ON MATRIX CHANNELS FOR NETWORK CODING

Jessica Claridge

Royal Holloway, University of London

Thesis submitted to
The University of London
for the degree of
Doctor of Philosophy
2017.

Declaration

These doctoral studies were conducted under the supervision of Professor Simon R. Blackburn.

The work presented in this thesis is the result of original research carried out by myself, in collaboration with others, whilst enrolled in the School of Mathematics and Information Security as a candidate for the degree of Doctor of Philosophy. This work has not been submitted for any other degree or award in any other university or educational establishment.

Jessica Claridge
September 2016

Abstract

In 2000, Ahlswede, Cai and Li introduced network coding, a technique used to improve the efficiency of information flow through networks by allowing intermediate nodes to compute with and modify data. In practice random linear network coding is used, where the nodes transmit random linear combinations of their incoming packets. This thesis is concerned with several mathematical problems motivated by network coding.

We first consider partial decoding in random linear network coding. By noting the equivalence to an enumeration problem in linear algebra, we compute the exact probability of a receiver decoding a fraction of the source message. We investigate the consequences when using both systematic and non-systematic network coding.

We then consider mathematical models for network coding. Silva, Kschischang and Kötter studied certain classes of finite field matrix channels in order to model random linear network coding where exactly t random errors are introduced. We introduce a generalisation of these channels that allow the modelling of channels where a variable number of random errors are introduced. For special cases of our channel we improve on previous analysis of the channel capacity.

For the general case we show that a capacity-achieving input distribution can always be taken to have a very restricted form (the distribution should be uniform given the rank of the input matrix). Nobrega, Silva and Uchoa-Filho proved a similar result for a class of matrix channels that model network coding with link erasures. Our result leads to an expression for the capacity of these channels as a maximisation over probability distributions on the set of possible ranks of input matrices (rather than the set of all input matrices): a set of linear rather than exponential size. Thus we give an efficient method to find optimal input distributions and compute the exact capacity for any channel parameters.

Acknowledgements

Firstly, I would like to thank Prof. Simon Blackburn for being the most supportive supervisor I could possibly hope for. His vast knowledge and wisdom has been a constant source of inspiration for me and his guidance, advice and encouragement have been invaluable. I will forever be grateful to him for making my time as a post-graduate so rich and enjoyable.

I would also like to thank Dr. Ioannis Chatzigeorgiou for coming to Royal Holloway and initiating our collaboration. Collaborating with him has enriched my researching experience and opened my eyes to a new area of problems.

I would like to thank my PhD examiners, Dr. Maximilien Gadouleau and Dr. Siaw-Lynn Ng, for their extremely useful comments and suggestions and for making my viva a truly enjoyable discussion of my work.

I would like to acknowledge the EU COST Action IC1104 and the EPSRC. I am very grateful for their generous financial support and the opportunities they made available to me.

Last but not least I would like to thank my family and friends for their immense support throughout this journey. In particular, my mum, Marit, for proof reading my thesis (whilst skipping the mathematics) and Sam for the continual supply of peppermint tea through many long nights of work.

Contents

Chapter 1

Introduction

1.1 Overview

The essential foundation of communication networks is mathematics. This thesis is concerned with several linear algebra problems that are motivated by network coding. The application of this work is explored in Chapter 2 where we define network coding and review the literature.

The linear algebra problems we consider involve systems of equations and matrix channels over finite fields. In this chapter we define these problems and give an overview of our results. Section 1.2 defines notation we adopt throughout this work. Section 1.3 outlines our results on systems of equations. We then move on to matrix channels: Section 1.4 defines the *Multiplicative Matrix Channel (MMC)*; Section 1.5 defines the *Additive Matrix Channel (AMC)* and Section 1.6 defines the *Generalised Additive Multiplicative MAtrix channel (Gamma channel)*. Finally in Section 1.7 we give an overview of the structure of the thesis.

1.2 Notation

Let q be a prime power and $\mathbb{F}_q$ be the finite field of q elements. We write $\mathbb{F}_q^{n\times m}$ to denote the set of all $n \times m$ matrices over $\mathbb{F}_q$, and $\mathbb{F}_q^{n\times m,r}$ to denote the set of matrices in $\mathbb{F}_q^{n\times m}$ of rank r. We denote the set of all invertible matrices in $\mathbb{F}_q^{n\times n}$ (the general linear group) by $\mathrm{GL}(n, q)$.

The following definition gives notation which is due to Nobrega, Silva and Uchoa-Filho [33].

Definition 1.2.1. Let $\mathcal{R}$ be a probability distribution on the set $\{0, 1, \ldots, \min\{m, n\}\}$ of possible ranks of matrices $M \in \mathbb{F}_q^{n\times m}$. We define a distribution on the set $\mathbb{F}_q^{n\times m}$ of matrices by choosing r according to $\mathcal{R}$, and then once r is fixed choosing a matrix $M \in \mathbb{F}_q^{n\times m,r}$ uniformly at random. We say that this distribution is *Uniform Given Rank (UGR) with rank distribution* $\mathcal{R}$. We say a distribution on $\mathbb{F}_q^{n\times m}$ is *Uniform Given Rank (UGR)* if it is UGR with rank distribution $\mathcal{R}$ for some distribution $\mathcal{R}$.

We write $\mathcal{R}(r)$ for the probability of rank r under the distribution $\mathcal{R}$. So a distribution on $\mathbb{F}_q^{n\times m}$ is UGR with rank distribution $\mathcal{R}$ if and only if each $M \in \mathbb{F}_q^{n\times m}$ of rank r is chosen with probability $\mathcal{R}(r)/|\mathbb{F}_q^{n\times m,r}|$.

1.3 Linear systems of equations

In this section we summarise the mathematical result of Chapter 4, which is concerned with partially solving systems of random linear equations.

Consider a random linear system of r linearly independent equations in k unknowns over a finite field. Suppose the system is expressed in the matrix form

$$\boldsymbol{M}\underline{v} = \underline{u}, \tag{1.3.1}$$

where $\boldsymbol{M}$ is a random full rank $r \times k$ matrix, $\underline{v} = (v_1, \ldots, v_k)$, and $\underline{u}$ is a constant vector of length r. Given that (1.3.1) is consistent, it is possible to determine the i-th unknown v_i if and only if the i-th unit vector is contained in the rowspace of $\boldsymbol{M}$.

In Chapter 4, invoking the principle of inclusion and exclusion, we derive an exact expression for the probability that the rowspace of a random $r \times k$ matrix $\boldsymbol{M}$ contains at least some fixed number x unit vectors, for $x \leq r \leq k$. This is equal to the probability of determining the values of at least x of the k unknowns of the system in (1.3.1).

Drawing parallels between partially solving systems of random linear equations and partial decoding in random linear network coding, we analyse the implications of our result to this application (see Chapter 2 for further details of the application).

1.4 The multiplicative matrix channel

In this section we define the *Multiplicative Matrix Channel* and summarise the results of our analysis in Chapter 5.

Definition 1.4.1. The *Multiplicative Matrix Channel* (MMC) has input set $\mathcal{X}$ and output set $\mathcal{Y}$, where $\mathcal{X} = \mathcal{Y} = \mathbb{F}_q^{n \times m}$. The channel law is

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}$$

where $\boldsymbol{A} \in \mathrm{GL}(n, q)$ is chosen uniformly at random.

We assume the values q, n, m are fixed by the application. We refer to these values as the *channel parameters* of the MMC channel.

The MMC channel is considered by Silva, Kschischang and Kötter [39, §III], the authors use it to model random linear network coding with no errors (see

Chapter 2 for further details). In [39] the authors give bounds on the capacity of the MMC channel that converge for large field size or large channel input. We improve on this analysis, giving bounds on the capacity that differ by a small additive constant for any value of the channel parameters, thus giving insight on the behaviour of the capacity for all channel parameters.

1.5 The additive matrix channel

In this section we define the *Additive Matrix Channel* and summarise the results of our analysis in Chapter 6.

Definition 1.5.1. Let $\mathcal{R}$ be a probability distribution on the set $\{0, 1, \ldots, \min\{m, n\}\}$ of possible ranks of matrices $M \in \mathbb{F}_q^{n\times m}$. The *Additive Matrix Channel with rank error distribution* $\mathcal{R}$ *(*AMC($\mathcal{R}$)*)* has input set $\mathcal{X}$ and output set $\mathcal{Y}$, where $\mathcal{X} = \mathcal{Y} = \mathbb{F}_q^{n\times m}$. The channel law is

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{B}$$

where $\boldsymbol{B} \in \mathbb{F}_q^{n\times m}$ is UGR with rank distribution $\mathcal{R}$.

We assume that q, n, m and $\mathcal{R}$ are fixed by the application. We refer to these values as the *channel parameters* of the AMC($\mathcal{R}$) channel.

In Chapter 6 we consider two particular rank distributions $\mathcal{R}$ that give special cases of the AMC channel. Firstly we consider the case when the rank of the error matrix $\boldsymbol{B}$ is fixed and equal to t, *the* AMC *channel with fixed error rank.* This is exactly the channel considered by Silva, Kschischang and Kötter [39, §IV], which the authors use to model coherent random linear network coding (see Chapter 2 for further details). Next we consider the case when $\mathcal{R}$ is chosen to ensure that the error matrix $\boldsymbol{B}$ has a uniform distribution over all $n \times m$ matrices of rank $\leq t$, *the* AMC *channel with uniform error.* We show that the capacity of the AMC channel with uniform error gives a lower bound for the capacity of the general AMC channel.

For both of these special cases of the AMC channel we present bounds on the capacity that differ by a small additive constant for any values of q, n, m, t. This improves on the work of [39], that gives bounds on the capacity of the AMC channel with fixed error rank that only converge to a close value for large field size or large channel input. The lower bound for the AMC channel with uniform error capacity gives an immediate lower bound for the capacity of the general AMC channel when the rank of $\boldsymbol{B}$ is bounded by t. Our results show that the minimum capacity of the general AMC channel is very close to the capacity of the channel with fixed error rank, thus our generalisation of the channel from [39, §IV] covers a wider class of channels without any significant loss in capacity.

1.6 The Gamma channel

In this section we define the *Gamma channel* and summarise the results of our analysis in Chapter 8.

Definition 1.6.1. Let $\mathcal{R}$ be a probability distribution on the set $\{0, 1, \ldots, \min\{m, n\}\}$ of possible ranks of matrices $M \in \mathbb{F}_q^{n\times m}$. The *Generalised Additive Multiplicative MAtrix Channel with rank error distribution* $\mathcal{R}$ *(the Gamma channel* $\Gamma(\mathcal{R})$*)* has input set $\mathcal{X}$ and output set $\mathcal{Y}$, where $\mathcal{X} = \mathcal{Y} = \mathbb{F}_q^{n\times m}$. The channel law is

$$\boldsymbol{Y} = \boldsymbol{A}(\boldsymbol{X} + \boldsymbol{B})$$

where $\boldsymbol{A} \in \mathrm{GL}(n, q)$ is chosen uniformly, where $\boldsymbol{B} \in \mathbb{F}_q^{n\times m}$ is UGR with rank distribution $\mathcal{R}$, and where $\boldsymbol{A}$ and $\boldsymbol{B}$ are chosen independently.

We assume that q, n, m and $\mathcal{R}$ are fixed by the application. We refer to these values as the *channel parameters* of the Gamma channel $\Gamma(\mathcal{R})$).

Both the MMC and AMC channels can be seen as special cases of the Gamma channel. Indeed the MMC channel is the Gamma channel $\Gamma(\mathcal{R})$, when $\mathcal{R}$ is the rank distribution that fixes the rank of $\boldsymbol{B}$ to be zero; and the AMC channel is the Gamma channel when the matrix $\boldsymbol{A}$ is the identity matrix, or equivalently when $\boldsymbol{A}$ is known to the receiver and therefore can be removed by multiplying by $\boldsymbol{A}^{-1}$.

Silva, Kschischang and Kötter [39, §V] considered the special case of the Gamma channel when the error matrix $\boldsymbol{B}$ has fixed rank t. The authors used this channel to model the general case of random linear network coding (see Chapter 2) and presented bounds on the capacity that converge for large field size or large channel input. However, the exact capacity of the Gamma channel is hard to determine due to the many degrees of freedom involved: the naive formula maximises over a probability distribution on the set of possible input matrices, and this set is exponentially large.

In Chapter 8 we show that a capacity-achieving input distribution can always be taken to be UGR. In 2013 Nobrega, Silva and Uchoa-Filho [33] proved a similar result for a class of matrix channels that model network coding with link erasures (see Chapter 2 for further details) and our result generalises theirs to a new class of channels. We use our result to express the Gamma channel capacity as a maximisation over probability distributions on the set of possible ranks of input matrices: a set of linear rather than exponential size. This gives an efficient way to compute the exact channel capacity and find an optimal input distribution for any channel parameters.

1.7 Structure of the thesis

The remainder of the thesis is structured as follows. Chapter 2 motivates this work, reviewing the literature on network coding and describing the application of the matrix channels to network coding. Chapter 3 presents preliminary mathematical results needed in the work that follows. Chapter 4 considers the problem of partial decoding in random linear network coding. Chapter 5 analyses the MMC channel and Chapter 6 analyses the AMC channel. Chapter 7 derives expressions for several matrix functions which are needed in the analysis of the Gamma channel in Chapter 8. Finally Chapter 9 discusses the results of the thesis and considers future research.

Chapter 2

Motivation

2.1 Overview

In this chapter we introduce network coding and show how the mathematical topics of this thesis relate to this application. We begin in Section 2.2 by defining network coding and reviewing some key results from the literature. In Section 2.3 we discuss partial decoding in network coding, motivating the topics of Chapter 4. In Section 2.4 we show how the matrix channels defined in Chapter 1 can be used to model random linear network coding in a variety of situations. In turn we describe the relevance of the MMC, AMC and Gamma channels to this application. We end Section 2.4 with a discussion of appropriate error rank distributions $\mathcal{R}$ for the AMC and Gamma channels.

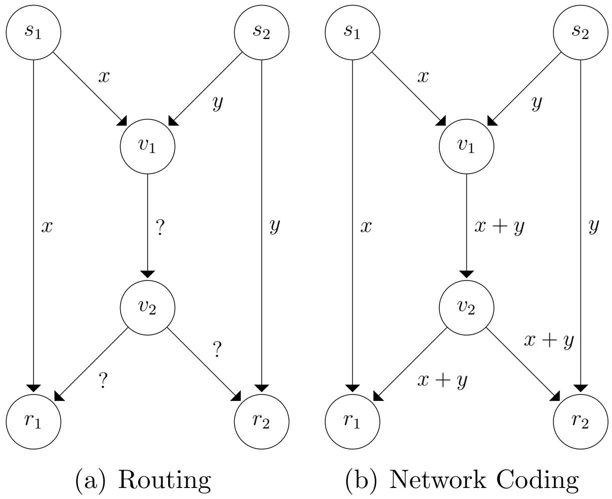
2.2 Network coding

In communication networks data has been traditionally viewed as a physical entity that must be routed through the network from a source to a receiver (*routing*). In this setting intermediate network nodes can only perform very limited operations of storing and replicating data. One may think of data as a fluid that flows through the network. Then intermediate nodes can be seen as ‘gates’ that either open, allowing the fluid to pass through, or remain closed. Whilst this may seem intuitively reasonable there is one big assumption: in order to obtain the original message, the receiving node (or sink) must receive (possibly subject to errors) an exact replica of the data sent.

In 2000 Ahlswede, Cai, Li and Yeung [1] presented an alternative approach: *network coding*. Exploiting the fact that data is not a physical entity (and doesn’t behave like a fluid), network coding allows *coding* at intermediate nodes, that is one can compute with and modify data as it travels through a network. Packets of data are injected into the network at the source (*source packets*). Intermediate nodes compute a function of their received packets and forward the resulting packet. The aim is no longer for the actual source packets to be received, instead it suffices to receive ‘evidence’ of the original packets. The evidence takes the form of the coded packets, from which the receiver must be able to recover the data.

The fundamental result of [1] is that network coding can increase the network throughput, that is the rate of information flow, when compared with routing.

The simplest network illustrating this benefit is the *butterfly network*, Fig. 2.1. The butterfly network has two sources s_1, s_2 and two sinks r_1, r_2. The sources s_1 and s_2 have packets x and y, respectively, and both sinks r_1, r_2 request both x and y. Each edge in the network can transmit one packet at a time. Using routing alone, Fig. 2.1(a) shows that it is impossible to satisfy the request of both sinks simultaneously: there is a bottleneck at the node v_1, which must choose between forwarding packet x or y. Therefore routing requires two uses of the edge $v_1 \to v_2$ and hence there is a delay. Network coding allows the node v_1 to compute the component-wise addition of x and



(a) Routing (b) Network Coding

Figure 2.1: The Butterfly Network

y, which then forwards the coded packet $x + y$. The result is that r_1 receives $x, x + y$, r_2 receives $x + y, y$ and both sinks can recover x and y (for example, r_1 computes y as $y = (x + y) - x$). The requests of both r_1 and r_2 have been satisfied with just one use of each edge in the network, hence network coding has increased the network throughput.

The benefit of allowing network coding can be seen clearly from the butterfly network example. However, real world networks can be extremely large with complex topology. In an arbitrary network, how do the intermediate nodes know what coding operations to perform? Do efficient schemes exist that allow for nodes with restrictions on power and storage? These questions become extremely complex for even moderate sized networks. However, it has been shown that for a large class of networking problems a significant simplification can be made without loss of throughput: coding is restricted to linear operations.

Linear network coding is the special case of network coding where packets

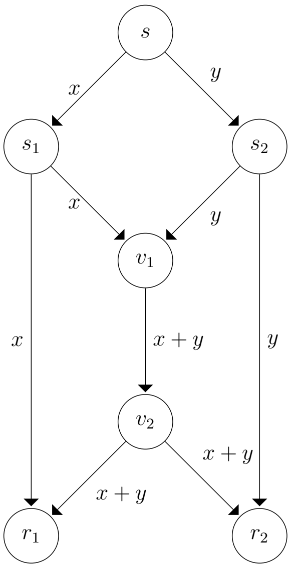


Figure 2.2: The single source variation of the butterfly network.

are thought of as vectors over some finite field and the coding at intermediate nodes is restricted to computing linear combinations of packets. As linear computation requires a small number of operations, linear network coding allows for low power intermediate nodes. In 2003 Li, Yeung and Cai [28] showed that linear network coding is sufficient to maximise network throughput in *multicast* problems, that is when there is one source with a data set that is to be transmitted in full to each member of a set of receivers. Building on this, later in 2003 Kötter and Médard [27] presented an algebraic framework for linear network coding.

Network problems with multiple sources and a set of receivers each wanting the full data set are called *multisource multicast* problems. Given a network with multiple sources, there is an equivalent network with a single source obtained by adding a superior node with one edge for each source packet directed to that packet's original source. For example the single source variation of the butterfly network is shown in Fig. 2.2 (this was the motivating example for network coding given in [1]). Therefore, when considering multisource multicast problems one may simplify to multicast problems with a single source without any loss of generality.

In the single source set up, also in 2003, Chou, Wu and Jain [11] took a practical look at network coding, presenting a robust coding system. The authors suggested transmitting packets with coding headers, used to record the particular linear combination of the source packets in each received packet. This system allows the receiver to have no knowledge of the network or coding coefficients, and decoding becomes straightforward. This practical approach has been adopted in much subsequent work e.g. [23], [32], [43].

In 2006 Ho *et al.* [23] investigate *random linear network coding*, allowing the coefficients of linear combinations to be chosen at random. This method means that intermediate nodes require no knowledge of the network and no storage ability as they simply compute and forward random linear combinations of their incoming packets. This allows network coding to be viewed as an end-to-end system, overcoming the complexity of uncertain network topology. The remarkable result of [23] is that for general multicast problems, random linear network coding achieves maximum network throughput with probability exponentially approaching 1 with the field size. Thus given that the field size is large, random linear network coding can be used without any loss in throughput.

To illustrate this result for the butterfly network (Fig 2.1), we note that the computation of any linear combination of x and y at node v_1 would be sufficient, as long as the coefficients of both x and y are non-zero. Indeed, this is the case with high probability if the field size is large.

Random linear network coding is efficient and simple to implement, but how does one deal with the possibility of errors? Due to packet mixing, if even a single error is injected into the network, this has the potential to corrupt all received packets. This phenomenon of error propagation in network coding means that classic error correction techniques will not work. However, all is not lost as the errors on the received packets are not independent of each other: they are all simply linear combinations of the injected error packets. Therefore the errors are contained in a subspace of small dimension (the dimension of the error subspace is equal to the number of linearly independent errors introduced into the network).

The following example is used to illustrate how errors affect the received packets when using random linear network coding. Consider the single source, single sink network in Fig. 2.3. Each edge in the network transmits one packet at a time. The source s transmits source packets x, y and z to the receiver at the sink r, using random linear network coding. Assuming no errors, the network throughput is shown in Fig. 2.3. The α_i are the random coding coefficients. The receiver r obtains the following linear combinations of the source packets

$$\begin{cases} (\alpha_3 + \alpha_4\alpha_1)x + \alpha_4\alpha_2 y \\ \alpha_1 x + \alpha_2 y \\ \alpha_5\alpha_1 x + \alpha_5\alpha_2 y + \alpha_6 z \end{cases}.$$

Suppose now that during the transmission two erroneous packets e and f were injected into the network. Suppose e was injected at the node v_1 and f was injected at the node v_6. Random linear network coding proceeds as before and the network throughput is shown in Fig. 2.4. The β_i are the new random coding coefficients introduced.

Now the receiver r obtains packets that are linear combinations of the

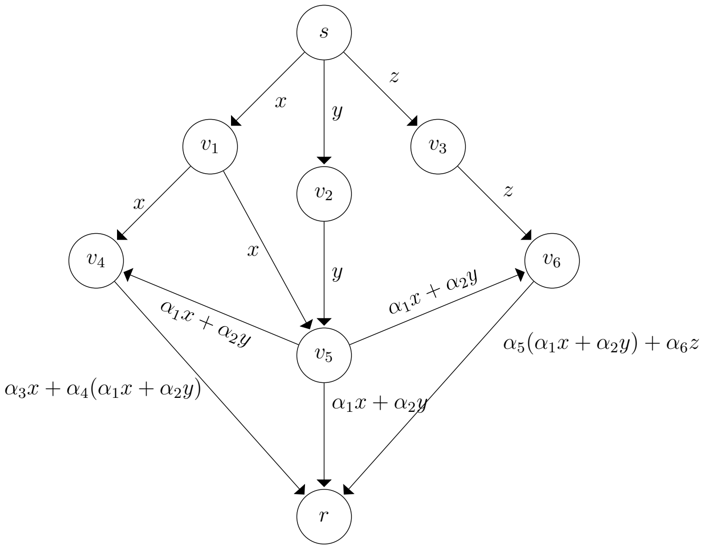

s
x
y
z
v_1
v_2
v_3
x
x
y
z
v_4
v_6
$\alpha_1 x + \alpha_2 y$
$\alpha_1 x + \alpha_2 y$
v_5
$\alpha_5(\alpha_1 x + \alpha_2 y) + \alpha_6 z$
$\alpha_3 x + \alpha_4(\alpha_1 x + \alpha_2 y)$
$\alpha_1 x + \alpha_2 y$
r

Figure 2.3: An example of random linear network coding without errors.

