, 46:388–404, March 2000.

[GKZ94] I. M. Gelfand, M. M. Kapranov, and A. V. Zelevinsky. *Discriminants, Resultants and Multidimensional Determinants*. Mathematics: Theory and Applications, 1994.

[GSV06] D. Guo, S. Shamai (Shitz), and S. Verdu. Proof of entropy power inequalities via mmse. *IEEE Int. Symp. on Inf. Theory (ISIT)*, pages 1011–1015, July 2006.

[GT06a] K. Griffin and M. J. Tsatsomeros. Principal minors, part I: A method for computing all the principal minors of a matrix. *Linear Algebra and Its Applications*, 419(1):107–124, 2006.

[GT06b] K. Griffin and M. J. Tsatsomeros. Principal minors, part II: The principal minor assignment problem. *Linear Algebra and Its Applications*, 419:125–171, 2006.

[Han75] T. S. Han. Linear dependence structure of the entropy space. *Information and Control*, 29:337–368, 1975.

[Han81] T. S. Han. A uniqueness of Shannon's information distance and related nonnegativity problems. *J. Comb., Inform. Syst. Sci.*, 6(4):320–331, 1981.

[HJ] H. T. Hall and C. R. Johnson. Bounded ratios of products of principal minors of positive definite matrices. *arXiv:0806.2645v1.*

[HRSV00] D. Hammer, A. E. Romashchenko, A. Shen, and N. K. Vereshchagin. Inequalities for Shannon entropy and Kolmogorov complexity. *Journal of Computer and System Sciences*, 60(2):442–464, April 2000.

[HS07a] B. Hassibi and S. Shadbakht. Normalized entropy vectors, network information theory and convex optimization. In *Information Theory Workshop*, Bergen, Norway, 2007.

[HS07b] O. Holtz and B. Sturmfels. Hyperdeterminantal relations among symmetric principal minors. *Journal of Algebra*, 316:634–648, 2007.

[HSYY08] P. Huggins, B. Sturmfels, J. Yu, and D. S. Yuster. The hyperdeterminant and triangulations of the 4-cube. *Mathematics of Computation*, 77(263):1653–1679, 2008.

[Ing71] A. W. Ingleton. Representation of matroids. In *Combinatorial Mathematics and Its Applications, D. Welsh, Ed. London: Academic Press*, pages 149–167, 1971.

[JB93] C. R. Johnson and W. W. Barrett. Determinantal inequalities for positive definite matrices. *Discrete Mathematics*, 119:97–106, 1993.

[JE10] S. Jalali and M. Effros. On the separation of lossy source-network coding and channel coding in wireline networks. *IEEE Int. Symp. on Inf. Theory (ISIT)*, 2010.

[JS98] M. Jerrum and G. Sorkin. The metropilis algorithm for graph bisection. *Discrete Applied Mathematics*, 82:155–175, 1998.

[KEM09] R. Koetter, M. Effros, and M. Médard. On the theory of network equivalence. *IEEE Information Theory Workshop (ITW)*, 2009.

[Kin09] R. Kinser. New inequalities for subspace arrangements. *arXiv:0905.1519v3*, 2009.

[KM03] R. Koetter and M. Médard. An algebraic approach to network coding. *IEEE Transactions on Networking*, 11:782–795, October 2003.

[Kra03] G. Kramer. Capacity results for the discrete memoryless network. *IEEE Transactions on Information Theory*, 49:4–21, January 2003.

[LM07] R. Lněnička and F. Matúš. On Gaussian conditional independence structures. *Kybernetika*, 43(3):327–342, 2007.

[Lne03] R. Lnenička. On the tightness of the Zhang-Yeung inequality for Gaussian vectors. *Communications in Information and Systems*, 3(1):41–46, 2003.

[LT03] J-G Luque and J-Y Thibon. Polynomial invariants of four qubits. *Phys. Rev. A 67, 042303*, 2003.

[LYC03] S. Y. R. Li, R. W. Yeung, and N. Cai. Linear network coding. *IEEE Transactions on Information Theory*, 49:371–381, February 2003.

[LYCH78] S. Leung-Yan-Cheong and M. Hellman. The Gaussian wire-tap channel. *IEEE Tran. on Inf. Theory*, 24(4):451–456, Jul. 1978.

[Mac03] D. J. C. Mackay. *Information Theory, Inference, and Learning Algorithms*. Cambridge, 2003.

[Mat99] F. Matúš. Matroid represenatations by partitions. *Descrete Mathematics*, 203:169–194, 1999.

[Mat07a] F. Matúš. Infinitely many information inequalities. In *Intl. Symp. on Inf. Theory*, 2007.

[Mat07b] F. Matúš. Two constructions on limits of entropy functions. *IEEE Trans. on Information Theory*, 53(1):320–330, 2007.

[MEKH03] M. Médard, M. Effros, D. Karger, and T. Ho. On coding for non-multicast networks. *41st Allerton Conf. on Comm., Control, and Comp.*, Monticello, IL, Oct. 2003.

[MH09] W. Mao and B. Hassibi. Violating the Ingleton inequality with finite groups. *46th Allerton Conf. on Comm., Control, and Comp.*, 2009.

[MMRV02] K. Makarychev, Y. Makarychev, A. Romashchenko, and N. Vereshchagin. A new class of non-Shannon-type inequalities for entropies. *Communications in Information and Systems*, 2(2):147–166, 2002.