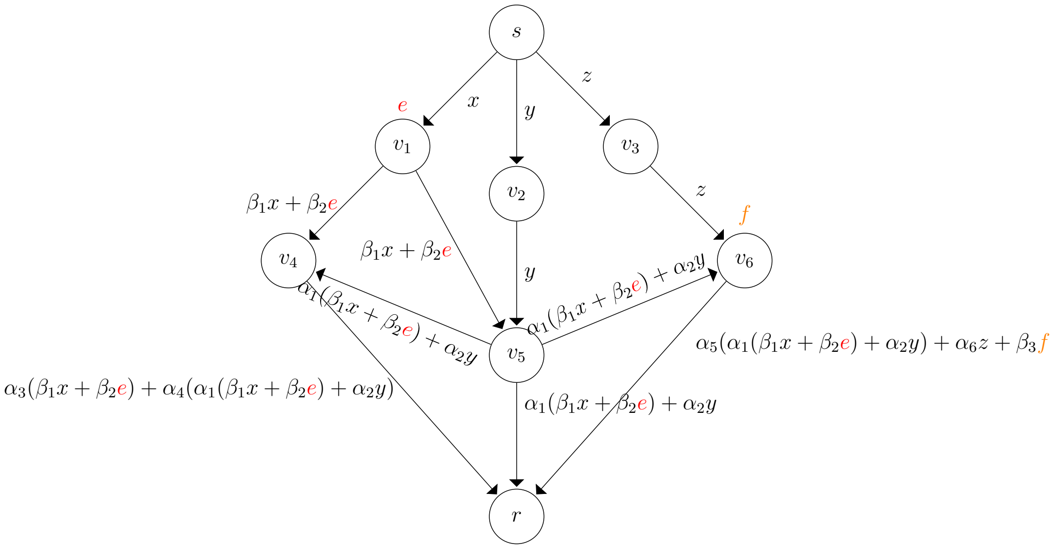

s
e
x
y
z
v_1
v_2
v_3
$\beta_1 x + \beta_2 e$
$\beta_1 x + \beta_2 e$
z
f
v_4
y
v_6
$\alpha_1(\beta_1 x + \beta_2 e) + \alpha_2 y$
$\alpha_1(\beta_1 x + \beta_2 e) + \alpha_2 y$
v_5
$\alpha_5(\alpha_1(\beta_1 x + \beta_2 e) + \alpha_2 y) + \alpha_6 z + \beta_3 f$
$\alpha_3(\beta_1 x + \beta_2 e) + \alpha_4(\alpha_1(\beta_1 x + \beta_2 e) + \alpha_2 y)$
$\alpha_1(\beta_1 x + \beta_2 e) + \alpha_2 y$
r

Figure 2.4: An example of random linear network coding with errors.

source packets x, y, z and the error packets e, f. The received packets are

$$\begin{cases} (\alpha_3 + \alpha_4\alpha_1\beta_1)x + \alpha_4\alpha_2 y + (\alpha_3 + \alpha_4\alpha_1)\beta_2 e \\ \alpha_1\beta_1 x + \alpha_2 y + \alpha_1\beta_2 e \\ \alpha_5\alpha_1\beta_1 x + \alpha_5\alpha_2 y + \alpha_6 z + \alpha_5\alpha_1\beta_2 e + \beta_3 f \end{cases},$$

(for clarity the errors are shown in different colours). We see that all three received packets contain errors. However the errors lie within the two-dimensional subspace spanned by the error vectors e and f.

In random linear network coding, although a small number of errors are able to corrupt all packets, since the errors on packets are contained within a small subspace it is possible to construct efficient methods for error correction.

In 2008 Kötter and Kschischang [26] studied random linear network coding, assuming adversarial errors (so the worst case was studied). They showed that it is optimal to encode information as a choice of subspace, transmitting packets that are vectors forming a generating set for that space. Indeed the subspace a set of vectors span is invariant under taking linear combinations. Then, since the errors are contained in a small subspace, the received subspace will be a small 'distance' from the input subspace. To decode one simply finds the 'closest' subspace from the set of possible inputs. The authors deduced coding bounds analogous to the sphere-packing, sphere-covering and Singleton bounds for classic codes and presented a Reed-Solomon-like code construction. The optimality of subspace coding puts into question the practical approach of [11] using coding headers. Although coding with headers can be viewed as subspace coding, the possible subspaces have a restricted form, resulting in a smaller, suboptimal coding space.

Following the work of [26], Montanari and Urbanke [32] and then Silva, Kschischang and Kötter [39] took a different approach, assuming random errors (as opposed to adversarial) and considered a probabilistic error model.

This is the approach we focus on for much of this work and we will discuss the probabilistic error model in detail in Section 2.4.

The subspace coding approach of [26] quickly became an active topic of research in network coding; for example an early survey of subspace coding is given in 2009 by Khaleghi, Silva and Kschischang [25]. This research area has since expanded and remains very active. For further surveys on network coding we refer the reader to the 2011 survey by Sanna and Izquierdo [35] and the 2013 survey by Bassoli *et al.* [3].

2.3 Partial decoding in random linear network coding

This section reviews the literature on partial decoding in random linear network coding, motivating the problems considered in Chapter 4. We conclude the section with an overview of the results of Chapter 4.

In random linear network coding, the packets received at the sink node (coded packets) are random linear combinations of source packets over a finite field. If k source packets are considered, decoding at a receiving node usually starts after k linearly independent coded packets have been collected. The probability of recovering all of the k source packets when at least k coded packets have been received has been derived by Trullols-Cruces, Barcelo-Ordinas and Fiore [40]. However, the requirement of decoders for the reception of a large number of coded packets could introduce undesirable delays at the receiving nodes.

In an effort to alleviate this problem, *rank-deficient* decoding was proposed by Yan, Xie, and Suter [42] for the recovery of a subset of source packets when fewer than k coded packets have been obtained. Whereas the literature on network coding defines *decoding success* as the recovery of 100% of the source packets with a certain probability, the authors of [42] presented numerical simulation results that measured the *fraction of decoding success*, that is, the recovery of a percentage of the source packets with a certain probability.

The fundamental problem that motivates the work of Chapter 4 is the characterisation of the probability of recovering some of the k source packets when n coded packets have been retrieved, where n can be smaller than, equal to or greater than k. This idea was considered independently by Gadouleau and Goupil [15] for random network communications over a matroid framework. The authors show that when transmitting only coded packets (so packets are randomised at the source), partial decoding is highly unlikely.

The problem has also been explored in the literature in the context of secure network coding, for example by Bhattad and Narayanan [6], and by Lima, Médard and Barros [29]. Strict information-theoretic security (in an appropriate model) is defined in [8] to be achieved if and only if the mutual information between the packets available to an eavesdropper and the source packets is zero. When network coding is used, the authors of [6] define a notion of *weak* security that can be achieved if the eavesdropper cannot obtain k linearly independent coded packets and, therefore, cannot recover any meaningful information about the k source packets. The authors obtained bounds on the probability of random linear network coding being weakly secure and showed that the adoption of large finite fields greatly improves security. A different setting but a similar problem was investigated in [29]. Intermediate relay nodes between transmitting nodes and receiving nodes were treated as potentially malicious, and criteria for characterising the algebraic security of random linear network coding were defined. The authors demonstrated that the probability of an intermediate node recovering a strictly positive number of source packets tends to zero as the field size and the number of source packets go to infinity.

In Chapter 4, we revisit this problem and make two key contributions:

- As explained in Section 1.3, we consider a random linear system of r linearly independent equations over a finite field and derive the probability of determining the values of at least x of the k unknowns for $x \leq r \leq k$.

- We draw parallels between systems of random linear equations and random linear network coding, and we obtain exact analytical expressions for the probability that a receiving node shall recover at least x of the k source packets if n random linear combinations of the k source packets are collected.

In addition to these contributions, the chapter investigates the impact of coding with headers when the headers of the first k packets transmitted are restricted to being the first k unit vectors and subsequent headers are random. This is known as *systematic* network coding and in practice means you are first transmitting source packets, followed by coded packets. We compare this to coding with headers, when all transmitted packets have random headers. This is *non-systematic* network coding, where in practice you are transmitting only coded packets. The asymptotic behavior of network coding over large finite fields is also studied.

2.4 Matrix channels for random linear network coding

In aim of this section is to show how the finite field matrix channels defined in Chapter 1 model various cases of random linear network coding.

In random linear network coding, the source injects packets into the network that can be thought of as vectors of length m with entries in a finite field $\mathbb{F}_q$ (where q is a fixed power of a prime). The packets flow through a network of unknown topology to a receiving node. Each intermediate node forwards packets that are random linear combinations of the packets it has received. A receiving sink node then attempts to reconstruct the message from these packets. In this context, k source packets can be represented as an $k \times m$ matrix over $\mathbb{F}_q$, $\boldsymbol{X}$, where the rows of $\boldsymbol{X}$ are the source packets. Similarly, if n packets are received by the sink, these can be represented as an $n \times m$ matrix over $\mathbb{F}_q$, $\boldsymbol{Y}$, where the rows of $\boldsymbol{Y}$ are the received packets. Thus a channel with input $\boldsymbol{X}$ and output $\boldsymbol{Y}$ is a model for network coding. The aim is to determine a channel law relating $\boldsymbol{Y}$ to $\boldsymbol{X}$ that accurately describes the effects of random linear network coding.

For simplicity (and in contrast to Section 2.3) we will assume the number of received packets is equal to the number of source packets (i.e. $k = n$). This is a widely considered practical case since a receiver simply waits until it has received $k = n$ packets and then begins decoding. However, the requirement of decoders for the reception of a large number of coded packets could introduce undesirable delays at the receiving nodes: see Section 2.3 and Chapter 4 for an alternative approach.

In Subsection 2.4.1 we show that the MMC is a model for error free network coding. Subsection 2.4.2 shows the AMC is a model for network coding when the coding coefficients are known to the receiver. Subsection 2.4.3 shows that the Gamma channel is a model for the general case of random linear network coding. Finally in Subsection 2.4.4 we discuss appropriate error rank distributions $\mathcal{R}$ for the AMC and Gamma channels, that lead to accurate models of error patterns in network coding.

2.4.1 The error free model

As in the work of Silva, Kschischang and Kötter [39], we use the Multiplicative Matrix Channel, defined in Definition 1.4.1, to model random linear network coding in the special case with no noise during transmission. In this error free case, since no erroneous packets are injected into the network, the sink node receives packets that are linear combinations of the original packets. Recall the MMC channel law is

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}$$

where $\boldsymbol{A} \in \mathrm{GL}(n, q)$ is chosen uniformly at random. This gives the rows of the output $\boldsymbol{Y}$ (the received packets) to be random linear combinations of the rows of the input $\boldsymbol{X}$ (the source packets). The matrix $\boldsymbol{A}$ describes the linear transformations the packets undergo during transmission.

When considering the model this channel describes, it is important to consider whether the distribution on the transfer matrix $\boldsymbol{A}$ is realistic. The distribution of $\boldsymbol{A}$ will depend on the underlying network topology and the choice of coding coefficients. The network topology will affect the transfer matrix distribution since the coding coefficients at a given node will be restricted by which packets that node has access to. For example a sparse network may be more likely to produce a sparse transfer matrix. However, large, deep and interconnected networks are likely to produce matrices A that are more uniform, therefore we assume the effect of the specific network topology is small enough that it can be ignored. Then, since the coding coefficients are chosen randomly, the transfer matrix $\boldsymbol{A}$ should be a uniform random matrix. The biggest assumption made is that $\boldsymbol{A}$ is invertible. We must consider whether it is reasonable to assume a randomly chosen $n \times n$ matrix over $\mathbb{F}_q$ is non-singular. This assumption becomes more and more realistic as the field size q grows, since you are less likely to get linear dependences. In network coding the field size is usually large to allow *feasibility* (that is the ability to deliver the input to all destinations when there are no errors). Therefore the MMC is a realistic model.

Both Nobrega, Silva, and Uchoa-Filho [33] and Siavoshani, Mohajer, Fragouli, and Diggavi [37] consider (different) generalisations of the MMC channel that do not necessarily have a square full rank transfer matrix. Such channels allow modelling of network coding when no erroneous packets are injected into the network, but there may be link erasures. In [33], the authors consider the following channel:

Definition 2.4.1. The *Generalised Multiplicative Matrix Channel* (GMMC) has input set $\mathcal{X} = \mathbb{F}_q^{k\times m}$ and output set $\mathcal{Y} = \mathbb{F}_q^{n\times m}$. The channel law is

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}$$

where $\boldsymbol{A} \in \mathbb{F}_q^{n\times k}$ is chosen according to a uniform given rank (UGR) distribution with some rank distribution $\mathcal{R}_{\boldsymbol{A}}$.

Note that the MMC is a special case of the GMMC, when $k = n$ and $\mathcal{R}_{\boldsymbol{A}}$ choses the matrix $\boldsymbol{A}$ to have full rank with probability 1.

In the GMMC the rank distribution on $\boldsymbol{A}$ is determined by the network topology, the random choices of coding coefficients and the link erasure probability. The authors prove that there exists a UGR optimal input distribution for the GMMC channel. Moreover, they show given any optimal input distribution, a UGR distribution with the same rank distribution will also be optimal. This implies that to find an optimal input distribution for the GMMC one must find an optimal distribution on the rank of the input $\boldsymbol{X}$ and then choose matrices uniformly once their rank is determined.

In this work we will focus on the MMC channel as its simplicity allows an elegant analysis of the channel capacity, which can be achieved via an intuitive coding scheme. We conclude Chapter 5 with a discussion comparing our results to those of [33]. However, we choose to consider a different generalisation of the MMC channel, the Gamma channel, which models a different class of cases in network coding; see Subsection 2.4.3.

2.4.2 The coherent model

Consider the case when the receiver knows the particular combinations source packets undergo during transmission; this is *coherent* network coding. For the error free MMC model discussed in Section 2.4.1 this is equivalent to the receiver knowing the transfer matrix $\boldsymbol{A}$. Since $\boldsymbol{A}$ is invertible the receiver can compute its inverse and multiply the received output by $\boldsymbol{A}^{-1}$ to obtain

$$\boldsymbol{A}^{-1}\boldsymbol{Y} = \boldsymbol{X}.$$

Since $\boldsymbol{A}^{-1}$ is known, up to relabeling this is equivalent to considering the channel with the simple law

$$\boldsymbol{Y} = \boldsymbol{X}. \tag{2.4.1}$$

However, unlike the error free MMC model, we now consider the introduction of errors. Following the approach of Montanari and Urbanke [32] and Silva, Kschischang and Kötter [39] we consider a probabilistic error model. We assume a small number of random erroneous packets may be injected at any point in the network. These packets are combined with legitimate packets and coding proceeds as before. The result is that each received packet will be a linear combination of the source packets and error packets. Supposing t' errors are introduced to the network, in the matrix model, for input $\boldsymbol{X} \in \mathbb{F}_q^{n\times m}$ the receiver obtains the output as

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{D}\boldsymbol{E} \tag{2.4.2}$$

where $\boldsymbol{A} \in \mathrm{GL}(n, q)$ is chosen uniformly and describes the linear combinations the source packets undergo, $\boldsymbol{E}$ is a $t' \times m$ matrix whose rows are the erroneous packets and $\boldsymbol{D}$ is an $n \times t'$ full rank random matrix that describes the linear transformations the error packets undergo during transmission. In general, the number of errors t' will not be fixed but will be determined by some distribution. The matrix $\boldsymbol{D}\boldsymbol{E}$ is then a $n \times m$ matrix whose rank gives the number of linearly independent errors. Since we are assuming errors are introduced randomly it is reasonable to assume the errors are distributed uniformly. As errors can be introduced at any point in the network they will encounter random linear transformations that are unknown to the receiver and different to those of the source packets. Due to the randomness of the errors and packet mixing, it is reasonable to assume that $\boldsymbol{D}\boldsymbol{E}$ has a UGR distribution.

Let $\boldsymbol{B} = \boldsymbol{A}^{-1}\boldsymbol{D}\boldsymbol{E}$, then since $\boldsymbol{A}^{-1}$ has full rank and is distributed uniformly, $\boldsymbol{B}$ is an $n \times m$ matrix with the same distribution as $\boldsymbol{D}\boldsymbol{E}$. It is then equivalent to write (2.4.2) as

$$\boldsymbol{Y} = \boldsymbol{A}(\boldsymbol{X} + \boldsymbol{B}) \tag{2.4.3}$$

where $\boldsymbol{B}$ is an $n \times m$ matrix chosen from a UGR distribution with rank distribution $\mathcal{R}$, where $\mathcal{R}$ gives the number of linearly independent errors. The matrix $\boldsymbol{B}$ is the error matrix: the rows of $\boldsymbol{B}$ represent linear combinations of the erroneous packets.

In coherent network coding, as in the error free case described above, we can assume the matrix $\boldsymbol{A}$ is known to the receiver, who can invert $\boldsymbol{A}$ and multiply the output by $\boldsymbol{A}^{-1}$ to obtain

$$\boldsymbol{A}^{-1}\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{B}.$$

Multiplying by $\boldsymbol{A}^{-1}$, inverting the known linear combinations, is essentially taking linear combinations of the received packets. This changes the linear combinations of errors on the received packets, but the distribution of these errors is the same. Then, up to relabeling, it is equivalent to consider the channel described by the law

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{B}, \tag{2.4.4}$$

where $\boldsymbol{B} \in \mathbb{F}_q^{n \times m}$ is UGR with some given rank distribution $\mathcal{R}$. Recall, this is precisely the channel law for the Additive Matrix Channel (AMC), defined in Definition 1.5.1. Therefore the AMC can be used to model coherent random linear network coding, when random errors are introduced into the network. In the model this gives the received packets (rows of $\boldsymbol{Y}$) to be source packets (rows of $\boldsymbol{X}$) subject to additive errors that are linear combinations of error packets (rows of $\boldsymbol{B}$).

Silva, Kschischang and Kötter [39] used the special case of the AMC channel with fixed error rank to model coherent network coding. That is they fixed the rank of $\boldsymbol{B}$ to equal exactly t. This channel allowed them to model the introduction of exactly t linearly independent errors, however there is no flexibility to allow the modelling of a different number of errors since the rank of $\boldsymbol{B}$ is the dimension of the error space. Indeed, they present a coding scheme that relies on an 'error-trapping' method to decode, which fails if less than t errors are 'trapped'. This scheme can be adapted to different error patterns but it is then possible for errors to go undetected, see [39, §VI. D] for further details.

A more natural restriction may be that $\mathrm{rk}(\boldsymbol{B}) \leq t$; allowing the modelling of situations when at most t errors are introduced, or when the errors are not necessarily linearly independent, or both. Different applications will lead to different distributions. Our generalised AMC channel allows for this, enabling the modelling of channels with different error patterns. The task is then to find appropriate distributions on the rank of $\boldsymbol{B}$ that lead to realistic models of error patterns. Such distributions are explored in Subsection 2.4.4.

2.4.3 The general model

The final channel we consider, the Gamma channel $\Gamma(\mathcal{R})$ (Definition 1.6.1), is used to model the general case of random linear network coding. In the general case we assume that coding is *non-coherent*, so the coding coefficients are unknown to the receiver and that random errors may be introduced into the network. Recall the channel law is

$$\boldsymbol{Y} = \boldsymbol{A}(\boldsymbol{X} + \boldsymbol{B}), \tag{2.4.5}$$

where $\boldsymbol{A} \in \mathrm{GL}(n, q)$ is chosen uniformly at random, $\boldsymbol{B} \in \mathbb{F}_q^{n\times m}$ is UGR with some given rank distribution $\mathcal{R}$ and $\boldsymbol{A}$ and $\boldsymbol{B}$ are chosen independently.

The rows of the output $\boldsymbol{Y}$ (received packets) are now random linear combinations of the rows of the input $\boldsymbol{X}$ (source packets) with the addition of rows of $\boldsymbol{B}$. So $\boldsymbol{A}$ describes the linear transformations the packets undergo during transmission and the rows of $\boldsymbol{B}$ represent random linear combinations of the erroneous packets.

Note that the MMC is exactly the Gamma channel with zero errors, that is the MMC is the special case of $\Gamma(\mathcal{R})$ when $\mathcal{R}$ is the distribution choosing rank 0 with probability 1. Similarly, the AMC($\mathcal{R}$) is exactly the $\Gamma(\mathcal{R})$ when the transfer matrix $\boldsymbol{A}$ is equal to the $n \times n$ identity matrix.

The validity of the Gamma channel as a model for network coding depends on the distributions of $\boldsymbol{A}$ and $\boldsymbol{B}$. As explained for the MMC, since we assume $\boldsymbol{A}$ is invertible, the Gamma channel becomes more and more realistic as the field size q grows. Then, as for the AMC, the error rank distribution $\mathcal{R}$ must be chosen to give a realistic model of error patterns. Silva, Kschischang and Kötter [39] used the special case of $\Gamma(\mathcal{R})$ with fixed error rank (i.e. $\mathcal{R}$ being the distribution choosing rank t with probability 1) to model the general case of random linear network coding. As discussed for the AMC channel this model is restrictive in the error patterns as it assumes that there are always exactly t linearly independent random errors introduced.

In the following subsection we discuss sensible distributions on the error rank to model naturally occurring error patterns.

2.4.4 Error rank distributions

As seen above, Silva, Kschischang and Kötter [39] used special cases of the AMC and Gamma channels with constant error rank to model random linear network coding when exactly t random linearly independent errors are introduced into the network.

Our models cover several other very natural situations. Both the AMC and Gamma channels allow for any distribution $\mathcal{R}$ on error ranks, which can be chosen to model a given error pattern. Some possible distributions are discussed below.

Firstly, suppose we assume the random errors are all linearly independent. Unlike the model in [39], there may not be exactly t errors and therefore we can model a distribution on the number of errors. Suppose in a given network every link has a small positive probability of producing an erroneous packet. In a large network it will be unlikely to have errors close to the sink, therefore one can assume sufficient packet mixing to ensure the error matrix is random and its distribution is only dependent on ranks. The error rank, or number of errors, could then be modeled by a binomial distribution. Or perhaps, if the probability of having more than t errors is extremely small, one could ignore this and use a truncated binomial distribution that defines the probability of more than t errors to be zero.

A different but related model may allow for the possibility that random errors may not be linearly independent. That is, with a certain probability, we may get lucky and errors may be introduced that are linear combinations of previous errors. If errors are introduced independently of each other this will mean that, given t' total errors, the probability that the error space will have dimension $r \leq t'$ will be equal to the probability that t' vectors span a subspace of dimension r. To model this situation, the total number of errors could be modeled as above with some binomial (perhaps truncated) distribution. Then, one could define $\mathcal{R}(r)$ to be the probability that t' vectors span a subspace of dimension r, averaged over the total number of errors t'.

Another possible model is an adversarial model, where it is assumed that exactly some fixed number of nodes are corrupt and produce random outputs. This could be extended to allow corrupt nodes to produce more than one output packet, meaning they could introduce multiple errors. This could become very dependent on the specific network topology and the location of corrupt nodes.

In practice, given a particular network, one may run tests on the network to see the actual error patterns produced and define an empirical distribution on ranks. It may also be sensible to consider some combination of the models described.

We have shown that there are several naturally occurring situations that could be represented under our model and therefore our generalisation of the matrix channels considered by [39] allow the modelling of a wider class of cases in random linear network coding.

Chapter 3

Preliminaries

3.1 Overview

This chapter presents preliminary results needed for the work that follows. We begin in Section 3.2 by considering finite-dimensional vector spaces. In Section 3.3 we discuss the theory of Möbius inversion. Finally, in Section 3.4 we review basic concepts from information theory.

3.2 Finite-dimensional vector spaces

In this section we consider vector spaces over finite fields. We begin by defining notation that will be adopted throughout this work.

Let q be a prime power and $\mathbb{F}_q$ be the finite field of q elements. For a positive integer m, we write $\mathbb{F}_q^m$ to denote the vector space composed of all m-tuples over $\mathbb{F}_q$. For a vector space V, we denote the dimension of V by $\dim(V)$. Let M be an $n \times m$ matrix over $\mathbb{F}_q$. We write $\mathrm{rk}(M)$ to denote the rank of M, suppose $\mathrm{rk}(M) = d$. In Chapter 1 we defined $\mathbb{F}_q^{n\times m}$ to be the set of all $n \times m$ matrices over $\mathbb{F}_q$, and $\mathbb{F}_q^{n\times m,d}$ to be the set of matrices in $\mathbb{F}_q^{n\times m}$ of rank d. The rowspace of M, denoted $\mathrm{Row}(M)$, is the d-dimensional subspace of $\mathbb{F}_q^m$ given by the span of the rows of M. Counting problems involving matrices and their rowspaces will be a recurring theme in this work. In this section we present the standard definitions and results on counting vector spaces that are needed for what follows.

The remainder of this section is organised as follows. In Subsection 3.2.1 we define and calculate a combinatorial constant which will be used in Subsection 3.2.2, where we discuss Gaussian binomial coefficients. In Subsection 3.2.3 we define quotient spaces and consider several subspace counting problems.

3.2.1 A combinatorial constant

In this subsection we define and calculate a combinatorial constant Q_0 which will be used in Subsection 3.2.2 to give a bound on the Gaussian binominal coefficient.

Consider the function $f(x) = \prod_{k=1}^{\infty}\left(1 - x^k\right)$. This function has lots of combinatorial applications, indeed it appears in Euler's pentagonal number theorem and its reciprocal is the generating function of integer partitions, see e.g. [2, Ch. 14]. For $x = \frac{1}{2}$, this function results in the combinatorial constant

$$Q_0 = f\left(\frac{1}{2}\right) = \prod_{k=1}^{\infty}\left(1 - 2^{-k}\right), \tag{3.2.1}$$

which is shown in [5] to give the probability that a random large square binary matrix is invertible. The authors of [5] compute the value of Q_0 by noting its equivalence to a rapidly converging series. The following lemma gives an alternative method for computing this value.

Lemma 3.2.1. *Let Q_0 be as defined in* (3.2.1)*. For $n \geq 1$, the value of Q_0 is bounded as follows*

$$\left(\prod_{k=1}^{n}\left(1 - 2^{-k}\right)\right)\exp\left(-2^{-n+1}\right) < Q_0 < \left(\prod_{k=1}^{n}\left(1 - 2^{-k}\right)\right)\exp\left(-2^{-n}\right), \tag{3.2.2}$$

in particular

$$Q_0 = 0.288788 \tag{3.2.3}$$

to six significant figures.

Proof. Note that for any integer $n \geq 1$,

$$Q_0 = \prod_{k=1}^{\infty} \left(1 - 2^{-k}\right) = \left(\prod_{k=1}^{n} \left(1 - 2^{-k}\right)\right) \exp\left(\ln\left(\prod_{j=n+1}^{\infty} \left(1 - 2^{-j}\right)\right)\right). \tag{3.2.4}$$

By the exponential expansion (e.g. [21, p. 104]) for any x such that $0 < x < 1$,

$$\begin{aligned} 1 - x < \exp(-x) &< 1 - x + \frac{x^2}{2!} \\ &< 1 - x + \frac{x}{2} \\ &= 1 - \frac{x}{2}, \end{aligned}$$

(where $m! = m \times (m-1) \times \cdots \times 1$ denotes the factorial of the integer m). Thus, given x with $0 < x < 1$,

$$\ln(1 - x) < -x < \ln\left(1 - \frac{x}{2}\right). \tag{3.2.5}$$

Therefore, for $n \geq 1$

$$\begin{aligned} \ln\left(\prod_{j=n+1}^{\infty} \left(1 - 2^{-j}\right)\right) &= \sum_{j=n+1}^{\infty} \ln\left(1 - 2^{-j}\right) \\ &< -\sum_{j=n+1}^{\infty} 2^{-j} \qquad (3.2.6) \\ &= -2^{-n} \qquad (3.2.7) \end{aligned}$$

where (3.2.6) follows from (3.2.5). Similarly

$$\begin{aligned}
\ln\left(\prod_{j=n+1}^{\infty}\left(1-2^{-j}\right)\right) &= \sum_{j=n+1}^{\infty}\ln\left(1-2^{-j}\right) \\
&= \sum_{j=n+1}^{\infty}\ln\left(1-\frac{2^{-j+1}}{2}\right) \\
&> -\sum_{j=n+1}^{\infty}2^{-j+1} \qquad (3.2.8) \\
&= -\sum_{i=n}^{\infty}2^{-i} \\
&= -2^{-n+1} \qquad (3.2.9)
\end{aligned}$$

where (3.2.8) follows from (3.2.5). Substituting (3.2.7) and (3.2.9) into (3.2.4) gives (3.2.2). For $n \geq 30$ the upper and lower bounds are equal to at least six significant figures, giving (3.2.3). □

3.2.2 Gaussian binomial coefficients

In this subsection we define Gaussian binomial coefficients, the q-analog of binomial coefficients and an essential tool in subspace counting arguments. We begin by recalling the definition of binomial coefficients.

Definition 3.2.1. Let m and d be non-negative integers. The *binomial coefficient*, denoted $\binom{m}{d}$, is defined to be the number of d-element subsets of an m-element set. It is given by (e.g. [9, §3.2])

$$\binom{m}{d} = \begin{cases} \dfrac{m!}{d!(m-d)!}, & \text{for } d \leq m \\ 0, & \text{for } d > m. \end{cases} \qquad (3.2.10)$$

Binomial coefficients are extremely important in combinatorics due to the wide range of counting problems they apply to. However their use in counting problems is restricted to those involving sets. For problems involving vector spaces over finite fields it is necessary to consider the q-analog of the binomial coefficient, the *Gaussian binomial coefficient*, defined below.

Definition 3.2.2. Let q be a prime power and let m and d be non-negative integers. The *Gaussian binomial coefficient*, denoted $\begin{bmatrix} m \\ d \end{bmatrix}_q$, is defined to be the number of d-dimensional subspaces of an m-dimensional space over $\mathbb{F}_q$. It is given by (e.g. [9, §9.2])

$$\begin{bmatrix} m \\ d \end{bmatrix}_q = \begin{cases} \displaystyle\prod_{i=0}^{d-1} \frac{(q^m - q^i)}{(q^d - q^i)}, & \text{for } d \leq m \\ 0, & \text{for } d > m. \end{cases} \tag{3.2.11}$$

Remark. The reason Gaussian binomial coefficients are considered the q-analog of binomial coefficients is that for fixed integers m and d (e.g. [9, §9.2]),

$$\lim_{q \to 1} \begin{bmatrix} m \\ d \end{bmatrix}_q = \binom{m}{d}.$$

Gaussian binomial coefficients apply to many counting problems involving vector spaces over finite fields, as such they will appear repeatedly throughout this work. Therefore, to simplify notation, when the underlying field size q is clear from the context, we omit the subscript q and write $\begin{bmatrix} m \\ d \end{bmatrix} = \begin{bmatrix} m \\ d \end{bmatrix}_q$ to denote the Gaussian binomial coefficient.

The remainder of this subsection establishes bounds on Gaussian binomial coefficients and discusses asymptotic behaviour. We begin by stating a known bound on the Gaussian binomial coefficient, which appears within the proof of [26, Lemma 4].

Lemma 3.2.2. *Let q be a prime power. For integers m, d with $d \leq m$ the following bound on the Gaussian binomial coefficient holds,*

$$q^{(m-d)d} < \begin{bmatrix} m \\ d \end{bmatrix} < q^{(m-d)d} \prod_{i=1}^{\infty} \frac{1}{1 - q^{-i}}. \tag{3.2.12}$$

Remark. The upper bound in Lemma 3.2.2 also appears in [16, Corollary 1].

Lemma 3.2.2 shows that the value of $\begin{bmatrix} m \\ d \end{bmatrix}$ is 'close to' $q^{(m-d)d}$ for large q, indeed the product $\prod_{i=1}^{\infty} \frac{1}{1-q^{-i}}$ approaches 1 quickly as q grows. The following lemma gives a constant factor bound on the Gaussian binomial coefficient, showing the behaviour for all values of q. (Note that a similar result is shown in [26, Lemma 4].)

Lemma 3.2.3. *Let q be a prime power. For integers m, d with $d \leq m$ the following bound on the Gaussian binomial coefficient holds.*

$$q^{(m-d)d} < \begin{bmatrix} m \\ d \end{bmatrix} < 3.5q^{(m-d)d}. \tag{3.2.13}$$

Proof.

$$\begin{aligned}\prod_{i=1}^{\infty} \frac{1}{1-q^{-i}} &\leq \prod_{i=1}^{\infty} \frac{1}{1-2^{-i}} \\ &= \frac{1}{Q_0} < 3.5, \end{aligned} \tag{3.2.14}$$

where (3.2.14) follows from Lemma 3.2.1. Substituting (3.2.14) into (3.2.12) gives the result. □

Lemma 3.2.3 shows that $\left[\begin{smallmatrix} m \\ d \end{smallmatrix}\right]$ is within a (small) constant factor of $q^{(m-d)d}$ for all values of q. Lemmas 3.2.2 and 3.2.3 show that for large q, the Gaussian binomial coefficient can be approximated as

$$\begin{bmatrix} m \\ d \end{bmatrix} \approx q^{(m-d)d}.$$

Indeed, we obtain the following lemma.

Lemma 3.2.4. *Let m and d be integers with $0 \leq d \leq m$. Then*

$$\lim_{q\to\infty} \begin{bmatrix} m \\ d \end{bmatrix} = q^{(m-d)d}.$$

Proof. Taking the limit as $q \to \infty$ in (3.2.12) gives the result. □

3.2.3 Quotient spaces

In this subsection we discuss quotient spaces and consider several subspace counting problems. An overview of quotient spaces is given in [20, §21 - 22]. We begin by recalling the definition of a quotient space.

Definition 3.2.3. Let V be a vector space and let U be a subspace of V. Define an equivalence relation $\sim$ on V such that for $v_1, v_2 \in V$, $v_1 \sim v_2$ if $v_1 - v_2 \in U$. The equivalence class of v_1 is

$$[v_1] = \{v_1 + u : u \in U\}.$$

The *quotient space* V/U is defined to be the set of all equivalence classes over V by $\sim$. Scalar multiplication and addition are defined as follows.

$$\alpha[v_1] = [\alpha v_1],$$

where α is any element of the base field, and

$$[v_1] + [v_2] = [v_1 + v_2].$$

Its not hard to show that this definition is well defined and that all elements of U map to zero in the quotient space. The following lemma gives the dimension of V/U.

Lemma 3.2.5. *Let V be a vector space of dimension d_V, and let U be a d_U-dimensional subspace of V. The quotient space V/U has dimension $d_V - d_U$.*

Proof. See [20, §22]. □

Definition 3.2.4. Let V be a vector space and let U be a subspace of V. The *quotient map* is defined to be the map that takes vectors in V to their image in the quotient space V/U,

$$\begin{array}{l} \pi : V \to V/U \\ \pi(v) = [v] \end{array}.$$

For a subspace W of V we shall denote the image of W under the quotient map as $[W]$.

The following lemma states a standard result from the theory of finite dimensional vector spaces.

Lemma 3.2.6. *Let V be a vector space and let U be a fixed subspace of V. For any subspace W of V*

$$\dim([W]) = \dim(W) - \dim(W \cap U),$$

where $[W]$ denotes the image of W in the quotient space V/U.

Proof. Note that

$$[W] = [W + U] = (W + U)/U, \tag{3.2.15}$$

since in the quotient U maps to zero. We have

$$\dim(W + U) = \dim(W) + \dim(U) - \dim(W \cap U), \tag{3.2.16}$$

and also by Lemma 3.2.5

$$\dim((W + U)/U) = \dim(W + U) - \dim(U). \tag{3.2.17}$$

Substituting (3.2.15) and (3.2.16) into (3.2.17) gives the result. □

The following lemma gives the number of subspaces U of V, when, given some subspace V_1 of V, the intersection of U and V_1 is fixed and the image of U in the quotient space V/V_1 is fixed.

Lemma 3.2.7. *Let V be a d_V-dimensional vector space, and let V_1, V_2 be subspaces of V, of dimensions d_{V_1} and d_{V_2} respectively, such that $V_2 \subseteq V_1$. The number of d_U-dimensional subspaces $U \subseteq V$ such that $U \cap V_1 = V_2$ and* *the image of U in the quotient space V/V_1 is the fixed $d_U - d_{V_2}$ dimensional space U', is given by*

$$q^{(d_U-d_{V_2})(d_{V_1}-d_{V_2})}.$$

Proof. Fix a basis for V_2, say $\{b_{1,1},\dots,b_{1,d_{V_2}}\}$. Let $\pi : V \to V/V_1$ be the map which takes vectors in V to their image in V/V_1. For $d_{U'} = d_U - d_{V_2}$, let $\{y_1,\dots y_{d_{U'}}\}$ be a basis for U', and let $\{b_{2,1},\dots,b_{2,d_{U'}}\}$ be some vectors in V such that $\pi(b_{2,i}) = y_i$, for $i = 1,\dots,d_{U'}$.

The set $\{b_{1,1},\dots,b_{1,d_{V_2}},b_{2,1},\dots,b_{2,d_{U'}}\}$ is a linearly independent set that forms the basis of a space U of the required form. Moreover, every space U of the required form has a basis that can be constructed in this way. The general construction takes the basis for V_2 and extends this to a basis for a space U by adding any set of $d_{U'}$ vectors in V whose image under π is $\{y_1,\dots,y_{d_{U'}}\}$.

Given the set $\{b_{2,1},\dots,b_{2,d_{U'}}\}$, all other sets of $d_{U'}$ vectors in V whose image under π is $\{y_1,\dots y_{d_{U'}}\}$ can be written in the form $\{v_1+b_{2,1},\dots,v_{d_{U'}}+b_{2,d_{U'}}\}$ for some $v_1,\dots,v_{d_{U'}} \in V_1$. Therefore, any space U of the required form has a basis that can be written as $B = \{b_{1,1},\dots,b_{1,d_{V_2}},v_1+b_{2,1},\dots,v_{d_{U'}}+b_{2,d_{U'}}\}$ for some $v_1,\dots,v_{d_{U'}} \in V_1$.

Given some set $v'_1,\dots,v'_{d_{U'}} \in V_1$, let $B' = \{b_{1,1},\dots,b_{1,d_{V_2}},v'_1+b_{2,1},\dots,v'_{d_{U'}}+b_{2,d_{U'}}\}$. Now $\mathrm{Span}(B) = \mathrm{Span}(B')$ if and only if $v_i - v'_i \in V_2$ for $i = 1,\dots,d_{U'}$. That is, the sets B and B' give rise to the same space if and only if $[v_i] = [v'_i]$ for $i = 1,\dots,d_{U'}$, where $[v]$ denotes the image of a vector v in the quotient space V_1/V_2.

Therefore there is a bijection between spaces U of the required form and ordered sets $\{[v_1],\dots,[v_{d_{U'}}]\}$ of elements in the quotient space V_1/V_2.

For $i = 1,\dots d_{U'}$, there are $q^{d_{V_1}-d_{V_2}}$ choices for $[v_i] \in V_1/V_2$, thus there are

$$q^{d_{U'}(d_{V_1}-d_{V_2})} = q^{(d_U-d_{V_2})(d_{V_1}-d_{V_2})}. \tag{3.2.18}$$

choices for the ordered set $\{[v_1], \ldots, [v_{d_{U'}}]\}$. The result follows. □

Given a vector space V and a subspace $V_1 \subseteq V$, Lemma 3.2.7 can be used to count subspaces U of V when either $U \cap V_1$ is fixed, or the image of U in V/V_1 is fixed, or when only the dimensions of these spaces are fixed. These results are given in the following three corollaries.

Corollary 3.2.8. *Let V be a d_V-dimensional vector space, and let V_1, V_2 be subspaces of V, of dimensions d_{V_1} and d_{V_2} respectively, such that $V_2 \subseteq V_1$. The number of d_U-dimensional subspaces $U \subseteq V$ such that $U \cap V_1 = V_2$, is given by*

$$q^{(d_U - d_{V_2})(d_{V_1} - d_{V_2})} \begin{bmatrix} d_V - d_{V_1} \\ d_U - d_{V_2} \end{bmatrix}.$$

Proof. Consider the quotient space V/V_1, this is a space of dimension $d_V - d_{V_1}$. Let U' be a $(d_U - d_{V_2})$-dimensional subspace of V/V_1. There are

$$\begin{bmatrix} d_V - d_{V_1} \\ d_U - d_{V_2} \end{bmatrix} \tag{3.2.19}$$

possible choices for U', fix one such space.

By Lemma 3.2.7 there are $q^{(d_U - d_{V_2})(d_{V_1} - d_{V_2})}$ possibilities for the space U whose image in the quotient V/V_1 is the fixed space U'. Multiplying by the number of possibilities for U', yields the result. □

Corollary 3.2.9. *Let V be a d_V-dimensional vector space, and let V_1 be a d_{V_1}-dimensional subspace of V. The number of d_U-dimensional subspaces $U \subseteq V$ such that the image of U in the quotient space V/V_1 is some fixed $d_{U'}$ dimensional space U', is given by*

$$q^{d_{U'}(d_{V_1} - (d_U - d_{U'}))} \begin{bmatrix} d_{V_1} \\ d_U - d_{U'} \end{bmatrix}. \tag{3.2.20}$$

Proof. Let V_2 be a $(d_U - d_{U'})$-dimensional subspace of V_1, there are

$$\begin{bmatrix} d_{V_1} \\ d_U - d_{U'} \end{bmatrix}$$

possible choices for V_2, fix one.

By Lemma 3.2.7 there are $q^{(d_U-(d_U-d_{U'}))(d_{V_1}-(d_U-d_{U'}))} = q^{d_{U'}(d_{V_1}-(d_U-d_{U'}))}$ possibilities for the space U whose intersection with V_1 is the fixed space V_2. Multiplying by the number of possibilities for V_2, yields the result. □

Corollary 3.2.10. *Let V be a d_V-dimensional vector space, and let V_1 be a d_{V_1}-dimensional subspace of V. The number of d_U-dimensional subspaces $U \subseteq V$ such that* $\dim(U \cap V_1) = d_{UV_1}$ *is equal to the number of d_U-dimensional subspaces U such that* $\dim([U]) = d_U - d_{UV_1}$*, where $[U]$ denotes the image of U in the quotient space V/V_1. The value is given by*

$$q^{(d_U-d_{UV_1})(d_{V_1}-d_{UV_1})} \begin{bmatrix} d_V - d_{V_1} \\ d_U - d_{UV_1} \end{bmatrix} \begin{bmatrix} d_{V_1} \\ d_{UV_1} \end{bmatrix}.$$

Proof. Note that $\dim(U \cap V_1) = d_{UV_1}$ if and only if $\dim([U]) = d_U - d_{UV_1}$ hence the equivalence in the statement of the lemma holds. Let V_2 be a (d_{UV_1})-dimensional subspace of V_1, there are

$$\begin{bmatrix} d_{V_1} \\ d_{UV_1} \end{bmatrix}$$

possible choices for V_2, fix one. Next let U' be a $(d_U - d_{UV_1})$-dimensional subspace of V/V_1. There are

$$\begin{bmatrix} d_V - d_{V_1} \\ d_U - d_{UV_1} \end{bmatrix} \tag{3.2.21}$$

possible choices for U', fix one.

By Lemma 3.2.7 there are $q^{(d_U-d_{UV_1})(d_{V_1}-d_{UV_1})}$ possibilities for the space U whose intersection with V_1 is the fixed space V_2, and image in the quotient V/V_1 is the fixed space U'. Multiplying by the number of possibilities for V_2 and U', yields the result. □

3.3 Möbius theory

In this section we discuss Möbius inversion, an important tool in combinatorics allowing the calculation of many complex counting problems. The fundamental paper by Rota [34], shows the importance of Möbius theory in combinatorics, exploring the connections in depth. Bender and Goldman [4] complement [34], giving an exposition of many applications of Möbius inversion in combinatorics. We recall several of these results below.

We begin by defining partially ordered sets (posets) and the Möbius function of a poset.

Definition 3.3.1. A *partially ordered set* or *poset* $(S, \leq)$ consists of a set S and a binary relation (ordering) $\leq$ on S that is

- reflexive: $a \leq a$ for all $a \in S$,
- transitive: if $a \leq b$ and $b \leq c$ then $a \leq c$,
- anti-symmetric: if $a \leq b$ and $b \leq a$ then $a = b$.

The poset $(S, \leq)$ is *locally finite* if for all $a, b \in S$ the closed interval

$$[a, b] = \{s \in S : a \leq s \leq b\}$$

is finite. For simplicity, we denote the poset $(S, \leq)$ by S and the ordering $\leq$ is understood by the context.

Definition 3.3.2. The *Möbius function* of a locally finite poset S is an integer valued function of two variables on S defined by $\mu(x, z) = 0$ if $x \not\leq z$ and when $x \leq z$,

$$\sum_{y : x \leq y \leq z} \mu(x, y) = \begin{cases} 1 & \text{if } x = z \\ 0 & \text{otherwise.} \end{cases}$$

Next we state the Möbius inversion formula, given in [4, Theorem 1].

Lemma 3.3.1. ***Möbius inversion.*** *Let S be a locally finite poset. Let $f(x)$ be a real valued function defined for all $x \in S$. Suppose there exists $s \in S$ such that $f(x) = 0$ for all $x < s$. Define*

$$g(x) = \sum_{y \leq x} f(y).$$

Then

$$f(x) = \sum_{y \leq x} \mu(y, x) g(y),$$

where μ is the Möbius function of S.

Remark. If S is a finite poset, as opposed to locally finite, given any real valued function $f(x)$ on S, one can define $f(x) = 0$ for all $x < \min\{s : s \in S\}$ (since no such x exist). Therefore, for finite posets it is possible to remove the constraint that there exists $s \in S$ such that $f(x) = 0$ for all $x < s$ in the statement of Lemma 3.3.1.

Let S be a finite set and let $P(S)$ denote the power set of S, that is the set of all subsets of S, ordered by inclusion. It is shown in [4, §3] that the Möbius function of $P(S)$ is given by

$$\mu(I, J) = (-1)^{|J \setminus I|} \tag{3.3.1}$$

for $I, J \in P(S)$ with $I \subseteq J$. Applying the Möbius inversion formula to $P(S)$ gives the following result, which is the basic principle of inclusion and exclusion.

Lemma 3.3.2. ***Principle of inclusion and exclusion.*** *Let $f(J)$ and $f'(J)$ be real valued functions defined for all $J \in P(S)$. If*

$$g(I) = \sum_{J \subseteq I} f(J)$$

then

$$f(I) = \sum_{J \subseteq I} (-1)^{|J \setminus I|} g(J). \tag{3.3.2}$$

Similarly, if

$$g'(I) = \sum_{J \supseteq I} f'(J)$$

then

$$f'(I) = \sum_{J \supseteq I} (-1)^{|I \setminus J|} g'(J). \tag{3.3.3}$$

Proof. Firstly, (3.3.2) follows by applying the Möbius inversion formula and substituting in (3.3.1). Then (3.3.3) follows by setting $f''(I) = f'(S \setminus I)$, $g''(I) = g'(S \setminus I)$ and applying (3.3.2) to the functions f'' and g''. □

Let $\mathrm{Po}(\mathbb{F}_q^m)$ denote the poset of all subspaces of $\mathbb{F}_q^m$ ordered by containment. It is shown in [4, §5] that the Möbius function of $\mathrm{Po}(\mathbb{F}_q^m)$ is given by

$$\mu(V, U) = (-1)^{\dim(U) - \dim(V)} q^{\binom{\dim(U) - \dim(V)}{2}}. \tag{3.3.4}$$

for $U, V \in \mathrm{Po}(\mathbb{F}_q^m)$ with $V \subseteq U$. Applying the Möbius inversion formula to $\mathrm{Po}(\mathbb{F}_q^m)$ gives the following result.

Lemma 3.3.3. *Let $f(U)$ be a real valued function defined for all subspaces U of $\mathbb{F}_q^m$. If*

$$g(U) = \sum_{V \subseteq U} f(V)$$

then

$$f(U) = \sum_{V \subseteq U} (-1)^{\dim(U) - \dim(V)} q^{\binom{\dim(U) - \dim(V)}{2}} g(V).$$

Proof. Applying the Möbius inversion formula and substituting in (3.3.4) gives the result. □

We now move on to consider direct products. Let P and Q be two posets. The direct product $P \times Q$ is the poset where $(p_1, q_1) \leq (p_2, q_2)$ if and only if $p_1 \leq p_2$ and $q_1 \leq q_2$, where $p_1, p_2 \in P$ and $q_1, q_2 \in Q$. In the following lemma we give the Möbius function of the direct product of two posets ([4, Theorem 3]).

Lemma 3.3.4. *If P has Möbius function μ_P and Q has Möbius function μ_Q, the Möbius function of $P \times Q$ is given by*

$$\mu((p_1, q_1), (p_2, q_2)) = \mu_P(p_1, p_2)\mu_Q(q_1, q_2). \tag{3.3.5}$$

Applying the Möbius inversion formula to the direct product $\mathrm{Po}(\mathbb{F}_q^m) \times \mathrm{Po}(\mathbb{F}_q^m)$ gives the following result.

Lemma 3.3.5. *Let $f((U, V))$ be a real valued function defined for all pairs $(U, V) \in Po(\mathbb{F}_q^m) \times Po(\mathbb{F}_q^m)$. If*

$$g((U, V)) = \sum_{(U', V') \leq (U, V)} f((U', V'))$$

then

$$f((U, V)) = \sum_{(U', V') \leq (U, V)} (-1)^{u-u'+v-v'} q^{\binom{u-u'}{2} + \binom{v-v'}{2}} g((U', V')),$$

where $\dim(U) = u, \dim(U') = u', \dim(V) = v$ *and* $\dim(V') = v'$.

Proof. By the Möbius inversion formula

$$f((U, V)) = \sum_{(U', V') \leq (U, V)} \mu((U', V'), (U, V)) g((U', V')). \tag{3.3.6}$$

Substituting (3.3.4) into (3.3.5) gives

$$\mu((U', V'), (U, V)) = (-1)^{u-u'+v-v'} q^{\binom{u-u'}{2} + \binom{v-v'}{2}}. \tag{3.3.7}$$

Substituting (3.3.7) into (3.3.6) gives the result. □

3.4 Information theory

Information theory is fundamental in this work as we consider channels to model network coding. The classic book by Thomas and Cover [13] gives a detailed exposition of information theory. In this section we extract the necessary concepts from [13], in particular we focus on entropy and mutual information [13, Ch. 2] in Subsection 3.4.1 and on channels [13, Ch. 8] in Subsection 3.4.2.