[MS95] F. Matúš and M. Studeny. Conditional independences among four random variables I. *Combin., Prob. Comput.*, 4:269–278, 1995.

[MT10] M. Madiman and P. Tetali. Information inequalities for joint distributions, with interpretations and applications. *IEEE Tran. on Inf. Theory*, 56(6):2699–2713, June 2010.

[MTH10] W. Mao, M. Thill, and B. Hassibi. On group network codes: Ingleton-bound violations and independent sources. *IEEE Int. Symp. on Inf. Theory (ISIT)*, 2010.

[Oxl04] J. Oxley. What is a matroid? *Workshop on Combinatorics and its Applications*, July 2004.

[Oxl06] J. G. Oxley. *Matroid Theory*. Oxford University Press, 2006.

[POR] http://www2.iwr.uni-heidelberg.de/groups/comopt/software/porta/.

[PR04] M. Passare and H. Rullgård. Amoebas. Monge-Ampère measures, and triangulations of the Newton polytope. *Duke Math. J.*, 121(3):481–507, 2004.

[RSG] S. El Rouayheb, A. Sprintson, and C. Georghiades. On the relation between the index coding and the network coding problems. *arXiv:0802.0179v2*.

[SA98] J. Simonis and A. Ashikhmin. Almost affine codes. *Designs, Codes and Cryptography*, 14:179–197, 1998.

[Sha48] C. E. Shannon. A mathematical theory of communication. In *Bell System Technical Journal*, volume 27, pages 623–656, 1948.

[Sha61] C. E. Shannon. Two-way communications channels. In *Proceedings of the 4th Berkeley Symposium on Mathematical Staistics and Probabililty*, pages 611–644. University of California Press, 1961.

[Shl52] L. Shläfli. Über die resultante eines systemes mehrerer algebraischen gleichungen. *Denkschr. der Kaiserlichen Akad. der Wiss., Math-Naturwiss. Klasse*, 4, 1852.

[SJH09] S. Shadbakht, A. Jafarian, and B. Hassibi. Scalar linear network coding for networks with two sources. *Intl Conf. on Comm., Cape Town, South Africa*, Sep 2009.

[Smi84] M. C. Smith. Applications of algebraic function theory in multivariable control. *Multivariable control: New concepts and tools, S.G. Tzafestas, Ed., Reidel Pub.*, 1984.

[Sta59] A. J. Stam. Some inequalities satisfied by the quantitties of information of Fisher and Shannon. *Information and Control*, 2:101–112, June 1959.

[SYC06] L. Song, R. W. Yeung, and N. Cai. A separation theorem for single-source network coding. *IEEE Tran. on Inf. Theory*, 52(5):1861–1871, May 2006.

[TW09] S. P Tsarev and T. Wolf. Hyperdeterminants as integrable discrete systems. *J. Phys. A: Math. Theor.*, 42(45), 2009.

[vdM77] E. C. van der Meulen. A survey of multi-way channels in information theory: 1961-1976. *IEEE Transactions on Information Theory*, 23:1–37, January 1977.

[VG06] S. Verdu and D. Guo. A simple proof of the entropy-power inequality. *IEEE Tran. on Inf. Theory*, 52(5):2165–2166, May 2006.

[WS10] C-C Wang and N. B. Shroff. Pairwise intersession network coding on directed networks. *IEEE Tran. on Inf. Theory*, 56(8):3879–3900, 2010.

[WSS06] H. Weingarten, Y. Steinberg, and S. Shamai. The capacity region of the Gaussian multiple-input multiple-output broadcast channel. *IEEE Tran. on Inf. Theory*, 52(9):3936–3964, Sept 2006.

[WW09] J. M. Walsh and S. Weber. A recursive construction of the set of binary entropy vectors and related inner bounds for the entropy region. *IEEE Tran. on Inf. Theory*, submitted 2009.

[Yeu91] R. W. Yeung. A new outlook on Shannon's information measures. *IEEE Tran. on Inf. Theory*, 37(3):466–474, May 1991.

[Yeu97] R. W. Yeung. A framework for linear information inequalities. *IEEE Trans. on Information Theory*, 43(6):1924–1934, 1997.

[Yeu02] R. W. Yeung. *A first course in information theory.* Kluwer, 2002.

[YLCZ06] R. W. Yeung, S. Y. R. Li, N. Cai, and Z. Zhang. *Network Coding Theory.* now Publishers, Inc., 2006.

[YYZ07] X. Yan, R. W. Yeung, and Z. Zhang. The capacity region for multi-source multi-sink network coding. *IEEE Int. Symp. on Inf. Theory (ISIT)*, pages 116–120, 2007.

[YZ99] R. W. Yeung and Z. Zhang. Distributed source coding for satellite communications. *IEEE Trans. on Information Theory*, 45(4):1111–1120, 1999.

[Zha03] Z. Zhang. On a new non-Shannon type information inequality. *Communications in Information and Systems*, 3(1):47–60, 2003.

[ZY97] Z. Zhang and R. Yeung. A non-Shannon-type conditional inequality of information quantities. *IEEE Trans. on Information Theory*, 43(6):1982–1986, 1997.

[ZY98] Z. Zhang and R. Yeung. On characterization of entropy function via information inequalities. *IEEE Trans. on Information Theory*, 44(4):1440–1452, 1998.