For the definitions and results we use it is standard to take logarithms to the base 2, giving the expressions in terms of binary bits. However, since we will look at channels whose inputs and outputs are matrices over a finite field of order q, it is more natural to take logarithms to the base q, giving expressions in q-ary bits. Throughout this section we will omit the base in order to state results in their generality, however for the remainder of this work one may assume logarithms are taken to the base q and 'bits' are q-ary bits, unless specifically stated otherwise.

3.4.1 Entropy and mutual information

In this subsection we recall several definitions and results about the entropy and mutual information of discrete random variables. For the proofs of these results see [13, Ch. 2].

We begin by defining notation for probabilities. For random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ let

- $\Pr(\boldsymbol{X} = X)$ be the probability that $\boldsymbol{X}$ takes the value X, that is $\boldsymbol{X} = X$,
- $\Pr(\boldsymbol{X} = X, \boldsymbol{Y} = Y)$ be the joint probability that $\boldsymbol{X} = X$ and $\boldsymbol{Y} = Y$,
- $\Pr(\boldsymbol{Y} = Y \mid \boldsymbol{X} = X)$ be the probability that $\boldsymbol{Y} = Y$ given that $\boldsymbol{X} = X$.

Next we define entropy, which is a measure of the uncertainty of a discrete random variable.

Definition 3.4.1. Let $\boldsymbol{X}$ be a discrete random variable with alphabet $\mathcal{X}$. The *entropy* of $\boldsymbol{X}$ is defined to be

$$H(\boldsymbol{X}) = -\sum_{X \in \mathcal{X}} \Pr(\boldsymbol{X} = X) \log \Pr(\boldsymbol{X} = X).$$

The following two lemmas state some properties about entropy.

Lemma 3.4.1. *Let $\boldsymbol{X}$ be a discrete random variable. Then*

$$H(\boldsymbol{X}) \geq 0.$$

Lemma 3.4.2. *Let $\boldsymbol{X}$ be a discrete random variable with alphabet $\mathcal{X}$ of cardinality $|\mathcal{X}|$. Then*

$$H(\boldsymbol{X}) \leq \log|\mathcal{X}|,$$

with equality if and only if $\boldsymbol{X}$ has a uniform distribution over $\mathcal{X}$.

Lemma 3.4.2 shows that the entropy of $\boldsymbol{X}$ is maximal when $\boldsymbol{X}$ has a uniform distribution. Intuitively this makes sense since the uncertainty is greatest when the distribution is uniform.

The following definition defines the joint entropy of discrete random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$, that is a measure of the uncertainty of $\boldsymbol{X}$ and $\boldsymbol{Y}$.

Definition 3.4.2. Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be discrete random variables with alphabets $\mathcal{X}$ and $\mathcal{Y}$ respectively. The *joint entropy* of $\boldsymbol{X}$ and $\boldsymbol{Y}$ is defined to be

$$H(\boldsymbol{X}, \boldsymbol{Y}) = -\sum_{X \in \mathcal{X}} \sum_{Y \in \mathcal{Y}} \Pr(\boldsymbol{X} = X, \boldsymbol{Y} = Y) \log \Pr(\boldsymbol{X} = X, \boldsymbol{Y} = Y).$$

Next we define the conditional entropy $H(\boldsymbol{Y} \,|\, \boldsymbol{X})$, a measure of the uncertainty of $\boldsymbol{Y}$ given knowledge of $\boldsymbol{X}$.

Definition 3.4.3. Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be discrete random variables with alphabets $\mathcal{X}$ and $\mathcal{Y}$ respectively. The *conditional entropy* of $\boldsymbol{Y}$ by $\boldsymbol{X}$ is defined to be

$$\begin{aligned} H(\boldsymbol{Y} \mid \boldsymbol{X}) &= \sum_{X \in \mathcal{X}} \Pr(\boldsymbol{X} = X) H(\boldsymbol{Y} \mid \boldsymbol{X} = X) \\ &= -\sum_{X \in \mathcal{X}} \Pr(\boldsymbol{X} = X) \sum_{Y \in \mathcal{Y}} \Pr(\boldsymbol{Y} = Y \mid \boldsymbol{X} = X) \log \Pr(\boldsymbol{Y} = Y \mid \boldsymbol{X} = X). \end{aligned}$$

These natural definitions of joint and conditional entropy lead to the *chain rule for entropy*, stated below.

Lemma 3.4.3. ***Chain rule for entropy.*** *Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be discrete random variables. Then*

$$H(\boldsymbol{X}, \boldsymbol{Y}) = H(\boldsymbol{X}) + H(\boldsymbol{Y} \mid \boldsymbol{X}).$$

The following lemma shows that conditioning reduces entropy. This is true intuitively, since knowledge of one random variable can only decrease the uncertainty of another.

Lemma 3.4.4. *Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be discrete random variables. Then*

$$H(\boldsymbol{Y} \mid \boldsymbol{X}) \leq H(\boldsymbol{Y}).$$

Now we have covered the basic properties of entropy, we move on to define the mutual information of two discrete random variables.

Definition 3.4.4. Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be discrete random variables with alphabets $\mathcal{X}$ and $\mathcal{Y}$ respectively. The *mutual information* of $\boldsymbol{X}$ and $\boldsymbol{Y}$ is defined to be

$$I(\boldsymbol{X}; \boldsymbol{Y}) = \sum_{X \in \mathcal{X}} \sum_{Y \in \mathcal{Y}} \Pr(\boldsymbol{X} = X, \boldsymbol{Y} = Y) \log \frac{\Pr(\boldsymbol{X} = X, \boldsymbol{Y} = Y)}{\Pr(\boldsymbol{X} = X) \Pr(\boldsymbol{Y} = Y)}.$$

Mutual information is closely related to entropy, as demonstrated in the following lemma.

Lemma 3.4.5. *Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be discrete random variables. Then the following equalities hold:*

$$\begin{aligned} I(\boldsymbol{X};\boldsymbol{Y}) &= I(\boldsymbol{Y};\boldsymbol{X}) && (3.4.1)\\ &= H(\boldsymbol{X}) - H(\boldsymbol{X}\,|\,\boldsymbol{Y}) && (3.4.2)\\ &= H(\boldsymbol{Y}) - H(\boldsymbol{Y}\,|\,\boldsymbol{X}) && (3.4.3)\\ &= H(\boldsymbol{X}) + H(\boldsymbol{Y}) - H(\boldsymbol{X},\boldsymbol{Y}). && (3.4.4) \end{aligned}$$

Moreover, the mutual information of $\boldsymbol{X}$ with itself, is just the entropy of $\boldsymbol{X}$, that is

$$I(\boldsymbol{X};\boldsymbol{X}) = H(\boldsymbol{X}). \qquad (3.4.5)$$

Expressing the mutual information $I(\boldsymbol{X};\boldsymbol{Y})$ as in (3.4.2) shows that $I(\boldsymbol{X};\boldsymbol{Y})$ measures the loss in uncertainty of $\boldsymbol{X}$ from the knowledge of $\boldsymbol{Y}$, or equivalently the amount of information $\boldsymbol{Y}$ gives about $\boldsymbol{X}$. Furthermore, the symmetry of (3.4.3) shows that $\boldsymbol{X}$ gives as much information about $\boldsymbol{Y}$ as $\boldsymbol{Y}$ gives about $\boldsymbol{X}$.

We conclude this subsection by stating some properties of the mutual information when considered as a function over possible probability distributions.

Definition 3.4.5. A function $f(x)$ is *convex* on an interval (a,b) if for every $x_1, x_2 \in (a,b)$ and $0 \leq \lambda \leq 1$

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2).$$

A function $g(x)$ is *concave* on an interval (a,b) if for every $x_1, x_2 \in (a,b)$ and $0 \leq \lambda \leq 1$

$$g(\lambda x_1 + (1-\lambda)x_2) \geq \lambda g(x_1) + (1-\lambda)g(x_2).$$

Let $\mathcal{P}_{\boldsymbol{X}}$ denote the probability distribution of a random variable $\boldsymbol{X}$, so that $\mathcal{P}_{\boldsymbol{X}}(X) = \Pr(\boldsymbol{X} = X)$. Similarly, let $\mathcal{P}_{\boldsymbol{Y},\boldsymbol{X}}$ and $\mathcal{P}_{\boldsymbol{Y}|\boldsymbol{X}}$ denote the joint distribution of $\boldsymbol{X}$ and $\boldsymbol{Y}$ and the conditional distribution of $\boldsymbol{Y}$ given $\boldsymbol{X}$, respectively.

Lemma 3.4.6. *Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be random variables distributed according to some distribution $\mathcal{P}_{\boldsymbol{Y},\boldsymbol{X}} = \mathcal{P}_{\boldsymbol{X}}\mathcal{P}_{\boldsymbol{Y}|\boldsymbol{X}}$. The mutual information $I(\boldsymbol{X};\boldsymbol{Y})$ is a concave function of $\mathcal{P}_{\boldsymbol{X}}$ for a fixed distribution $\mathcal{P}_{\boldsymbol{Y}|\boldsymbol{X}}$ and a convex function of $\mathcal{P}_{\boldsymbol{Y}|\boldsymbol{X}}$ for a fixed distribution $\mathcal{P}_{\boldsymbol{X}}$.*

3.4.2 Channels

This subsection recalls some definitions and results on channels and channel capacity. For further information and for proofs of these results see [13, Ch. 8].

We begin with the definition of a discrete memoryless channel.

Definition 3.4.6. A *discrete channel*, denoted $(\mathcal{X}, \mathcal{Y}, \mathcal{P}_{\boldsymbol{Y}|\boldsymbol{X}})$, is a system consisting of an input alphabet $\mathcal{X}$, an output alphabet $\mathcal{Y}$ and a conditional probability distribution $\mathcal{P}_{\boldsymbol{Y}|\boldsymbol{X}}$, giving the probability of the channel output being $Y \in \mathcal{Y}$ for a given input $X \in \mathcal{X}$. The channel is *memoryless* if the probability distribution of the output $\mathcal{P}_{\boldsymbol{Y}}$ depends only on the input at that time and is conditionally independent of all previous channel inputs and outputs.

Next we define the channel capacity of a discrete memoryless channel in terms of the mutual information of the channel input and output.

Definition 3.4.7. The *channel capacity* of a discrete memoryless channel $(\mathcal{X}, \mathcal{Y}, \mathcal{P}_{\boldsymbol{Y}|\boldsymbol{X}})$ is defined to be

$$C = \max_{\mathcal{P}_{\boldsymbol{X}}} I(\boldsymbol{X};\boldsymbol{Y}),$$

where $\mathcal{P}_{\boldsymbol{X}}$ denotes the probability distribution of the channel input.

This is a natural definition since the mutual information measures the amount of information $\boldsymbol{Y}$ gives about $\boldsymbol{X}$ and hence the channel capacity gives the maximum amount of information we can send through the channel.

Given a coding scheme for encoding and decoding information sent over a channel, the *rate* of that scheme is (imprecisely) the amount of information conveyed per channel use. It is then equivalent to define the channel capacity to be the supremum of all achievable rates (Shannon's second theorem).

The capacity is a key parameter of a channel, in particular, as Shannon's second theorem shows, the capacity gives the maximum rate at which we can transmit information over the channel.

Chapter 4

Probability of Partially Solving Random Linear Systems in Network Coding

4.1 Overview

This chapter considers the problem of partial decoding in random linear network coding, as introduced in Section 2.3. In the literature there exist analytical expressions for the probability of a receiver decoding a transmitted source message that has been encoded using random linear network coding. In this chapter, we look into the probability that the receiver will decode at least a fraction of the source message. We present an exact solution to this problem for both non-systematic and systematic network coding, by rephrasing the problem as the enumeration problem from linear algebra presented in Section 1.3. Based on our exact solution, we investigate the potential of these two implementations of network coding for information-theoretic secure communication and progressive recovery of data.

This work was done in collaboration with Ioannis Chatzigeorgiou and resulted in the preprint [12], which forms the basis for this chapter. My main contribution to the paper was the development of the formula in Section 4.2.

The rest of the chapter has been organised as follows. Section 4.2 introduces the notation and derives the probability that the rows of a random matrix over a finite field define a subspace containing a particular number of unit vectors. After the analogy between unit vectors and source packets is explained, Section 4.3 focuses on both non-systematic and systematic network coding, and obtains the probability of recovering a fraction of a network-coded message. Results and trends are discussed in Section 4.4, while conclusions are summarised in Section 4.5.

4.2 The elementary unit vectors in a rowspace

In this section we find the probability that the rowspace of a random matrix contains some fixed number of unit vectors. We begin by defining notation that will be adopted throughout this chapter. We then present the problem solution and conclude the section by noting the equivalence to the probability of partially solving a random linear system, which is related to the probability of a receiver decoding a fraction of a source message that has been encoded using random linear network coding.

Let $\boldsymbol{M}$ be a random $n \times k$ matrix over $\mathbb{F}_q$. We define the random variable R to give the rank of the matrix $\boldsymbol{M}$. For $i = 1, \ldots, k$, we write e_i to denote the i-th unit vector of length k, that is the vector with 1 in the i-th position and zeros elsewhere. Let X be the set of indices that correspond to the unit vectors that are contained in the rowspace of $\boldsymbol{M}$, so that $X = \{i : e_i \in \mathrm{Row}(\boldsymbol{M})\}$. We write $|X|$ to denote the random variable giving the cardinality of the set X.

Given the matrix $\boldsymbol{M}$ has rank r, let $P(|X| = x \,|\, R = r)$ denote the probability of $\mathrm{Row}(\boldsymbol{M})$ containing *exactly* $x \leq r$ unit vectors, and $P(|X| \geq x \,|\, R = r)$

denote the probability of Row($\boldsymbol{M}$) containing *at least* $x \leq r$ unit vectors. Expressions for the probabilities $P(|X| = x \,|\, R = r)$ and $P(|X| \geq x \,|\, R = r)$ are derived in the remainder of this section.

Theorem 4.2.1. *Given a random $n \times k$ matrix $\boldsymbol{M}$ of rank r, the probability that the rowspace of $\boldsymbol{M}$ contains exactly $x \leq r$ unit vectors is given by*

$$P(|X| = x \,|\, R = r) = \frac{\binom{k}{x}}{\begin{bmatrix} k \\ r \end{bmatrix}} \sum_{j=0}^{k-x} (-1)^j \binom{k-x}{j} \begin{bmatrix} k-x-j \\ r-x-j \end{bmatrix}. \tag{4.2.1}$$

Proof. Let X be the set of indices that correspond to the unit vectors that are contained in the rowspace of $\boldsymbol{M}$. For $S \subseteq \{1, \ldots k\}$, let $g(S)$ be the probability that $\{e_i : i \in S\} \subseteq \text{Row}(\boldsymbol{M})$, that is the probability that $S \subseteq X$. This is just the probability that Row($\boldsymbol{M}$) contains a fixed $|S|$-dimensional subspace, namely the space $V = \text{Span}\{e_i : i \in S\}$. By considering the quotient space $\mathbb{F}_q^k/V$, we see there is a direct correspondence between r-dimensional subspaces of $\mathbb{F}_q^k$ containing V, and $(r-|S|)$-dimensional subspaces of a $(k-|S|)$-dimensional space. Hence, there are

$$\begin{bmatrix} k-|S| \\ r-|S| \end{bmatrix} \tag{4.2.2}$$

r-dimensional subspaces of $\mathbb{F}_q^k$ containing V. Therefore, the probability that Row($\boldsymbol{M}$) contains the space V is equal to (4.2.2) divided by the number of r-dimensional subspaces of $\mathbb{F}_q^k$, that is

$$g(S) = \frac{\begin{bmatrix} k-|S| \\ r-|S| \end{bmatrix}}{\begin{bmatrix} k \\ r \end{bmatrix}}. \tag{4.2.3}$$

Now let $f(S)$ be the probability that $S = X$, that is the probability that

$\{e_i : i \in S\} \subseteq \mathrm{Row}(\boldsymbol{M})$ and $\{e_i : i \notin S\} \nsubseteq \mathrm{Row}(\boldsymbol{M})$. It follows that

$$g(S) = \sum_{S \subseteq J \subseteq \{1,\ldots,k\}} f(J).$$

Then, by the Principle of Inclusion and Exclusion (Lemma 3.3.2), we can write

$$f(S) = \sum_{S \subseteq J \subseteq \{1,\ldots,k\}} (-1)^{|J \setminus S|} g(J). \tag{4.2.4}$$

Substituting (4.2.3) into (4.2.4) gives

$$\begin{aligned} f(S) &= \sum_{S \subseteq J \subseteq \{1,\ldots,k\}} (-1)^{|J \setminus S|} \frac{\begin{bmatrix} k-|J| \\ r-|J| \end{bmatrix}}{\begin{bmatrix} k \\ r \end{bmatrix}} \\ &= \frac{1}{\begin{bmatrix} k \\ r \end{bmatrix}} \sum_{J' \subseteq \{1,\ldots,k\} \setminus S} (-1)^{|J'|} \begin{bmatrix} k-|S|-|J'| \\ r-|S|-|J'| \end{bmatrix} \qquad (4.2.5) \\ &= \frac{1}{\begin{bmatrix} k \\ r \end{bmatrix}} \sum_{j=0}^{k-|S|} (-1)^j \binom{k-|S|}{j} \begin{bmatrix} k-|S|-j \\ r-|S|-j \end{bmatrix} \qquad (4.2.6) \end{aligned}$$

where (4.2.5) follows by setting $J' = J \setminus S$, and (4.2.6) follows since there are $\binom{k-|S|}{j}$ sets J' of size j.

Then since $f(S)$ is the probability that $X = S$,

$$\begin{aligned} P(|X| = x \mid R = r) &= \sum_{S : |S| = x} f(S) \\ &= \binom{k}{x} f(S'), \qquad (4.2.7) \end{aligned}$$

where S' is any subset of $\{1,\ldots,k\}$ of size x, and (4.2.7) holds since there are $\binom{k}{x}$ sets $S \subseteq \{1,\ldots,k\}$ of size x. Substituting (4.2.6) in (4.2.7) gives the result. □

Remark. Theorem 4.2.1 appears as a special case of [15, Proposition 6], where the authors consider the probability of partially decoding in non-systematic random linear network coding given a matroid framework. Our work was developed independently and relies on an alternative proof strategy.

Corollary 4.2.2. *Given a random $n \times k$ matrix $\boldsymbol{M}$ of rank r, the probability that the rowspace of $\boldsymbol{M}$ contains at least $x \leq r$ unit vectors is given by*

$$P(|X| \geq x \,|\, R = r) = \frac{1}{\begin{bmatrix} k \\ r \end{bmatrix}} \sum_{i=x}^{r} \binom{k}{i} \sum_{j=0}^{k-i} (-1)^j \binom{k-i}{j} \begin{bmatrix} k-i-j \\ r-i-j \end{bmatrix}. \qquad (4.2.8)$$

Proof. By definition

$$P(|X| \geq x \,|\, R = r) = \sum_{i=x}^{r} P(|X| = i \,|\, R = r). \qquad (4.2.9)$$

Substituting (4.2.1) into (4.2.9) gives the result. □

Note that, although $\boldsymbol{M}$ is defined to be an $n \times k$ matrix, the probabilities $P(|X| = x \,|\, R = r)$ and $P(|X| \geq x \,|\, R = r)$ are independent of n since they depend only on the row space of $\boldsymbol{M}$. Thus the expressions given in (4.2.1) and (4.2.8) hold for any value of $n \geq r$.

Remark. Our results also address the equivalent problem of finding the probability of partially solving random underdetermined linear systems over finite fields. Suppose a random linear system of r equations in $k \geq r$ unknowns, $v_1, \ldots, v_k$, is expressed in the matrix form

$$\boldsymbol{M}\underline{v} = \underline{u}, \qquad (4.2.10)$$

where $\boldsymbol{M}$ is a random full rank $r \times k$ matrix, $\underline{v} = (v_1, \ldots, v_k)$, and $\underline{u}$ is a constant vector of length r. Given that (4.2.10) is consistent, it is possible to determine the i-th unknown v_i if and only if e_i is contained in the rowspace of $\boldsymbol{M}$. Hence, $P(|X| = x \,|\, R = r)$ gives the probability of determining exactly x of the unknowns and $P(|X| \geq x \,|\, R = r)$ gives the probability of determining at least x of the unknowns.

4.3 Partial and full recovery of network-coded messages

Whereas Section 4.2 provided a generic formulation of the considered problem, this section puts the derived expressions into the context of network coding in order to characterise the probability of recovering a fraction of a network-coded message. For convenience and clarity, the notation that was introduced in Section 4.2 is also used in this section.

Let us consider a receiving network node, which collects n packets and attempts to reconstruct a message that consists of k source packets. In the case of *non-systematic* communication, transmitted packets are generated from the k source packets using random linear network coding over $\mathbb{F}_q$ [23]. A coding vector of length k, which contains the weighting coefficients used in the generation of a coded packet, is transmitted along with each coded packet[1]. Note that a coding vector that is equal to unit vector e_i, as defined in Section 4.2, represents the i-th source packet. At the receiving node, the coding vectors of the n successfully retrieved coded packets form the rows of a matrix $\boldsymbol{M} \in \mathbb{F}_q^{n \times k}$. The k source packets can be recovered from the n coded packets if and only if k of the n coding vectors are linearly independent, implying that $\mathrm{rk}(\boldsymbol{M}) = k$ for $n \geq k$. The probability that the $n \times k$ random matrix $\boldsymbol{M}$ has rank k and, thus, the receiving node can reconstruct the entire message is given by the

[1] In practice, a coding vector can be represented by the seed value of a predetermined pseudo-random function [10] or shortened using simple compression methods [22] before it is appended to the header of the associated coded packet.

well-known expression[2] [41]

$$P_{\mathrm{ns}}(R = k \mid N = n) = \prod_{i=0}^{k-1} \left(1 - q^{-n+i}\right). \tag{4.3.1}$$

The probability in (4.3.1) is conditioned on N, the random variable giving the number of received packets. If the distribution of the successfully delivered packets over a packet erasure channel is known, the marginal probability $P_{\mathrm{ns}}(R = k)$ can be obtained from $P_{\mathrm{ns}}(R = k \mid N = n)$ by averaging over all possible values of n.

In the case of *systematic* network coding, a sequence of n_{T} transmitted packets consists of the k source packets followed by $n_{\mathrm{T}} - k$ coded packets, which have been generated as in the non-systematic case. If the receiving node collects $n \geq k$ packets, let h of them be source packets and the remaining $n-h$ be coded packets. Elementary row and column operations can split the $n \times k$ matrix $\boldsymbol{M}$ into four sub-matrices, with the top left being the $h \times h$ identity matrix, the bottom right being an $(n-h) \times (k-h)$ random matrix and the remaining entries being zero. The probability that $\mathrm{rk}(\boldsymbol{M}) = k$ is [30, 36]

$$P_{\mathrm{s}}(R = k \mid N = n) = \frac{1}{\binom{n_{\mathrm{T}}}{n}} \sum_{h=h_{\min}}^{k} \binom{k}{h} \binom{n_{\mathrm{T}} - k}{n - h} \prod_{i=0}^{k-h-1} \left(1 - q^{-n+h+i}\right) \tag{4.3.2}$$

where $h_{\min} = \max(0,\, n - n_{\mathrm{T}} + k)$ [24]. Again, N is the random variable for the number of received packets. Note that for $k - h - 1 < 0$, the product in the right-hand side of (4.3.2) becomes an empty product and is equal to 1.

[2] Expression (4.3.1) assumes that any of the q^k coding vectors can be transmitted but practical implementations of network coding consider only the $q^k - 1$ non-zero vectors. For coding vectors in $\mathbb{F}_q^k \backslash \mathbf{0}$, the probability of recovering the whole message has been derived in [40, 44] but (4.3.1) converges to it even for small values of k and q. For example, for $n = k = 10$ and $q = 2$, the probability given by (4.3.1) is within 2.8×10^{-3} of the exact probability and is significantly closer for larger values of k or q. For this reason and owing to its simplicity, (4.3.1) is often used regardless of whether coding vectors take values in $\mathbb{F}_q^k$ or $\mathbb{F}_q^k \backslash \mathbf{0}$.

Both (4.3.1) and (4.3.2) provide the probability that the receiving node will recover the *entire* message from the n delivered packets in the non-systematic and systematic implementations of network coding, respectively. The following propositions consider both network coding schemes and derive the probability that the receiving node will reconstruct a *fraction* of the source message.

Proposition 4.3.1. *If a receiving node collects n random linear combinations of k source packets, the probability that at least $x \leq k$ source packets will be recovered is*

$$P_{\text{ns}}(|X| \geq x \,|\, N = n) = \frac{1}{q^{nk}} \sum_{r=x}^{\min(n,k)} \left(\sum_{i=x}^{r} \binom{k}{i} \sum_{j=0}^{k-i} (-1)^j \binom{k-i}{j} \begin{bmatrix} k-i-j \\ r-i-j \end{bmatrix} \right) \prod_{\ell=0}^{r-1} (q^n - q^\ell). \tag{4.3.3}$$

Proof. In accordance with Section 4.2, let X be the set of indices that correspond to the unit vectors in $\text{Row}(\boldsymbol{M})$, or equivalently the recoverable source packets, and let $|X|$ be the cardinality of that set. Provided that the matrix $\boldsymbol{M}$ has rank r, the probability that X contains the indices of at least x of the k source packets, denoted by $P(|X| \geq x \,|\, R = r)$, is given by (4.2.8). Let $P(R = r)$ denote the probability that the $n \times k$ matrix $\boldsymbol{M}$ has rank r. This is equivalent to the probability that the receiving node has collected r linearly independent random linear combinations of the k source packets, given that n random linear combinations have been received in total. The average probability that at least x of the k source packets will be recovered can be obtained as follows:

$$P_{\text{ns}}(|X| \geq x \,|\, N = n) = \sum_{r=x}^{\min(n,k)} P(R = r)\; P(|X| \geq x \,|\, R = r). \tag{4.3.4}$$

The probability $P(R = r)$ is equal to [41, p. 338]

$$P\left(R = r\right) = \frac{1}{q^{nk}} \begin{bmatrix} k \\ r \end{bmatrix} \sum_{\ell=0}^{r} (-1)^{r-\ell} \begin{bmatrix} r \\ \ell \end{bmatrix} q^{n\ell + \binom{r-\ell}{2}} \tag{4.3.5}$$

but can be further reduced to [14, Equation 13]

$$P(R = r) = \frac{1}{q^{nk}} \begin{bmatrix} n \\ r \end{bmatrix} \prod_{\ell=0}^{r-1} \left(q^k - q^\ell\right). \tag{4.3.6}$$

Substituting (4.2.8) and (4.3.6) into (4.3.4) and taking into account that

$$\frac{\left[\begin{smallmatrix} n \\ r \end{smallmatrix}\right]}{\left[\begin{smallmatrix} k \\ r \end{smallmatrix}\right]} \prod_{\ell=0}^{r-1} (q^k - q^\ell) = \prod_{\ell=0}^{r-1} (q^n - q^\ell) \tag{4.3.7}$$

which follows from the definition of the Gaussian binomial coefficient in (3.2.11), we obtain (4.3.3). □

Remark. The factor $1/q^{nk}$ in (4.3.6) implies that the realisations of all $n \times k$ random matrices over $\mathbb{F}_q$ are uniformly distributed. If random matrices having the same rank follow a rank distribution $P(R = r)$ other than that in (4.3.6), the general expression (4.3.4) can be used instead.

Proposition 4.3.2. *If k source packets and $n_\mathrm{T}-k$ random linear combinations of those k source packets are transmitted over single-hop links, the probability that a receiving node will recover at least $x \le k$ source packets from $n \le n_\mathrm{T}$ received packets is*

$$\begin{aligned} P_\mathrm{s}(|X| \ge x \,|\, N = n) = \frac{1}{\binom{n_\mathrm{T}}{n}} \sum_{r=x}^{\min(n,k)} \sum_{h=h_\mathrm{min}}^{r} \Bigg(\binom{k}{h} \binom{n_\mathrm{T}-k}{n-h} q^{-(n-h)(k-h)} \prod_{\ell=0}^{r-h-1} (q^{n-h} - q^\ell) \\ \cdot \sum_{i=x_\mathrm{min}}^{r-h} \binom{k-h}{i} \sum_{j=0}^{k-h-i} (-1)^j \binom{k-h-i}{j} \begin{bmatrix} k-h-i-j \\ r-h-i-j \end{bmatrix} \Bigg) \end{aligned} \tag{4.3.8}$$

where $h_\mathrm{min} = \max(0,\, n - n_\mathrm{T} + k)$ and $x_\mathrm{min} = \max(0,\, x-h)$.

Proof. Let us assume that some or none of the k transmitted source packets have been received and let $X' \subseteq X$ be the set of indices of the remaining source packets that can be recovered from the received coded packets. If n' coded packets have been received and k' source packets remain to be recovered, the respective coding vectors will form an $n' \times k'$ random matrix $\boldsymbol{M}'$. The probability that $r' \leq \min(k', n')$ coding vectors are linearly independent and at least $x' \leq r'$ source packets can be recovered is given by

$$P(|X'| \geq x',\, R' = r' \,|\, N' = n') = P(R' = r')\; P(|X'| \geq x' \,|\, R' = r') \quad (4.3.9)$$

where the two terms of the product can be obtained from (4.3.6) and (4.2.8), respectively. Here the random variables N' and R' denote the number of coded packets received and the rank of the matrix $\boldsymbol{M}'$ respectively. If n of the n_{T} transmitted packets are received, the probability that h of them are source packets and the remaining $n-h$ are coded packets is

$$P(N' = n - h \,|\, N = n) = \frac{\dbinom{k}{h}\dbinom{n_{\mathrm{T}} - k}{n - h}}{\dbinom{n_{\mathrm{T}}}{n}}. \quad (4.3.10)$$

The coding vectors of the n received packets compose a matrix of rank r, based on which x or more source packets can be recovered when h of the n received packets are source packets. Parameters x', r', k' and n', which are concerned with the received *coded* packets only, can be written as $x-h$, $r-h$, $k-h$ and $n-h$, respectively. Therefore, the probability of recovering at least x source packets for all valid values of r and h is

$$\begin{aligned} P_{\mathrm{s}}(|X| \geq x \,|\, N = n) = \sum_{r=x}^{\min(n,k)} \sum_{h=h_{\min}}^{r} & P(N' = n - h \,|\, N = n) \\ & \cdot P\big(|X'| \geq \max(0,\, x{-}h),\, R' = r - h \,|\, N' = n{-}h\big) \quad (4.3.11) \end{aligned}$$

which can be expanded into (4.3.8). Note that $\max(0,\, x-h)$ ensures that the first input to the second term of the product in (4.3.11) is a non-negative integer when $h > x$. □

Remark. Proposition 4.3.2 assumes that the receiving node attempts to recover a part of or the entire source message after the k source packets have been transmitted, i.e., $n_\mathrm{T} > k$. If the objective of the receiving node is to identify recoverable source packets as soon as the transmission is initiated, i.e., $n_\mathrm{T} \leq k$, at least x source packets will certainly be recovered if $n \geq x$ source packets are received. Thus, for $n_\mathrm{T} > k$ we can use (4.3.8) but for $n_\mathrm{T} \leq k$ we can write

$$P_\mathrm{s}(|X| \geq x \,|\, N = n) = \begin{cases} 1, & \text{if } \; n_\mathrm{T} \leq k \; \text{ and } \; x \leq n \\ 0, & \text{if } \; n_\mathrm{T} \leq k \; \text{ and } \; x > n. \end{cases} \tag{4.3.12}$$

As is well-established [31], random linear network coding over large finite fields can deliver optimal rate, that is, the k source packets can be recovered as soon as $n = k$ packets are received.

Recall from Section 3.2.2, for finite fields of large size q, the Gaussian binomial coefficient can be approximated as

$$\begin{bmatrix} m \\ d \end{bmatrix} \approx q^{(m-d)d}. \tag{4.3.13}$$

Using this approximation in (4.2.1) gives

$$P(|X| = x \,|\, R = r) \approx \binom{k}{x} q^{-(k-r)x} \sum_{j=0}^{k-x} (-1)^j \binom{k-x}{j} q^{-(k-r)j}. \tag{4.3.14}$$

The summation in (4.3.14) is the binomial expansion of $\left[1 - q^{-(k-r)}\right]^{k-x}$. Therefore, (4.3.14) can be reduced to

$$P(|X| = x \,|\, R = r) \approx \binom{k}{x} \left[q^{-(k-r)}\right]^x \left[1 - q^{-(k-r)}\right]^{k-x} \tag{4.3.15}$$

which corresponds to the probability mass function of a binomial distribution with $q^{-(k-r)}$ being the probability of obtaining a single source packet. Observe that for $q \to \infty$, the reception of fewer than k linearly independent coded packets does not help in the recovery of any of the source packets, i.e., $P(|X| = 0 \,|\, R = r) = 1$ if $r < k$. Substituting (4.3.15) into (4.2.8), then into both (4.3.3) and (4.3.8) and taking limits as $q \to \infty$, leads to expressions (4.3.16) and (4.3.17) below, respectively. These give the asymptotic behaviour of non-systematic and systematic network coding.

Firstly for non-systematic network coding we obtain the limit of (4.3.3) as $q \to \infty$ to be

$$\lim_{q\to\infty} P_{\mathrm{ns}}(|X| \geq x \,|\, N = n) = \begin{cases} 1, & \text{if } \ k \leq n \\ 0, & \text{otherwise.} \end{cases} \tag{4.3.16}$$

In other words, if non-systematic network coding is used, the entire source message can indeed be decoded if k or more coded packets are received. However, if fewer than k coded packets are collected, not even a single source packet of the original message can be recovered, provided that $q \to \infty$. This result is in accordance with [29, Theorem 1] but without the requirement for $k \to \infty$. It also confirms the conclusions of [6] about the potential of non-systematic network coding for weak information-theoretic security when large finite fields are used. On the other hand, if systematic network coding is employed, the limit of (4.3.8) and (4.3.12) as $q \to \infty$ is

$$\lim_{q\to\infty} P_{\mathrm{s}}(|X| \geq x \,|\, N = n) = \begin{cases} 1, & \text{if } \ k \leq n \leq n_{\mathrm{T}} \ \text{ or } \ n \leq n_{\mathrm{T}} \leq k \\ \dfrac{1}{\dbinom{n_{\mathrm{T}}}{n}} \displaystyle\sum_{h=h_x}^{n} \binom{k}{h}\binom{n_{\mathrm{T}}-k}{n-h}, & \text{if } \ n < k < n_{\mathrm{T}} \\ 0, & \text{otherwise} \end{cases} \tag{4.3.17}$$

where $h_x = \max\,(x,\, n - n_{\mathrm{T}} + k)$ and $x \leq \min(n, k)$. In this case, as the top and middle branches of (4.3.17) dictate, x or more source packets can be recovered when fewer than k packets are received if and only if they are among the k transmitted source packets. As in the previous case, the complete set of k source packets will be recovered if k or more transmitted packets are collected. The asymptotic behavior of non-systematic and systematic network coding will be further discussed in the following section.

4.4 Results and discussion

Section 4.3 derived closed-form expressions for the probability that a receiving node will recover some or all of the source packets that compose a message and, consequently, will be able to reconstruct part of or the entire source message. To demonstrate the accuracy of the derived expressions, Monte Carlo simulations for random linear combinations of $k = 20$ source packets using arithmetic operations in $\mathbb{F}_2$ were carried out. The probability that a receiving node using Gaussian elimination will recover at least x source packets, given that n packets are received, was measured for $x \in \{2, 4, \ldots, 20\}$ and $x \leq n \leq 25$. Fig. 4.1(a) and Fig. 4.1(b) compare probability measurements obtained through simulations to probability calculations obtained from (4.3.3) and (4.3.8) for non-systematic and systematic network coding, respectively. In the case of systematic network coding, the length of the transmitted sequence of source and coded packets also needs to be considered and is taken to be $n_{\mathrm{T}} = 30$. The plots clearly show that the simulations and the exact expressions (4.3.3) and (4.3.8) are in agreement. They also confirm the intuitive expectation that the number of received packets has a more pronounced effect on the probability of partly recovering the source message ($x < 20$) when systematic network coding is employed as opposed to non-systematic network coding.

Fig. 4.2 and Fig. 4.3 consider the simple case of network-coded transmission over a broadcast erasure channel. If the transmission of n_{T} packets is modeled as a sequence of n_{T} Bernoulli trials whereby ε signifies the probability that a transmitted packet will be erased, the probability that a receiving node shall recover *at least* x of the k source packets can be expressed as

$$P(|X| \geq x) = \sum_{n=x}^{n_{\mathrm{T}}} \binom{n_{\mathrm{T}}}{n} (1-\varepsilon)^{n} \, \varepsilon^{n_{\mathrm{T}}-n} \, P(|X| \geq x \,|\, N=n). \tag{4.4.1}$$

The conditional probability $P(|X| \geq x \,|\, N = n)$ is equal to (4.3.3) for non-systematic network coding or (4.3.8) and (4.3.12), depending on the value of n_{T}, for systematic network coding.

Fig. 4.2 focuses on non-systematic network coding and uses a colour map to depict $P(|X| \geq x)$ in terms of parameters n_{T} and x, which have been normalised by the considered value of k. Results have been obtained for $k \in \{20, 30\}$, $q \in \{2, 8\}$ and $\varepsilon \in \{0.05, 0.2\}$. For $q = 2$, we observe in Figs. 4.2(a)–4.2(d) that fractions of the transmitted message can be recovered with different probabilities when fewer than k coded packets have been transmitted ($n_{\mathrm{T}}/k < 1$). However, non-systematic network coding starts to exhibit the asymptotic behavior reported in Section 4.3 for values of q as low as 8. As shown in Fig. 4.2(e) and Fig. 4.2(f), only a very small number of source packets can be recovered with low probability for a small number of transmitted coded packets. The long single-coloured vertical stripes for $n_{\mathrm{T}}/k \geq 1$ imply that a receiving node will recover the entire message ($x/k = 1$) with a certain probability but will be unable to recover large fractions of the message with a higher probability. Bear in mind that if $P(|X| \geq x_1) = P(|X| \geq x_2)$ for $x_1 < x_2$, the probability of recovering *exactly* $x \in \{x_1, x_1 + 1, \ldots, x_2 - 1\}$ source packets is zero. A comparison between the plots on the left-hand and right-hand sides of Fig. 4.2 confirms that an increase in the erasure probability significantly affects the gradient of $P(|X| \geq x)$, as is evident by the sharp transition from low to high values of $P(|X| \geq x)$ for an increasing value of n_{T}/k on the left-hand side plots and the smoother transition on the right-hand side plots. The effect that the number of source packets, which constitute the message to be delivered, has on $P(|X| \geq x)$ can be noticed in Fig. 4.2(b) and Fig. 4.2(d). For small values of n_{T}/k, dividing the message into $k = 20$ source packets permits the receiving node to recover a higher fraction of the message (x/k) with a non-zero probability than dividing the same message into $k = 30$ source packets. On the other hand, if n_{T}/k takes values in the high regime of $(0, 1.4]$, segmentation into $k = 30$ packets offers a small improvement in the probability of recovering a fraction of the message over segmentation into $k = 20$ packets.

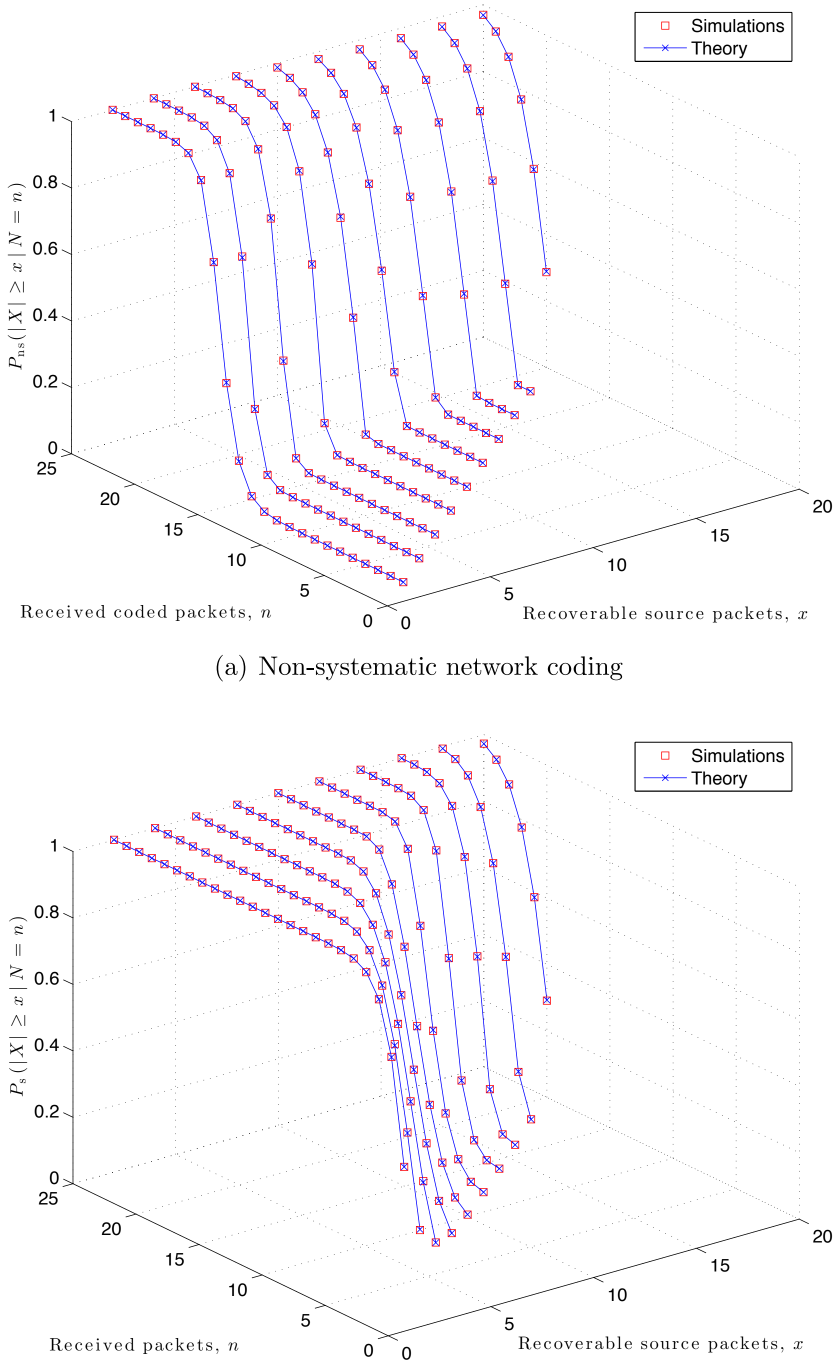
The same settings as in Fig. 4.2 are used in Fig. 4.3 but systematic network coding is considered. Besides the reduced decoding complexity reported in [31], we observe that the systematic implementation of network coding enables the receiving node to reveal an increasingly larger portion of the message as the number of transmitted packets grows. For $\varepsilon = 0.2$, the plots on the right-hand side of Fig. 4.3 show that a small finite field (e.g., $q = 2$) and even a small number of source packets can be used to progressively recover the message. The adoption of high-order finite fields (e.g., $q \geq 8$) impairs the progressive recovery of the message for $n_{\mathrm{T}}/k \geq 1$ but enables the recovery of the entire message for a smaller number of transmitted packets.

Both Fig. 4.2 and Fig. 4.3 illustrate that the choice of the network coding scheme and the corresponding design parameters are strongly dependent on the system requirements. If secrecy is of importance and legitimate nodes experience better average channel conditions than eavesdroppers, non-systematic network coding over large finite fields can be used to segment each secret message into a large number of source packets. The number of transmitted coded packets can then be tuned to the average channel conditions to achieve a balance between the probability that legitimate nodes can reconstruct the entire message and the probability that eavesdroppers cannot decode even a small portion of the message. On the other hand, if the objective of the communication system is to maximise the number of nodes that will recover at least a large part of a transmitted message, systematic network coding over small finite fields can be used to segment data into a relatively small number of packets. If the receiving nodes do not suffer from limited computational capabilities or energy constraints, the size of the finite field used in systematic network coding can be increased in order to improve the probability of recovering the entire transmitted message.

4.5 Conclusions

Previous work had shown that the probability of decoding a fraction of a network-coded source message can be made infinitesimal by coding over large finite fields. However, exact probability expressions for fields of any size and network parameters of any value were not available in the literature. This chapter derived the probability of recovering a fraction of the source message, conditioned on the reception of a specific number of linearly independent coded packets. The obtained conditional probability laid the foundation for the derivation of the probability of decoding a fraction of the source message upon reception of an arbitrary number of packets, when non-systematic or systematic random linear coding is used. Results confirmed that non-systematic network coding offers weak information-theoretic security because it does not allow for the decoding of sizeable portions of the source message with high probability, unless the number of collected coded packets is sufficiently large, even when operations are over finite fields of small size. By contrast, systematic network coding allows for the progressive recovery of the source message as the number of received packets increases, especially when the size of the finite field is small.

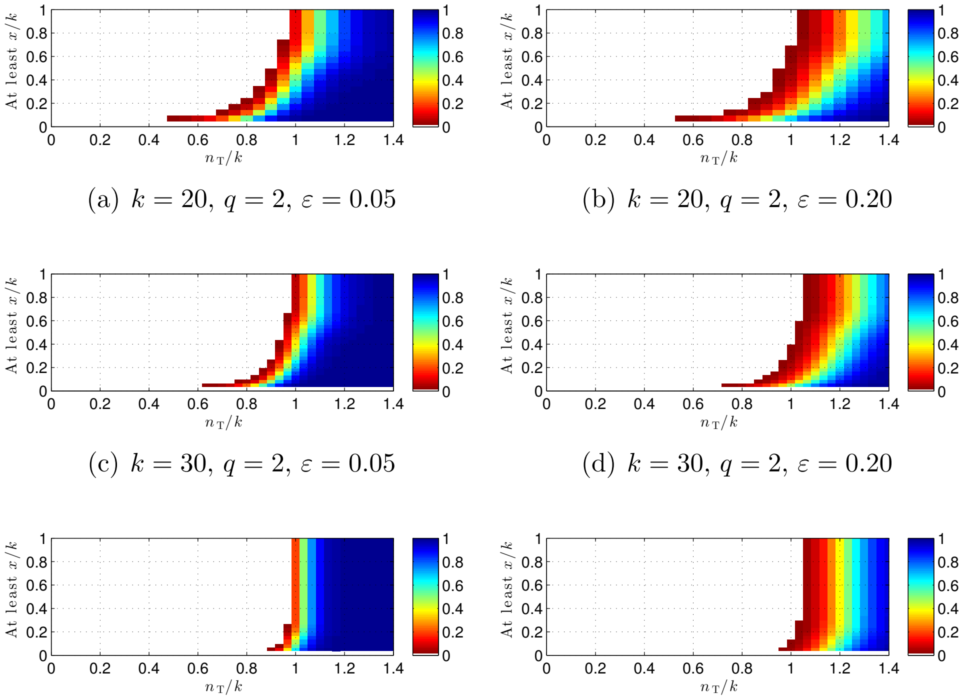
The derived exact expressions can prove useful in network design and system-level optimisation. For example, the objective of a system could be the minimisation of the probability of malicious nodes recovering one or more source packets, without resorting to unnecessarily large field sizes that would impose a prohibitive computational cost on legitimate receiving nodes. On the other hand, the objective of a broadcast system could be the provision of a guaranteed service quality, which could be translated as the recovery of either an entire set of data with a certain probability or a fraction of the data with a higher probability.



(a) Non-systematic network coding

(b) Systematic network coding ($n_{\mathrm{T}} = 30$)

Figure 4.1: Comparison of results obtained through simulations and from theoretical expressions (4.3.3), (4.3.8) for (a) non-systematic network coding and (b) systematic network coding, respectively. Source messages consist of $k = 20$ packets and arithmetic operations are in $\mathbb{F}_2$. The probability of recovering at least x source packets has been plotted for $x = 2, 4, 6, \ldots, 20$.



(a) $k = 20$, $q = 2$, $\varepsilon = 0.05$

(b) $k = 20$, $q = 2$, $\varepsilon = 0.20$

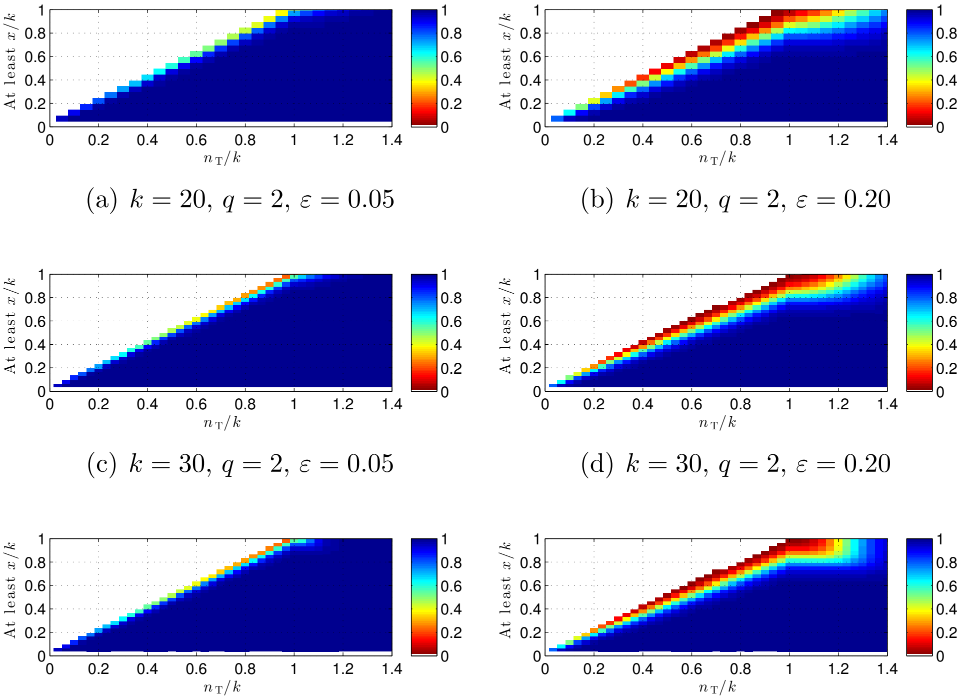
(c) $k = 30$, $q = 2$, $\varepsilon = 0.05$

(d) $k = 30$, $q = 2$, $\varepsilon = 0.20$

(e) $k = 30$, $q = 8$, $\varepsilon = 0.05$

(f) $k = 30$, $q = 8$, $\varepsilon = 0.20$

Figure 4.2: Colour-coded depiction of the probability of recovering at least x source packets when n_{T} coded packets have been transmitted over a packet erasure channel. *Non-systematic* network coding has been assumed and various values for the number of source packets k, the field size q and the erasure probability ε have been used.



(a) $k = 20$, $q = 2$, $\varepsilon = 0.05$

(b) $k = 20$, $q = 2$, $\varepsilon = 0.20$

(c) $k = 30$, $q = 2$, $\varepsilon = 0.05$

(d) $k = 30$, $q = 2$, $\varepsilon = 0.20$

(e) $k = 30$, $q = 8$, $\varepsilon = 0.05$

(f) $k = 30$, $q = 8$, $\varepsilon = 0.20$

Figure 4.3: Colour-coded depiction of the probability of recovering at least x source packets when n_{T} packets have been transmitted over a packet erasure channel. *Systematic* network coding has been assumed and various values for the number of source packets k, the field size q and the erasure probability ε have been used.

Chapter 5

The Multiplicative Matrix Channel

5.1 Overview

In this chapter we consider the Multiplicative Matrix Channel (MMC), defined in Section 1.4, a channel used to model the special case of error-free network coding (see Section 2.4.1). Silva, Kschischang and Kötter [39] analyse the MMC channel capacity, giving upper and lower bounds on the capacity that converge for large field size or large channel input. We improve on the bounds from [39], giving upper and lower bounds that are within a (small) additive constant for any channel parameters and not just in certain asymptotic cases. Thus we determine the behaviour of the channel capacity for all channel parameters.

The chapter is organised as follows. In Section 5.2 we review previously known results on the MMC channel capacity. Section 5.3 gives preliminary results bounding sums of Gaussian binomial coefficients and Section 5.4 uses these results to give a bound on the MMC channel capacity. Finally Section 5.5 concludes the chapter with a discussion about optimal coding schemes.

5.2 Known results on capacity

Recall the definition (Definition 1.4.1) of the MMC channel:

Definition. The *Multiplicative Matrix Channel* (MMC) has input set $\mathcal{X}$ and output set $\mathcal{Y}$, where $\mathcal{X} = \mathcal{Y} = \mathbb{F}_q^{n\times m}$. The channel law is

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} \tag{5.2.1}$$

where $\boldsymbol{A} \in \mathrm{GL}(n, q)$ is chosen uniformly at random.

Since the MMC channel is defined for matrices over a base field of size q, the mutual information of the channel input and output $I(\boldsymbol{X}; \boldsymbol{Y})$ is naturally given in q-ary units. The capacity of a channel is defined as the maximum mutual information over possible input distributions (see Definition 3.4.7), therefore we define the MMC channel capacity in terms of q-ary units.

Definition 5.2.1. Define $C_{\mathrm{MMC}} = \max_{P_{\boldsymbol{X}}} I(\boldsymbol{X}; \boldsymbol{Y})$ to be the capacity of the MMC channel in q-ary units per channel use.

Recall the channel capacity gives the maximum rate at which we can transmit information over the channel, hence C_{MMC} gives the maximum number of q-ary units we can transmit in one use of the MMC channel.

Consider the MMC channel law (5.2.1). Since $\boldsymbol{A}$ is invertible, $\boldsymbol{X}$ and $\boldsymbol{Y}$ share the same rowspace, thus the rowspace of the input is preserved under the channel law. Intuitively this suggests that information should be encoded as a choice of subspace and a maximal set of codewords would contain a unique matrix $\boldsymbol{X}$ of each possible rowspace. In [39] this is shown to be the case and the capacity of the channel is computed in q-ary units per channel use to be

$$C_{\mathrm{MMC}} = \log_q \sum_{k=0}^{n} \begin{bmatrix} m \\ k \end{bmatrix}. \tag{5.2.2}$$

The authors go on to show that for $\frac{n}{m} \leq \frac{1}{2}$ the following bound on the capacity holds,

$$(m-n)n < C_{\mathrm{MMC}} < (m-n)n + \log_q 4(n+1). \tag{5.2.3}$$

This bound is sufficient for the authors to compute the capacity in certain asymptotic cases as follows. Fix n and m, with $\frac{n}{m} \leq \frac{1}{2}$, then

$$\lim_{q \to \infty} C_{\mathrm{MMC}} = (m-n)n. \tag{5.2.4}$$

Now fix q and let $n = \lambda m$ for some constant $\lambda \leq \frac{1}{2}$. Since the capacity naturally scales with nm, the authors in [39] define the normalised capacity to be $\overline{C_{\mathrm{MMC}}} = \frac{1}{nm} C_{\mathrm{MMC}}$. They show

$$\lim_{\substack{m \to \infty \\ n = \lambda m}} \overline{C_{\mathrm{MMC}}} = 1 - \lambda. \tag{5.2.5}$$

The limits (5.2.4) and (5.2.5) follow from (5.2.3) since the last term on the right vanishes in both limiting cases. However in any case of practical interest it will be necessary for the parameter values to be finite. In this case the bounds in (5.2.3) differ by a factor of $\mathcal{O}(\log(n))$, giving little information about the true size of the capacity. Even when we let m and n grow, without normalisation the bounds in (5.2.3) become far apart and we obtain little information.

In the remainder of this chapter, we present an improved upper bound on the MMC channel capacity, which replaces the last term on the right in (5.2.3) with a (small) constant. This decreases the gap between the bounds from $\mathcal{O}(\log(n))$ to a constant, determining that the true capacity of the channel is (very) ‘close’ to $(m-n)n$ for all parameter values q, n and m, given $\frac{n}{m} \leq \frac{1}{2}$. There is no need for normalisation and the distance between the upper and lower bound becomes negligible as any of the parameters grow. Even for small parameter values the distance between the bounds is small, hence our result gives insight on the behaviour of the channel capacity for all parameter values.

5.3 Bounds on sums of Gaussian binomial coefficients

In this section we present results that lead to a constant bound on the sum of Gaussian binomial coefficients. These results will be used in the following section to bound the capacity of the MMC channel.

Lemma 5.3.1. *Let n and m be integers with $\frac{n}{m} \leq \frac{1}{2}$. Then*

$$\sum_{k=0}^{n} \begin{bmatrix} m \\ k \end{bmatrix} < q^{(m-n)n} \prod_{j=1}^{\infty} \frac{1}{1-q^{-j}} \sum_{i=0}^{\infty} \frac{1}{q^{i^2}}. \tag{5.3.1}$$

Proof. It is shown in Lemma 3.2.2 that

$$\begin{bmatrix} m \\ k \end{bmatrix} < q^{(m-k)k} \prod_{j=1}^{\infty} \frac{1}{1-q^{-j}}.$$

Therefore

$$\begin{aligned} \sum_{k=0}^{n} \begin{bmatrix} m \\ k \end{bmatrix} &< \prod_{j=1}^{\infty} \frac{1}{1-q^{-j}} \sum_{k=0}^{n} q^{(m-k)k} \\ &= \prod_{j=1}^{\infty} \frac{1}{1-q^{-j}} \sum_{i=0}^{n} q^{(m-(n-i))(n-i)} \\ &= q^{(m-n)n} \prod_{j=1}^{\infty} \frac{1}{1-q^{-j}} \sum_{i=0}^{n} \frac{1}{q^{mi-2ni+i^2}}. \end{aligned} \tag{5.3.2}$$

Since $\frac{n}{m} \leq \frac{1}{2}$,

$$\begin{aligned} \sum_{i=0}^{n} \frac{1}{q^{mi-2ni+i^2}} &\leq \sum_{i=0}^{n} \frac{1}{q^{i^2}} \\ &\leq \sum_{i=0}^{\infty} \frac{1}{q^{i^2}}. \end{aligned} \tag{5.3.3}$$

Substituting (5.3.3) into (5.3.2) gives the result. □

Remark. The independent work of Gadouleau and Yan [19] considers related bounds on sums of Gaussian binomial coefficients, in the context of studying packing and covering properties of subspace codes. In fact the bound in

Lemma 5.3.1 appears in the proof of [19, Proposition 1] and is a special case, with tighter constants, of [19, Proposition 1 and Proposition 11].

Lemma 5.3.2. *Let n and m be integers with $\frac{n}{m} \leq \frac{1}{2}$. Then*

$$q^{(m-n)n} < \sum_{k=0}^{n} \begin{bmatrix} m \\ k \end{bmatrix} < 7q^{(m-n)n}. \tag{5.3.4}$$

Proof. It is shown in Lemma 3.2.2 that

$$q^{(m-n)n} < \begin{bmatrix} m \\ n \end{bmatrix},$$

the lower bound follows since $\left[\begin{smallmatrix} m \\ n \end{smallmatrix}\right] < \sum_{k=0}^{n} \left[\begin{smallmatrix} m \\ k \end{smallmatrix}\right]$.

By Lemma 5.3.1,

$$\sum_{k=0}^{n} \begin{bmatrix} m \\ k \end{bmatrix} < q^{(m-n)n} \prod_{j=1}^{\infty} \frac{1}{1-q^{-j}} \sum_{i=0}^{\infty} \frac{1}{q^{i^2}}. \tag{5.3.5}$$

Since $q \geq 2$,

$$\begin{aligned} \sum_{i=0}^{\infty} \frac{1}{q^{i^2}} &\leq \sum_{i=0}^{\infty} \frac{1}{2^{i^2}} \\ &< \sum_{i=0}^{\infty} \frac{1}{2^{i}} = 2. \end{aligned} \tag{5.3.6}$$

Furthermore, it is shown in the proof of Lemma 3.2.3 that

$$\prod_{j=1}^{\infty} \frac{1}{1-q^{-j}} \leq \frac{1}{Q_0} < 3.5, \tag{5.3.7}$$

where Q_0 is as defined in Subsection 3.2.1. Substituting (5.3.6) and (5.3.7) into (5.3.5) gives the upper bound. □

Remark. Note that $\sum_{k=0}^{n} \left[\begin{smallmatrix} m \\ k \end{smallmatrix}\right]$ is equal to the number of subspaces of $\mathbb{F}_q^m$ of dimension $\leq n$. Thus Lemma 5.3.2 tells us that the number of subspaces of $\mathbb{F}_q^m$ with dimension bounded by n lies within a small factor of $q^{(m-n)n}$.

The constant factor in the upper bound in Lemma 5.3.2 can be reduced by using numerical computation, giving a tighter bound. As the upper and lower bound already differ by a (small) constant this improvement does not improve our overall knowledge of the behaviour of the sums of Gaussian binomial coefficients. However it does suggest that the true value of the sum lies closer to the lower bound $q^{(m-n)n}$, we show this to be the case particularly as q grows. The computation is described below.

In the proof of Lemma 5.3.2 the following bound is used

$$\sum_{i=0}^{\infty} \frac{1}{2^{i^2}} < \sum_{i=0}^{\infty} \frac{1}{2^i} = 2,$$

however using numerical computation it is possible to show that

$$\sum_{i=0}^{\infty} \frac{1}{2^{i^2}} \approx 1.56447 < 1.5645. \tag{5.3.8}$$

Replacing (5.3.6) by (5.3.8) in the proof of Lemma 5.3.2 gives

$$\sum_{k=0}^{n} \begin{bmatrix} m \\ k \end{bmatrix} < \frac{1.5645}{Q_0} q^{(m-n)n} < 5.4175 q^{(m-n)n}. \tag{5.3.9}$$

The constant can be marginally reduced further by evaluating Q_0 and $\sum_{i=0}^{\infty} \frac{1}{2^{i^2}}$ to a higher number of significant figures.

The bound in (5.3.9) is a marginal improvement over Lemma 5.3.2, however for $q > 2$, numerical computation can lead to a more noticeable improvement in the bound. For a given value q_*, by computationally evaluating the constant

$$\sum_{j=0}^{\infty} \frac{1}{1 - q_*^{-j}} \sum_{i=0}^{\infty} \frac{1}{q_*^{i^2}}$$

we can improve the bound for all $q \geq q_*$. For example with $q_* = 3$ it is possible to show

$$\sum_{j=0}^{\infty} \frac{1}{1 - 3^{-j}} \sum_{i=0}^{\infty} \frac{1}{3^{i^2}} < 1.7854 \cdot 1.3458 < 2.4028,$$

hence for $q \geq 3$

$$q^{(m-n)n} < \sum_{k=0}^{n} \begin{bmatrix} m \\ k \end{bmatrix} < 2.4028q^{(m-n)n} \tag{5.3.10}$$

Similarly, it is possible to show for $q \geq 64$

$$q^{(m-n)n} < \sum_{k=0}^{n} \begin{bmatrix} m \\ k \end{bmatrix} < 1.032q^{(m-n)n}. \tag{5.3.11}$$

This computation shows the upper bound approaches the lower bound very quickly as q grows. Hence, even for relatively small values of q, it is indeed the case that $\sum_{k=0}^{n} \left[\begin{smallmatrix} m \\ k \end{smallmatrix}\right] \approx q^{(m-n)n}$.

5.4 Capacity of the MMC channel

In this section we use the bounds given on q-binomial coefficients in Section 5.3 in order to bound the capacity of the MMC channel. The following theorem gives good bounds on the channel capacity.

Theorem 5.4.1. *Let C_{MMC} be the capacity of the MMC channel, in q-ary units per channel use. If $\frac{n}{m} \leq \frac{1}{2}$ then the channel capacity satisfies the following bound:*

$$(m-n)n < C_{\text{MMC}} < (m-n)n + 3. \tag{5.4.1}$$

Proof. The channel capacity is given in (5.2.2). Taking logarithms to the base q of each term in (5.3.4) gives

$$\begin{aligned} (m-n)n < C_{\text{MMC}} &< (m-n)n + \log_q 7 \\ &\leq (m-n)n + \log_2 7 \\ &< (m-n)n + 3. \end{aligned}$$ □

Theorem 5.4.1 gives bounds on the MMC channel capacity that differ by a small additive constant which is independent of parameter values. This is a significant improvement of the previously known bound (5.2.3), given in [39]. Our bounds determine that $C_{\mathrm{MMC}} \approx (m-n)n$ for all parameter values and not just in certain asymptotic cases.

Remark. The upper bound in Theorem 5.4.1 can be lowered slightly by using the tighter bounds on the sum of Gaussian coefficients, obtained via numerical computation in Section 5.3. Substituting (5.3.9) into the proof of Theorem 5.4.1 gives the improved bound on capacity,

$$\begin{aligned}(m-n)n < C_{\mathrm{MMC}} &< (m-n)n + \log_2(5.4175)\\ &< (m-n)n + 2.438,\end{aligned} \tag{5.4.2}$$

for all possible parameter values q, n and m with $\frac{n}{m} \leq \frac{1}{2}$. Furthermore for $q \geq 3$

$$\begin{aligned}(m-n)n < C_{\mathrm{MMC}} &< (m-n)n + \log_3(2.4028)\\ &< (m-n)n + 0.798,\end{aligned} \tag{5.4.3}$$

and for $q \geq 64$

$$\begin{aligned}(m-n)n < C_{\mathrm{MMC}} &< (m-n)n + \log_{64}(1.032)\\ &< (m-n)n + 0.0076.\end{aligned} \tag{5.4.4}$$

This shows the upper bound approaches the lower bound very quickly as q grows. Recall the MMC channel is most appropriate for modelling network coding when the field size is large (since we assume $\boldsymbol{A}$ is non-singular), in this case the channel capacity is indeed very close to $(m-n)n$. However, Theorem 5.4.1 is enough to show that $C_{\mathrm{MMC}} \approx (m-n)n$ for all parameter values q, n and m with $\frac{n}{m} \leq \frac{1}{2}$.

5.5 Discussion

The results of Section 5.4 give an intuitive interpretation of the channel capacity, by noting

$$C_{\mathrm{MMC}} \approx (m-n)n = mn - n^2,$$

so in the transmission of mn q-ary bits (the input matrix $\boldsymbol{X}$) you lose approximately n^2 q-ary bits which is precisely the amount needed to communicate the transfer matrix $\boldsymbol{A}$. Indeed Silva, Kschischang and Kötter [39] present the following coding scheme, which achieves the rate $(m-n)n$. Let the first n columns of an input matrix $\boldsymbol{X}$ be the $n \times n$ identity matrix so $\boldsymbol{X} = (I_n|\boldsymbol{X}')$ for some data matrix $\boldsymbol{X}'$ of size $n \times (m-n)$. Then

$$\boldsymbol{Y} = \boldsymbol{A}(I_n|\boldsymbol{X}') = (\boldsymbol{A}|\boldsymbol{A}\boldsymbol{X}').$$

The receiver gains knowledge of $\boldsymbol{A}$ and can easily compute $\boldsymbol{A}^{-1}$ in order to obtain $\boldsymbol{X}' = \boldsymbol{A}^{-1}(\boldsymbol{A}\boldsymbol{X}')$ and recover the message. (Note that in the network coding application this is equivalent to using coding headers, e.g. [11].) Our result shows that this coding scheme achieves very close to capacity for all parameter values q, n and m with $\frac{n}{m} \leq \frac{1}{2}$ and not just for certain asymptotic cases.

Recall that [39] shows that when communicating over the MMC channel, information should be encoded as a choice of subspace. Each matrix of the form $\boldsymbol{X} = (I_n|\boldsymbol{X}')$ has a unique rowspace, so this coding scheme conforms with the idea of encoding subspaces. However it does not give a maximal set of codewords as only a restricted set of subspaces are represented. Since this coding scheme is close to optimal, this observation shows that 'nearly all' subspaces of $\mathbb{F}_q^m$ with dimension $\leq n$ can be represented as the rowspace of a matrix of this form.

Note that in this scheme, the input matrix $\boldsymbol{X}$ always has full rank. Thus an (almost) optimal distribution on the input rank is for $\boldsymbol{X}$ to have constant rank $r_{\boldsymbol{X}} = n$. Nobrega, Silva and Uchoa-Filho [33, Th. 4] show within their proof that given an optimal input distribution, the UGR distribution with the same distribution on ranks is also optimal. This implies that a 'good' coding scheme would be to pick input matrices $\boldsymbol{X}$ uniformly from the set of all full rank $n \times m$ matrices, allowing a larger set of codewords than the previous scheme. Indeed it is shown in [33] that constant rank coding schemes are optimal asymptotically. However is it not immediately obvious or intuitive how one would decode with a uniform or UGR input coding scheme. For the simple MMC channel a UGR distribution may give the best rates, but may be unnecessarily complicated to implement. Therefore for practical purposes it is sensible to adopt the simple scheme from [39] described above, which will achieve close to optimal rates for any parameter values.

Chapter 6

The Additive Matrix Channel

6.1 Overview

In this chapter we consider the Additive Matrix Channel (AMC), defined in Section 1.5, a channel used to model coherent network coding (see Section 2.4.2). As discussed in Section 1.5, the AMC channel we present is a generalisation of the channel considered by Silva, Kschischang and Kötter [39, §IV], who assume a fixed constant error rank. In [39] the authors analyse the capacity of the AMC channel with fixed error rank, giving upper and lower bounds on the capacity that converge for large field size or large channel input. We improve on the bounds from [39], giving upper and lower bounds that are within a (small) additive constant for any channel parameters and not just in certain asymptotic cases, thus determining the capacity's behaviour for all channel parameters. We then present similar results for the AMC channel with uniform error, which leads to a lower bound for the capacity of the general AMC channel. Our results show that the minimum capacity of the general AMC channel is very close to the capacity of the channel with fixed error rank, thus our generalisation covers a wider class of cases without any significant loss in capacity.

The chapter is organised as follows. In Section 6.2 we express the AMC channel capacity in terms of the channel parameters. Section 6.3 considers the AMC channel with fixed error rank. We review the known results on the channel capacity and present our improved bound. Section 6.4 considers the AMC channel with uniform error matrix. The uniform error distribution is an important special case of the AMC channel as its capacity provides a lower bound for the AMC channel capacity (Lemma 6.4.1). We compute the channel capacity in this case and provide a constant bound similar to that for the capacity of the channel with fixed error rank. In Section 6.5 we use the results of Section 6.4 to give a lower bound on the capacity of the general AMC channel. Finally, in Section 6.6 we explain the consequences the results from Chapters 5 and 6 have on the analysis of the Gamma channel capacity (see Chapter 8).

6.2 The AMC channel capacity

Recall the definition (Definition 1.5.1) of the AMC channel:

Definition. Let $\mathcal{R}$ be a probability distribution on the set $\{0, 1, \dots, \min\{m, n\}\}$ of possible ranks of matrices $M \in \mathbb{F}_q^{n\times m}$. The *Additive Matrix Channel with rank error distribution* $\mathcal{R}$ *(*AMC($\mathcal{R}$)*)* has input set $\mathcal{X}$ and output set $\mathcal{Y}$, where $\mathcal{X} = \mathcal{Y} = \mathbb{F}_q^{n\times m}$. The channel law is

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{B}$$

where $\boldsymbol{B} \in \mathbb{F}_q^{n\times m}$ is UGR with rank distribution $\mathcal{R}$.

As for the MMC channel (see Section 5.2), since the AMC channel is defined for matrices over a base field of size q, we define the AMC channel capacity in terms of q-ary units.

Definition 6.2.1. Define $C_{\mathrm{AMC}(\mathcal{R})} = \max_{P_{\boldsymbol{X}}} I(\boldsymbol{X};\boldsymbol{Y})$ to be the capacity of the AMC($\mathcal{R}$) channel, in q-ary units per channel use.

Recall the channel capacity gives the maximum rate at which we can transmit information over the channel, hence $C_{\mathrm{AMC}(\mathcal{R})}$ gives the maximum number of q-ary units we can transmit in one use of the AMC($\mathcal{R}$) channel. The following lemma gives $C_{\mathrm{AMC}(\mathcal{R})}$ in terms of the channel parameters.

Lemma 6.2.1. *Let $C_{\mathrm{AMC}(\mathcal{R})}$ be the* AMC($\mathcal{R}$) *channel capacity in q-ary units per channel use. Then*

$$\begin{aligned} C_{\mathrm{AMC}(\mathcal{R})} &= nm - H(\boldsymbol{B}) && (6.2.1)\\ &= nm - \sum_{r=0}^{\min\{n,m\}} \mathcal{R}(r)\log_q \frac{|\mathbb{F}_q^{n\times m,r}|}{\mathcal{R}(r)}. && (6.2.2)\end{aligned}$$

Proof. Expanding the mutual information,

$$\begin{aligned} C_{\mathrm{AMC}(\mathcal{R})} &= \max_{P_{\boldsymbol{X}}}\{I(\boldsymbol{X}:\boldsymbol{Y})\}\\ &= \max_{P_{\boldsymbol{X}}}\{H(\boldsymbol{Y}) - H(\boldsymbol{Y}|\boldsymbol{X})\}\\ &= \max_{P_{\boldsymbol{X}}}\{H(\boldsymbol{Y})\} - H(\boldsymbol{B}) && (6.2.3)\\ &= nm - H(\boldsymbol{B}), && (6.2.4)\end{aligned}$$

where (6.2.3) holds since $\boldsymbol{X}$ and $\boldsymbol{B}$ are independent, so $H(\boldsymbol{Y}|\boldsymbol{X}) = H(\boldsymbol{B})$ and $H(\boldsymbol{B})$ does not depend on $P_{\boldsymbol{X}}$, and (6.2.4) holds by taking a uniform output distribution to maximise $H(\boldsymbol{Y}) = nm$. Hence (6.2.1) holds. Now, since $\boldsymbol{B}$ has a UGR distribution with rank distribution $\mathcal{R}$, we have

$$\Pr(\boldsymbol{B} = B) = \frac{\mathcal{R}(\mathrm{rk}(B))}{|\mathbb{F}_q^{n\times m,\mathrm{rk}(B)}|}.$$

Therefore

$$\begin{aligned}
H(\boldsymbol{B}) &= -\sum_{B\in\mathbb{F}_q^{n\times m}} \Pr(\boldsymbol{B}=B)\log_q \Pr(\boldsymbol{B}=B) \\
&= -\sum_{B\in\mathbb{F}_q^{n\times m}} \frac{\mathcal{R}(\mathrm{rk}(B))}{|\mathbb{F}_q^{n\times m,\mathrm{rk}(B)}|}\log_q \frac{\mathcal{R}(\mathrm{rk}(B))}{|\mathbb{F}_q^{n\times m,\mathrm{rk}(B)}|} \\
&= -\sum_{r=0}^{\min\{n,m\}} \mathcal{R}(r)\log_q \frac{\mathcal{R}(r)}{|\mathbb{F}_q^{n\times m,r}|} \\
&= \sum_{r=0}^{\min\{n,m\}} \mathcal{R}(r)\log_q \frac{|\mathbb{F}_q^{n\times m,r}|}{\mathcal{R}(r)} \qquad (6.2.5)
\end{aligned}$$

Substituting (6.2.5) into (6.2.4) gives (6.2.2). □

Remark. The same result is proved by Silva, Kschischang and Kötter [39] for the AMC channel with fixed error rank, which is considered in Section 6.3.

6.3 The AMC channel with fixed error rank

In this section we focus on the special case of the Additive Matrix Channel, where the rank of the error matrix $\boldsymbol{B}$ is fixed. This is exactly the AMC channel considered by Silva, Kschischang and Kötter [39, §IV]. We give bounds on the channel capacity that differ by a small additive constant which is independent of all channel parameters q, n, m and t, thus we determine the behaviour of the channel capacity for all parameter choices. Below we give the definition of the channel.

Definition 6.3.1. The *Additive Matrix Channel with fixed error rank (*AMC($\mathcal{R}_t$)*)* is the Additive Matrix Channel with rank distribution $\mathcal{R} = \mathcal{R}_t$ which takes $\mathrm{rk}(\boldsymbol{B}) = t$ with probability 1.

Remark. This model is more restrictive than the general AMC channel as it assumes the error matrix $\boldsymbol{B}$ has rank exactly t. As discussed in Subsection 2.4.2, for the network coding application this implies there are always exactly t linearly independent random errors introduced into the network.

The remainder of this section is organised as follows. In Subsection 6.3.1 we review the known results about the AMC$(\mathcal{R}_t)$ channel capacity. In Subsection 6.3.2 we present some preliminary results which are used in Subsection 6.3.3 to give bounds on the channel capacity. We conclude Subsection 6.3.3 with a discussion of the results.

6.3.1 Known results on capacity

In [39] the capacity of the AMC$(\mathcal{R}_t)$ channel is computed in q-ary units per channel use to be

$$C_{\text{AMC}(\mathcal{R}_t)} = (m-t)(n-t) + \log_q \prod_{i=0}^{t-1} \frac{(1-q^{i-t})}{(1-q^{i-n})(1-q^{i-m})}. \tag{6.3.1}$$

This is sufficient for the authors to compute the capacity in certain asymptotic cases as follows. Fix n and m, then

$$\lim_{q\to\infty} C_{\text{AMC}(\mathcal{R}_t)} = (m-t)(n-t).$$

Now fix q and let $n = \lambda m$, $t = \tau n$ for some constants $\lambda \leq 1$, $\tau < 1$. Since the capacity naturally scales with nm, the authors in [39] define the normalised capacity to be $\overline{C_{\text{AMC}(\mathcal{R}_t)}} = \frac{1}{nm} C_{\text{AMC}(\mathcal{R}_t)}$, then

$$\lim_{\substack{m\to\infty \\ n=\lambda m \\ t=\tau n}} \overline{C_{\text{AMC}(\mathcal{R}_t)}} = (1-\lambda\tau)(1-\tau).$$

These limits follow from (6.3.1) since the last term on the right vanishes in both limiting cases. However (6.3.1) gives little intuitive information about the size of the capacity for general parameters. This motivates the remainder of this section, which presents improved bounds on the capacity that are ‘close’ for any channel parameters.

6.3.2 Preliminary results

In this section we present several preliminary results which will be used in subsequent sections to bound the AMC channel capacity in various cases.

The following two lemmas bound the term $\prod_{i=0}^{t-1} \frac{(1-q^{i-t})}{(1-q^{i-n})(1-q^{i-m})}$ to lie within a small constant interval. This is used to give a constant bound on the last term on the right hand side of (6.3.1).

Lemma 6.3.1. *Let q, m, n, t be integers with $q \geq 2$ and $t < n, m$. Then*

$$\prod_{i=0}^{t-1} \frac{(1-q^{i-t})}{(1-q^{i-n})(1-q^{i-m})} < 1. \tag{6.3.2}$$

Proof. For $i \in \{0, \dots, t-1\}$,

$$\begin{aligned}
(1-q^{i-n})(1-q^{i-m}) &\geq (1-q^{i-\min\{n,m\}})^2 \\
&= 1 - 2q^{i-\min\{n,m\}} + q^{2(i-\min\{n,m\})} \\
&> 1 - 2q^{i-\min\{n,m\}} \\
&\geq 1 - q^{i-(\min\{n,m\}-1)} && (6.3.3) \\
&\geq 1 - q^{i-t}, && (6.3.4)
\end{aligned}$$

where (6.3.3) holds since $q \geq 2$ and (6.3.4) holds since $t < n, m$. Thus for $i \in \{0, \dots, t-1\}$

$$\frac{(1-q^{i-t})}{(1-q^{i-n})(1-q^{i-m})} < 1,$$

and therefore

$$\prod_{i=0}^{t-1} \frac{(1-q^{i-t})}{(1-q^{i-n})(1-q^{i-m})} < \prod_{i=0}^{t-1} 1 = 1. \qquad \square$$

Lemma 6.3.2. *Let q, m, n, t be integers with $q \geq 2$, then*

$$\prod_{i=0}^{t-1} \frac{(1-q^{i-t})}{(1-q^{i-n})(1-q^{i-m})} > Q_0, \tag{6.3.5}$$

where $Q_0 \approx 0.28879$ is as defined in Section 3.2.1.

Proof. Observe that

$$\begin{aligned}
\prod_{i=0}^{t-1} \frac{(1-q^{i-t})}{(1-q^{i-n})(1-q^{i-m})} &> \prod_{i=0}^{t-1} \left(1-q^{i-t}\right) \\
&\geq \prod_{i=0}^{t-1} \left(1-2^{i-t}\right) \qquad (6.3.6) \\
&= \prod_{k=1}^{t} \left(1-2^{-k}\right) \\
&\geq \prod_{k=1}^{\infty} \left(1-2^{-k}\right) \\
&= Q_0,
\end{aligned}$$

where (6.3.6) follows since $q \geq 2$. □

The following lemma gives a constant bound on the last term on the right hand side of (6.3.1).

Lemma 6.3.3. *Let q, m, n, t be integers with $q \geq 2$ and $t < n, m$. Then*

$$-2 < \log_q \prod_{i=0}^{t-1} \frac{(1-q^{i-t})}{(1-q^{i-n})(1-q^{i-m})} < 0. \qquad (6.3.7)$$

Proof. The upper bound follows from Lemma 6.3.1, by noting that $\log_q(1) = 0$. Then, by Lemma 6.3.2,

$$\begin{aligned}
\log_q \left(\prod_{i=0}^{t-1} \frac{(1-q^{i-t})}{(1-q^{i-n})(1-q^{i-m})} \right) &> \log_q(Q_0) \\
&\geq \log_2(Q_0) \\
&> -2,
\end{aligned}$$

hence the lower bound holds. □

Remark. Note that the constant lower bound on the left hand side of (6.3.7) can be increased slightly from -2 to -1.792 by evaluating $\log_2(Q_0)$ with higher accuracy. Further more, as in Section 5.3, by using numerical evaluation for higher values of q, the constant bound in (6.3.5) can be increased and quickly approaches the upper bound as q grows. As we are aiming for a bound that is independent of all parameters we omit these details.

6.3.3 Bounds on capacity

The following theorem gives a constant bound on the AMC($\mathcal{R}_t$) channel capacity.

Theorem 6.3.4. *Let $C_{\text{AMC}(\mathcal{R}_t)}$ be the* AMC($\mathcal{R}_t$) *channel capacity in q-ary units per channel use. Then*

$$(m-t)(n-t) - 2 < C_{\text{AMC}(\mathcal{R}_t)} < (m-t)(n-t). \tag{6.3.8}$$

Proof. Substituting the bounds from (6.3.7) into the expression for the channel capacity given in (6.3.1) gives the result. □

Remark. The constant 2 on the left hand side of (6.3.8) can be decreased slightly to tighten the bound as described in the remark at the end of Section 6.3.2. This does not improve our overall understanding of the channel behaviour, so we omit the details.

Theorem 6.3.4 shows that the capacity of the AMC($\mathcal{R}_t$) channel is 'close to' $(m-t)(n-t)$ for all parameter values q, n, m, t. This is a significant improvement of the results of [39] which focused on certain asymptotic cases. This result gives an intuitive interpretation of the channel capacity, by noting

$$C_{\text{AMC}(\mathcal{R}_t)} \approx (m-t)(n-t) = mn - (m+n-t)t,$$

so in the transmission of mn q-ary bits (the input matrix $\boldsymbol{X}$) you lose approximately $(m+n-t)t$ q-ary bits. This is shown to be approximately the amount needed to specify an $n \times m$ matrix of rank t in [39, §IV]. Thus, the capacity can be interpreted as the number of q-ary bits transmitted minus the number of bits needed to communicate $\boldsymbol{B}$. An efficient coding scheme that achieves the rate $(m-t)(n-t)$ for the asymptotic cases discussed in Subsection 6.3.1 is presented in [39, §IV. B].

In the following section we move on to consider the AMC channel in the special case when the error matrix $\boldsymbol{B}$ is chosen uniformly, given $\text{rk}(\boldsymbol{B}) \leq t$. We will consider the channel capacity in this case and how it relates to the general AMC channel capacity.

6.4 The AMC channel with uniform error matrix

In this section we consider the special case of the additive matrix channel when the error matrix $\boldsymbol{B}$ is chosen uniformly from the set of all $n \times m$ matrices with rank $\leq t$. We show that the capacity of this channel gives a lower bound for the AMC($\mathcal{R}$) capacity, with general $\mathcal{R}$. We then give bounds on the channel capacity, thus leading to a lower bound on the general AMC channel capacity.

The remainder of this section is organised as follows. In Subsection 6.4.1 we formally define the AMC channel with uniform error matrix and show that its capacity is a lower bound for the capacity of the general AMC channel. In Subsection 6.4.2 we calculate the exact channel capacity. Subsection 6.4.3 presents some preliminary results which are used in Subsection 6.4.4 to give bounds on the channel capacity.

6.4.1 Uniform error is a lower bound for capacity

Let $\mathcal{R}_U$ denote the probability distribution on the set $\{0, 1, \ldots, \min\{m, n\}\}$ of possible ranks of matrices $M \in \mathbb{F}_q^{n\times m}$ such that

$$\mathcal{R}_U(r) = \begin{cases} \frac{|\mathbb{F}_q^{n\times m,r}|}{\sum_{k=0}^{t} |\mathbb{F}_q^{n\times m,k}|} & \text{if } 0 \leq r \leq t \\ 0 & \text{otherwise.} \end{cases} \tag{6.4.1}$$

Observe that a UGR distribution with rank distribution $\mathcal{R}_U$ on $\mathbb{F}_q^{n\times m}$, is just the uniform distribution on the set of $n \times m$ matrices with rank $\leq t$.

Definition 6.4.1. The *Additive Matrix Channel with uniform error matrix (*AMC($\mathcal{R}_U$)*)* is the Additive Matrix Channel with rank distribution $\mathcal{R} = \mathcal{R}_U$, where $\mathcal{R}_U$ is given in (6.4.1).

The following lemma shows the AMC($\mathcal{R}_U$) channel capacity is a lower bound for the additive matrix channel capacity, given that the AMC error rank is bounded by t.

Lemma 6.4.1. *Given an integer t, between* 0 *and* $\min\{n, m\}$*, let $\mathcal{R}$ be any rank distribution such that $\mathcal{R}(r) = 0$ for $r > t$. Let $C_{\mathrm{AMC}(\mathcal{R})}$ and $C_{\mathrm{AMC}(\mathcal{R}_U)}$ denote the capacities (in q-ary units per channel use) of the* AMC($\mathcal{R}$) *and* AMC($\mathcal{R}_U$) *channels, respectively. The capacity of the additive matrix channel is bounded below by the capacity of the additive matrix channel with uniform error matrix, that is*

$$C_{\mathrm{AMC}(\mathcal{R})} \geq C_{\mathrm{AMC}(\mathcal{R}_U)}. \tag{6.4.2}$$

Proof. By Lemma 6.2.1, $C_{\mathrm{AMC}(\mathcal{R})} = nm - H(\boldsymbol{B})$, which is minimal when $H(\boldsymbol{B})$ is maximal. By Lemma 3.4.2, the entropy of $\boldsymbol{B}$ is maximal when $\boldsymbol{B}$ is chosen uniformly, hence when $\mathcal{R} = \mathcal{R}_U$. □

In the following section we compute the exact capacity of the additive matrix channel with uniform error matrix.

6.4.2 Channel capacity

The following lemma gives the capacity of the AMC($\mathcal{R}_U$) channel.

Lemma 6.4.2. *Let $C_{\mathrm{AMC}(\mathcal{R}_U)}$ be the* AMC($\mathcal{R}_U$) *channel capacity in q-ary units per channel use. Then*

$$C_{\mathrm{AMC}(\mathcal{R}_U)} = (m-t)(n-t) - \log_q \left(\sum_{j=0}^{t} \left(\frac{1}{q^{(m+n-2t+j)j}} \prod_{i=0}^{j-1} \frac{(1-q^{i-m})(1-q^{i-n})}{(1-q^{i-j})} \right) \right). \tag{6.4.3}$$

Proof. By Lemma 6.2.1,

$$\begin{aligned} C_{\mathrm{AMC}(\mathcal{R}_U)} &= nm - \sum_{r=0}^{\min\{n,m\}} \mathcal{R}_U(r) \log_q \frac{|\mathbb{F}_q^{n\times m,r}|}{\mathcal{R}_U(r)} \\ &= nm - \log_q \sum_{k=0}^{t} |\mathbb{F}_q^{n\times m,k}|, \end{aligned} \tag{6.4.4}$$

where the second equality follows by substituting in the value of $\mathcal{R}_U(r)$ from (6.4.1). It is well known (e.g. [39, §IV]) that

$$|\mathbb{F}_q^{n\times m,k}| = q^{(m+n-k)k} \prod_{i=0}^{k-1} \frac{(1-q^{i-m})(1-q^{i-n})}{(1-q^{i-k})},$$

thus

$$\sum_{k=0}^{t} |\mathbb{F}_q^{n\times m,k}| = \sum_{k=0}^{t} q^{(m+n-k)k} \prod_{i=0}^{k-1} \frac{(1-q^{i-m})(1-q^{i-n})}{(1-q^{i-k})}. \tag{6.4.5}$$

Now let $f(k)$ be any function of k. Then

$$\begin{aligned} \sum_{k=0}^{t} q^{(m+n-k)k} f(k) &= \sum_{j=0}^{t} q^{(m+n-(t-j))(t-j)} f(t-j) \\ &= q^{(m+n-t)t} \sum_{j=0}^{t} q^{jt-(m+n-(t-j))j} f(t-j) \\ &= q^{(m+n-t)t} \sum_{j=0}^{t} \frac{1}{q^{(m+n-2t+j)j}} f(t-j). \end{aligned} \tag{6.4.6}$$

Substituting (6.4.6) into (6.4.5), with $f(k) = \prod_{i=0}^{k-1} \frac{(1-q^{i-m})(1-q^{i-n})}{(1-q^{i-k})}$, gives

$$\begin{aligned}
&\log_q \sum_{k=0}^{t} |\mathbb{F}_q^{n\times m,k}| \\
&= \log_q \left(q^{(m+n-t)t} \sum_{j=0}^{t} \frac{1}{q^{(m+n-2t+j)j}} \prod_{i=0}^{t-j-1} \frac{(1-q^{i-m})(1-q^{i-n})}{(1-q^{i-(t-j)})} \right) \\
&= (m+n-t)t + \log_q \left(\sum_{j=0}^{t} \frac{1}{q^{(m+n-2t+j)j}} \prod_{i=0}^{t-j-1} \frac{(1-q^{i-m})(1-q^{i-n})}{(1-q^{i-(t-j)})} \right).
\end{aligned} \tag{6.4.7}$$

Substituting (6.4.7) into (6.4.4) gives the result. □

6.4.3 Preliminary results

This section presents preliminary results that will be used to bound the capacity of the AMC channel with uniform error matrix.

Lemma 6.4.3. *Let q, m, n, t be integers with $q \geq 2$ and $t < n, m$. The following holds:*

$$1 < \sum_{j=0}^{t} \frac{1}{q^{(m+n-2t+j)j}} \prod_{i=0}^{t-j-1} \frac{(1-q^{i-m})(1-q^{i-n})}{(1-q^{i-(t-j)})} < 7. \tag{6.4.8}$$

Proof. By Lemma 6.3.1 and Lemma 6.3.2, for $k < n, m$

$$Q_0 < \prod_{i=0}^{k-1} \frac{(1-q^{i-k})}{(1-q^{i-n})(1-q^{i-m})} < 1,$$

hence for $j = 0, \dots, t$, letting $k = t - j$, we see

$$1 < \prod_{i=0}^{t-j-1} \frac{(1-q^{i-m})(1-q^{i-n})}{(1-q^{i-(t-j)})} < \frac{1}{Q_0} < 3.5.$$

Therefore

$$\begin{aligned}
\sum_{j=0}^{t} \frac{1}{q^{(m+n-2t+j)j}} &< \sum_{j=0}^{t} \frac{1}{q^{(m+n-2t+j)j}} \prod_{i=0}^{t-j-1} \frac{(1-q^{i-m})(1-q^{i-n})}{(1-q^{i-(t-j)})} && (6.4.9) \\
&< 3.5 \sum_{j=0}^{t} \frac{1}{q^{(m+n-2t+j)j}}. && (6.4.10)
\end{aligned}$$

The lower bound follows immediately from the left hand side of (6.4.9), by observing that all terms in the sum are positive and the term for $j = 0$ is equal to 1. Now, since $t < n, m$,

$$\begin{aligned} \sum_{j=0}^{t} \frac{1}{q^{(m+n-2t+j)j}} &< \sum_{j=0}^{t} \frac{1}{q^{j^2}} \\ &\leq \sum_{j=0}^{\infty} \frac{1}{q^{j^2}} \\ &< 2. \end{aligned} \tag{6.4.11}$$

Where (6.4.11) follows from the argument of (5.3.6) in the proof of Lemma 5.3.2. Substituting (6.4.11) into (6.4.10) gives the upper bound. □

6.4.4 Bounds on capacity

The following theorem gives bounds on the AMC($\mathcal{R}_U$) channel capacity that differ by a small additive constant.

Theorem 6.4.4. *Let $C_{\mathrm{AMC}(\mathcal{R}_U)}$ be the capacity (in q-ary units per channel use) of the* AMC($\mathcal{R}_U$) *channel. Then $C_{\mathrm{AMC}(\mathcal{R}_U)}$ satisfies*

$$(m-t)(n-t) - 3 < C_{\mathrm{AMC}(\mathcal{R}_U)} < (m-t)(n-t). \tag{6.4.12}$$

Proof. Substituting the bound from Lemma 6.4.3 into the expression for the capacity given in Lemma 6.4.2, gives

$$(m-t)(n-t) - \log_q(7) < C_{\mathrm{AMC}(\mathcal{R}_U)} < (m-t)(n-t) - \log_q(1).$$

The upper bound follows immediately since $\log_q(1) = 0$. The lower bound follows since

$$\log_q(7) \geq \log_2(7) > 3.$$ □

Remark. As in the case of the fixed error rank AMC channel, the constant 3 on the left hand side of (6.4.12) can be decreased slightly by tightening the bound used from (6.4.8). This change is minimal so again the details are omitted.

Theorem 6.4.4 shows that $C_{\mathrm{AMC}(\mathcal{R}_U)} \approx (m-t)(n-t)$ for all parameter values q, m, n, t with $q \geq 2$ and $t < n, m$. Comparing this to Theorem 6.3.4 shows that the capacity of the AMC channel is very similar for both rank error distributions $\mathcal{R}_U$ and $\mathcal{R}_t$.

By Lemma 6.4.1 the capacity of the AMC($\mathcal{R}_U$) channel can be viewed as a lower bound for the capacity of the general AMC($\mathcal{R}$) channel when the error rank is bounded by t. Therefore Theorem 6.4.4 has an immediate consequence for the general AMC($\mathcal{R}$) channel, as is described in the following section.

6.5 Bounds on the AMC capacity

The following theorem gives a lower bound on the AMC channel capacity in the case when the error rank is bounded by t. Note that in any practical implementation, in order to recover the source message one would require the number of errors to be bounded, or at the very least the probability of having more than some fixed number of errors to tend to zero. Therefore this case can be viewed as the general case. The lower bound we present is 'close to' $(m-t)(n-t)$. For special cases of the AMC channel we have shown that the capacity is also bounded above by $(m-t)(n-t)$. Note that for the general channel, the tightest possible upper bound on capacity is $C_{\mathrm{AMC}(\mathcal{R})} \leq nm$. Indeed $C_{\mathrm{AMC}(\mathcal{R})} = nm$ when $\mathcal{R} = \mathcal{R}_0$ is the rank distribution that takes $\mathrm{rk}(\boldsymbol{B}) = 0$ with probability 1. Therefore we cannot hope for a similar upper bound in the general case.

Theorem 6.5.1. *Given an integer t such that $0 \leq t < \min\{n, m\}$, let $\mathcal{R}$ be a rank distribution such that $\mathcal{R}(r) = 0$ for all $r > t$. Let $C_{\mathrm{AMC}(\mathcal{R})}$ be the capacity (in q-ary units per channel use) of the AMC($\mathcal{R}$) channel. Then $C_{\mathrm{AMC}(\mathcal{R})}$ satisfies*

$$C_{\mathrm{AMC}(\mathcal{R})} > (m-t)(n-t) - 3.$$

Proof. Substituting the lower bound on the left hand side of (6.4.12) into (6.4.2) gives the result. □

Comparing Theorem 6.5.1 to Theorem 6.3.4 shows that the lower bound on the capacity of the general additive matrix channel is extremely close to that with a fixed error rank (indeed it drops by just 1 q-ary unit per channel use). Therefore our generalisation of the model studied by Silva, Kschischang and Kötter [39] gives a more realistic model for coherent network coding whilst (almost) maintaining the full channel capacity.

Note, the asymptotic coding scheme (mention in Subsection 6.3.3) for the AMC($\mathcal{R}_t$) channel of rate $(m-t)(n-t)$ presented in [39, §IV. B], can still be applied to the general AMC channel, although decoding errors may occur, see [39, §VI. D] for further details.

Recall in Subsection 6.3.3 we discuss an intuitive interpretation of the AMC($\mathcal{R}_t$) channel capacity as the number of q-ary bits transmitted ($= nm$) minus the number of bits needed to communicate an $n \times m$ matrix of rank t ($\approx (m+n-t)t$). Applying this interpretation to Theorem 6.5.1 shows that for the general AMC($\mathcal{R}$) channel, the maximum loss in capacity from the number of q-ary bits transmitted is still approximately the number of bits needed to communicate an $n \times m$ matrix of rank t. Intuitively this is the best we could hope for in general, since we allow the possibility of the error matrix $\boldsymbol{B}$ having rank t.

6.6 Consequences for more general models

Consider the Gamma channel $\Gamma(\mathcal{R})$ defined in Chapter 1. The results obtained in Chapters 5 and 6, bounding the capacity of the MMC and AMC channels, have an immediate consequence for the known bounds on the Gamma channel. Consider the special case of the Gamma channel with fixed error rank $\mathcal{R}_t$ which takes $\mathrm{rk}(\boldsymbol{B})) = t$ with probability 1, we shall denote this channel $\Gamma(\mathcal{R}_t)$. This is equivalent to the AMMC channel considered by Silva, Kschischang and Kötter [39, §V].

In [39] the following bound on the capacity of the $\Gamma(\mathcal{R}_t)$ channel is shown. Let $C_{\Gamma(\mathcal{R}_t)}$ denote the $\Gamma(\mathcal{R}_t)$ channel capacity, in q-ary units per channel use. Then for $\frac{n}{m} \leq \frac{1}{2}$, given any $\epsilon \geq 0$

$$(m-n)(n-t-\epsilon t) - \log_q 4 - \frac{2tnm}{q^{1+\epsilon t}} \leq C_{\Gamma(\mathcal{R}_t)} \leq (m-n)(n-t) + \log_q 4(1+n)(1+t). \tag{6.6.1}$$

This is sufficient for the authors to compute the capacity in certain asymptotic cases as follows. Fix n and m. Then

$$\lim_{q \to \infty} C_{\Gamma(\mathcal{R}_t)} = (m-n)(n-t).$$

Now define the normalised capacity to be $\overline{C_{\Gamma(\mathcal{R}_t)}} = \frac{1}{nm} C_{\Gamma(\mathcal{R}_t)}$. Fix q and let $n = \lambda m$, $t = \tau n$ for some constants $\lambda \leq 1$, $\tau < 1$,

$$\lim_{\substack{m \to \infty \\ n = \lambda m \\ t = \tau n}} \overline{C_{\Gamma(\mathcal{R}_t)}} = (1-\lambda)(1-\tau).$$

These limits follow from (6.6.1) by choosing an appropriate $\epsilon \geq 0$ in each case. However for general parameters, the bounds in (6.6.1) can be far apart and therefore give little information on the true behaviour of the capacity of the channel. Using the results from Chapters 5 and 6 together with the proof strategy from [39] it is possible to replace the upper bound in (6.6.1) by

$$C_{\Gamma(\mathcal{R}_t)} < (m-n)(n-t) + 6. \tag{6.6.2}$$

However a similar improvement is not immediately possible for the lower bound. As this results in bounds on the capacity that still differ by a factor of $\mathcal{O}(n)$, there is not a significant overall improvement to the known bounds. For this reason we omit the full details. In the following chapters we explore a different method for analysing the capacity of the Gamma channel, determining the exact channel capacity for all parameter values.

Chapter 7

Matrix Functions

7.1 Overview

Consider the following three matrix functions.

- Let U be a subspace of $\mathbb{F}_q^m$ of dimension u. Define $f_0(u)$ to be the number of $n \times m$ matrices whose rowspace is U.

- Let U and V be subspaces of $\mathbb{F}_q^m$ of dimensions u and v respectively. Let $h = \dim(U \cap V)$. Let $M \in \mathbb{F}_q^{n\times m}$ be a fixed matrix such that $\mathrm{Row}(M) = U$. Let r be a non-negative integer. Define $f_1(u, v, h; r)$ to be the number of matrices $B \in \mathbb{F}_q^{n\times m,r}$, such that the rowspace of $M+B$ is V.

- Let r, r_X and r_B be non-negative integers. Let $X \in \mathbb{F}_q^{n\times m}$ be a fixed matrix such that $\mathrm{rk}(X) = r_X$. Define $f_2(r, r_X, r_B)$ to be the number of matrices $B \in \mathbb{F}_q^{n\times m,r_B}$ such that the rank of $X + B$ is equal to r.

The aim of this chapter is to show that the functions f_0, f_1, f_2 are well defined (i.e. they depend only on their inputs and the parameters n, m, q) and to express them in terms of these parameters, showing that they can be efficiently computed. By an efficient computation, we mean a polynomial (in $\max\{n, m\}$) number of arithmetic operations.

The fact that f_0 and f_2 could be efficiently computed was already known from [14] and [18], respectively. However, the expressions given differ from those we develop here. We will further discuss these related works in Section 7.4.

The functions f_0, f_1, f_2 will be used in Chapter 8 to calculate the exact capacity of the Gamma channel.

We approach the problem of determine the expressions for f_0, f_1, f_2 by first exploring several combinatorial results that will be needed. In Section 7.2, for fixed subspaces $U, V \subseteq \mathbb{F}_q^m$, we count the number of subspaces $W \subseteq \mathbb{F}_q^m$ such that the dimension of W, $W \cap U$, $W \cap V$ and $W \cap U \cap V$ are all fixed. In Section 7.3, for fixed subspaces $U, V, W \subseteq \mathbb{F}_q^m$ with $W + V = U + V$, we count the number of pairs (V', W') such that $V' \subseteq V$, $W' \subseteq W$, $U + V' = W' + V'$ and the dimensions of V', W' and $V' \cap W'$ are fixed. Finally in Section 7.4 we calculate the values of the functions f_0, f_1, f_2.

7.2 Counting subspaces

Let U and V be fixed subspaces of $\mathbb{F}_q^m$. The aim of this section is to count the number of subspaces $W \subseteq \mathbb{F}_q^m$ such that the dimension of W, $W \cap U$, $W \cap V$ and $W \cap U \cap V$ are all fixed; this is a key result which will be used in the proof of a later result.

In order to count such spaces W we begin, in Subsection 7.2.1, by focusing on the special case when $W \subseteq U \oplus V$ and W intersects both U and V trivially. Then in Subsection 7.2.2 we relax the condition on the intersection of W with U and V. In Subsection 7.2.3 we consider the case when $W \subseteq U+V$ and finally in Subsection 7.2.4 we consider the general case of the counting problem.

7.2.1 The special case with a direct sum and trivial intersection

Lemma 7.2.1. *Let U, V be subspaces of $\mathbb{F}_q^m$ of dimensions d_U and d_V respectively. The number of d_W-dimensional spaces W such that $W \subseteq U \oplus V$ with $W \cap U = W \cap V = \{0\}$ is*

$$f(d_U, d_V, d_W) = \begin{bmatrix} d_U \\ d_W \end{bmatrix} \prod_{i=0}^{d_W-1} (q^{d_V} - q^i)$$

when $0 \leq d_W \leq \min\{d_U, d_V\}$, and $f(d_U, d_V, d_W) = 0$ otherwise.

Proof. Consider the natural map $\varphi_V : U \oplus V \to V$, note $\ker(\varphi_V) = U$. We have

$$\begin{aligned} \dim(\varphi_V(W)) &= \dim(W) - \dim(U \cap W) \\ &= d_W. \end{aligned}$$

Then since $\dim(\varphi_V(U \oplus V)) = d_V$ we have that $d_W \leq d_V$ and a similar argument shows $d_W \leq d_U$. Thus $0 \leq d_W \leq \min\{d_U, d_V\}$, as required.

Now assume that $0 \leq d_W \leq \min\{d_U, d_V\}$. There are $\left[\begin{smallmatrix} d_U \\ d_W \end{smallmatrix}\right]$ d_W-dimensional subspaces X of U. We fix one such space X and count the number of spaces W of the form above with the additional condition that the image of W under the natural map $\varphi_U : U \oplus V \to U$ is X.

Let $\{x_1, \ldots, x_{d_W}\}$ be a basis for X. Let $w_1, \ldots, w_{d_W} \in U \oplus V$ be such that $\varphi_U(w_i) = x_i$. Then the vectors $w_1, \ldots, w_{d_W}$ form a basis for some space W of the required form.

Define $v_1, \ldots, v_{d_W} \in V$ by $v_i = w_i - \varphi_U(w_i) = w_i - x_i$. We have $\varphi_V(W) = \mathrm{Span}\{v_1, \ldots, v_{d_W}\}$, so the vectors $v_1, \ldots, v_{d_W}$ form a basis of a d_W-dimensional subspace of V. Moreover, different choices for the vectors $v_1, \ldots, v_{d_W}$ will lead to different subspaces W. The number of possible choices for the vectors

$v_1, \ldots, v_{d_W}$ is $\prod_{i=0}^{d_W-1}(q^{d_V} - q^i)$, since they must span a space of dimension d_W. Therefore, there are $\prod_{i=0}^{d_W-1}(q^{d_V} - q^i)$ spaces W of the required form with $\varphi_U(W) = X$, multiplying this by the number of choices for X gives the total number of spaces W to be as claimed. □

7.2.2 The special case with a direct sum

Lemma 7.2.2. *Let U, V be subspaces of $\mathbb{F}_q^m$ of dimensions d_U and d_V respectively. The number of d_W-dimensional spaces W such that $W \subseteq U \oplus V$ with $\dim(W \cap U) = d_{UW}$ and $\dim(W \cap V) = d_{VW}$ is*

$$g(d_U, d_V, d_W, d_{UW}, d_{VW}) \\ = \begin{bmatrix} d_U \\ d_{UW} \end{bmatrix} \begin{bmatrix} d_V \\ d_{VW} \end{bmatrix} f(d_U - d_{UW}, d_V - d_{VW}, d_W - d_{UW} - d_{VW})$$

when

$$d_{UW} \leq \min\{d_U, d_W\}, \quad d_{VW} \leq \min\{d_V, d_W\} \tag{7.2.1}$$

and

$$d_W - d_{UW} - d_{VW} \leq \min\{du - d_{UW}, d_V - d_{VW}\}, \tag{7.2.2}$$

and $g(d_U, d_V, d_W, d_{UW}, d_{VW}) = 0$ otherwise.

Proof. The inequalities in (7.2.1) hold since the dimension of the intersection of two spaces is bounded above by the dimension of those spaces.

For a given space W of the required form, consider the space $U \oplus V$ quotiented out by $(U \cap W) \oplus (V \cap W)$. For a subspace X of $U \oplus V$, let X' denote the image of this space in the quotient. Then W' is a subspace of $U' \oplus V'$ that intersects both U' and V' trivially. We have $\dim(U') = d_U - d_{UW}$, $\dim(V') = d_V - d_{VW}$ and $\dim(W') = d_W - d_{UW} - d_{VW}$. Therefore the inequality in (7.2.2) follows from Lemma 7.2.1.

To count the number of possible spaces W when (7.2.1) and (7.2.2) hold, first note that there are $\left[{d_U \atop d_{UW}}\right]$ choices for $U \cap W$ and $\left[{d_V \atop d_{VW}}\right]$ choices for $V \cap W$. Once these spaces are fixed, there are $f(d_U - d_{UW}, d_V - d_{VW}, d_W - d_{UW} - d_{VW})$ possible choices for W'.

Finally, since W' is a quotient of W by $(U \cap W) \oplus (V \cap W) \subseteq W$, it follows that W is uniquely determined by W', $U \cap W$ and $V \cap W$, hence the lemma holds. □

7.2.3 The special case with a sum

Lemma 7.2.3. *Let U, V be subspaces of $\mathbb{F}_q^m$ of dimensions d_U and d_V respectively such that* $\dim(U \cap V) = d_{UV}$. *The number of d_W-dimensional spaces W such that $W \subseteq U + V$ with* $\dim(W \cap U) = d_{UW}$, $\dim(W \cap V) = d_{VW}$ *and* $\dim(U \cap V \cap W) = d_{UVW}$ *is*

$$\begin{aligned}&h(d_U, d_V, d_W, d_{UV}, d_{UW}, d_{VW}, d_{UVW})\\&\quad= g(d_U - d_{UV}, d_V - d_{UV}, d_W - d_{UVW}, d_{UW} - d_{UVW}, d_{VW} - d_{UVW})\\&\qquad\qquad\cdot \begin{bmatrix} d_{UV} \\ d_{UVW} \end{bmatrix} q^{(d_W - d_{UVW})(d_{UV} - d_{UVW})}.\end{aligned}$$

Proof. Let $X = U \cap V$. Consider the quotient $\mathbb{F}_q^m / X$, and let S' denote the image of a space S in the quotient. Then $(U + V)/X = U' \oplus V'$ where $\dim(U') = d_U - d_{UV}$, $\dim(V') = d_V - d_{UV}$ and given any space W of the required form, $(W + X)/X = W'$ where $\dim(W') = d_W - d_{UVW}$. Now

$$\begin{aligned}\dim(U' \cap W') &= \dim(((U + X)/X) \cap ((W + X)/W))\\&= \dim(((U \cap W) + X)/X) = d_{UW} - d_{UVW}\end{aligned}$$

and similarly $\dim(V' \cap W') = d_{VW} - d_{UVW}$.

Hence the number of possibilities for the image W' of W in the quotient by X is $g(d_U - d_{UV}, d_V - d_{UV}, d_W - d_{UVW}, d_{UW} - d_{UVW}, d_{VW} - d_{UVW})$.

There are $\left[{d_{UV} \atop d_{UVW}}\right]$ possibilities for the space $U \cap V \cap W = X \cap W$.

Once the spaces W' and $X \cap W$ are fixed, by Lemma 3.2.7, there are $q^{(d_W - d_{UVW})(d_{UV} - d_{UVW})}$ possibilities for W. Multiplying this by the number of choices for W' and $X \cap W$ gives the statement of the lemma. □

7.2.4 The general case

Lemma 7.2.4. *Let U, V be subspaces of $\mathbb{F}_q^m$ of dimensions d_U and d_V respectively such that* $\dim(U \cap V) = d_{UV}$. *The number of d_W-dimensional subspaces $W \subseteq \mathbb{F}_q^m$ with* $\dim(W \cap U) = d_{UW}$, $\dim(W \cap V) = d_{VW}$ *and* $\dim(U \cap V \cap W) = d_{UVW}$ *is*

$$\begin{aligned} &l(m, d_U, d_V, d_W, d_{UV}, d_{UW}, d_{VW}, d_{UVW}) \\ &\quad = \sum_{k=\max\{d_{UW}, d_{VW}\}}^{\min\{d_{UW}+d_{VW}-d_{UVW}, d_W\}} h(d_U, d_V, k, d_{UV}, d_{UW}, d_{VW}, d_{UVW}) \\ &\qquad\qquad \cdot \begin{bmatrix} m - (d_U + d_V - d_{UV}) \\ d_W - k \end{bmatrix} q^{(d_W - k)(d_U + d_V - d_{UV} - k)}. \end{aligned}$$

Proof. For a space W of the required form, let $W' = (U + V) \cap W$. Let $k = \dim(W')$, then $k \geq d_{UW}$ since $\dim(W' \cap U) = \dim(W \cap U) = d_{UW}$. Similarly $k \geq d_{VW}$, hence $k \geq \max\{d_{UW}, d_{VW}\}$.

Clearly $k \leq \dim(W) = d_W$. Let $\varphi : U + V \to (U + V)/V \oplus (V + U)/U$ be the natural map. Note that $\ker(\varphi) = U \cap V$. Now $\varphi(W') \subseteq ((U \cap W) + V)/V \oplus ((V \cap W) + U)/U$, and so $\dim(\varphi(W')) \leq (d_{UW} - d_{UVW}) + (d_{VW} - d_{UVW})$. Moreover $\dim(W' \cap \ker(\varphi)) = \dim(W' \cap U \cap V) = d_{UVW}$. Hence $k = \dim(W') \leq \dim(\varphi(W')) + \dim(W' \cap \ker(\varphi)) \leq d_{UW} + d_{VW} - d_{UVW}$.

For k such that $\max\{d_{UW}, d_{VW}\} \leq k \leq \min\{d_{UW} + d_{VW} - d_{UVW}, d_W\}$ we will count spaces W with $\dim(W') = \dim((U + V) \cap W) = k$.

Since $W' \subseteq U + V$ and $\dim(U \cap W') = d_{UW}$, $\dim(V \cap W') = d_{VW}$ and $\dim(U \cap V \cap W') = d_{UVW}$, by Lemma 7.2.3 the number of choices for W' is

$h(d_U, d_V, k, d_{UV}, d_{UW}, d_{VW}, d_{UVW})$.

Consider the quotient space $\mathbb{F}_q^m/(U+V)$, this has dimension $m-(d_U+d_V-d_{UV})$. The image of W in the quotient has dimension d_W-k, hence there are $\left[{m-(d_U+d_V-d_{UV}) \atop d_W-k}\right]$ choices for the image.

By Lemma 3.2.7, once this image and W' are fixed, there are $q^{(d_W-k)(d_U+d_V-d_{UV}-k)}$ choices for W. Summing over k, the lemma follows. □

7.3 Counting pairs of subspaces

Let $U, V, W \subseteq \mathbb{F}_q^m$ such that $W+V=U+V$. The aim of this section is to count the number of pairs (V', W') such that $V' \subseteq V$, $W' \subseteq W$, $U+V'=W'+V'$ and the dimensions of V', W' and $V' \cap W'$ are fixed. This will be used in the proof of a later result.

We will show that this count depends only on the dimensions of V, U, W their pairwise intersections and $U \cap V \cap W$, thus we define a function

$$c'(d_U, d_V, d_W, d_{UV}, d_{UW}, d_{VW}, d_{UVW}; d_{V'}, d_{W'}, d_{V'W'})$$

which, given any spaces U, V, W of dimension d_U, d_V, d_W respectively with $U+V=W+V$ and $\dim(U\cap V)=d_{UV}$, $\dim(U\cap W)=d_{UW}$, $\dim(V\cap U)=d_{VW}$, $\dim(U\cap V\cap W)=d_{UVW}$, outputs the number of pairs (V', W') where $\dim(V')=d_{V'}$, $\dim(W')=d_{W'}$ and $\dim(V'\cap W')=d_{V'W'}$. The remainder of this section works to determine the value of the function c'. In Subsection 7.3.1 we discuss some basic properties the dimensions of subspaces and their intersections must satisfy and define a function $c(\underline{d}_1, \underline{d}_2)$ which calculates the number of pairs (V', W') when all the various dimensions of intersection with U, V and W are fixed and given in $\underline{d}_1$ and $\underline{d}_2$. Subsection 7.3.2 calculates the value of $c(\underline{d}_1, \underline{d}_2)$ in the special case when $V'=\{0\}$, then Subsection 7.3.3

calculates $c(\underline{d}_1, \underline{d}_2)$ for the special case when V' and W' intersect trivially. Subsection 7.3.4 calculates $c(\underline{d}_1, \underline{d}_2)$ in the general case and finally Subsection 7.3.5 uses the function $c(\underline{d}_1, \underline{d}_2)$ to calculate the value of c'.

7.3.1 Basic dimension properties

Throughout this subsection and the remainder of this section there will be many dimensions to consider. Therefore, for simplicity it will be understood that the notation d_V denotes the dimension of a space V , d_{UV} denotes $\dim(U \cap V)$ and d_{UVW} denotes $\dim(U \cap V \cap W)$.

Let $U, V, W \subseteq \mathbb{F}_q^m$ such that $W + V = U + V$. The aim of this subsection is to define a function that calculates the number of pairs (V', W') when all the various dimensions of intersection with U, V and W are fixed. We begin by discussing some basic properties the dimensions of these subspaces and their intersections must satisfy.

Let

$$\begin{aligned}&\underline{d}_1 = (d_U, d_V, d_W, d_{UV}, d_{UW}, d_{VW}, d_{UVW}) \text{ and}\\ &\underline{d}_2 = (d_{V'}, d_{W'}, d_{UV'}, d_{UW'}, d_{VW'}, d_{WV'}, d_{V'W'}, d_{UVW'}, d_{UWV'}, d_{UV'W'})\end{aligned} \tag{7.3.1}$$

be vectors whose entries are positive integers corresponding to the dimensions of the spaces U, V, W, V', W' as shown. Note that in this section we think of the subspaces U, V and W as fixed, and the subspaces V' and W' as varying, so that $\underline{d}_1$ is fixed, and the vector $\underline{d}_2$ varies according to the choices of V' and W'.

Given any vector $\underline{d}_1$ of this form, since its entries correspond to dimensions of subspaces and the dimension of a subspace is bounded above by the dimension of a space, these entries must satisfy the following *basic dimension properties (1)* (BDP1):

- $d_U, d_V, d_W \leq m$
- $d_{UV} \leq d_U, d_V, \quad d_{UW} \leq d_U, d_W, \quad d_{VW} \leq d_V, d_W$
- $d_{UVW} \leq d_{UV}, d_{UW}, d_{VW}$
- $d_U + d_V - d_{UV} \leq m, \quad d_U + d_W - d_{UW} \leq m, \quad d_V + d_W - d_{VW} \leq m$
- $d_{UV} + d_{VW} - d_{UVW} \leq m, \quad d_{UW} + d_{VW} - d_{UVW} \leq m,$
  $d_{UV} + d_{UW} - d_{UVW} \leq m$
- $d_U + d_V + d_W - d_{UV} - d_{UW} - d_{VW} + d_{UVW} \leq m$

Given a vector $\underline{d}_1$ satisfying BDP1, let V, U, W be any fixed spaces whose dimensions correspond to those in $\underline{d}_1$ and such that $W + V = U + V$. Then for any vector $\underline{d}_2$, since the entries in $\underline{d}_2$ are dimensions related to those in $\underline{d}_1$ and the dimension of a subspace is bounded above by the dimension of a space, these entries must satisfy the following *basic dimension properties (2)* (BDP2):

- $d_{V'} \leq d_V, \quad d_{W'} \leq d_W$
- $d_{UV'} \leq d_{UV}, d_{V'}, \quad d_{UW'} \leq d_{UW}, d_{W'}$
- $d_{VW'} \leq d_{VW}, d_{W'} \quad d_{WV'} \leq d_{VW}, d_{V'}$
- $d_{V'W'} \leq d_{VW'}, d_{WV'}$
- $d_{UVW'} \leq d_{UVW}, d_{UW'}, d_{VW'}, \quad d_{UWV'} \leq d_{UVW}, d_{UV'}, d_{WV'}$
- $d_{UV'W'} \leq d_{UVW'}, d_{UWV'}, d_{V'W'}$
- $d_U + d_{V'} - d_{UV'} \leq d_U + d_V - d_{UV}, \quad d_U + d_{W'} - d_{UW'} \leq d_U + d_W - d_{UW}$
- $d_{V'} + d_{W'} - d_{V'W'} \leq d_V + d_{W'} - d_{VW'}, d_W + d_{V'} - d_{WV'} \leq d_V + d_W - d_{VW}$

- $d_{UV'} + d_{V'W'} - d_{UV'W'} \leq d_{UV} + d_{VW'} - d_{UVW'}, d_{UV'} + d_{V'W} - d_{UV'W}$
  $\leq d_{UV} + d_{VW} - d_{UVW}$

- $d_{UW'} + d_{V'W'} - d_{UV'W'} \leq d_{UW'} + d_{VW'} - d_{UVW'}, d_{UW} + d_{V'W} - d_{UV'W}$
  $\leq d_{UW} + d_{VW} - d_{UVW}$

- $d_{UV'} + d_{UW'} - d_{UV'W'} \leq d_{UV} + d_{UW'} - d_{UVW'}, d_{UV'} + d_{UW} - d_{UV'W}$
  $\leq d_{UV} + d_{UW} - d_{UVW}$

- $d_U + d_{V'} + d_{W'} - d_{UV'} - d_{UW'} - d_{V'W'} + d_{UV'W'}$
  $\leq d_U + d_V + d_{W'} - d_{UV} - d_{UW'} - d_{VW'} + d_{UVW'},\ d_U + d_{V'} + d_W - d_{UV'} -$
  $d_{UW} - d_{V'W} + d_{UV'W}$
  $\leq d_U + d_V + d_W - d_{UV} - d_{UW} - d_{VW} + d_{UVW}$

Given a pair of vectors $(\underline{d}_1, \underline{d}_2)$ that satisfy BDP1 and BDP2, we say the pair $(\underline{d}_1, \underline{d}_2)$ satisfies the *basic dimension properties* (BDP).

Given a pair of vectors $(\underline{d}_1, \underline{d}_2)$ satisfying BDP, let V, U, W be any fixed spaces whose dimensions correspond to those in $\underline{d}_1$ and such that $W + V = U + V$. Define $c(\underline{d}_1, \underline{d}_2)$ to be the number of pairs (V', W') whose dimensions correspond to the entries in $\underline{d}_2$, such that $V' \subseteq V$, $W' \subseteq W$ and $W' + V' = U + V'$. For a pair $(\underline{d}_1, \underline{d}_2)$ that doesn't satisfy BDP, define $c(\underline{d}_1, \underline{d}_2) = 0$.

The following three subsections show that the function c is well defined, and depends on the values of $\underline{d}_1$ and $\underline{d}_2$. We begin by establishing the formula in the special case when $d_{V'} = 0$ in Subsection 7.3.2, then for the special case $d_{V'W'} = 0$ in Subsection 7.3.3 and finally, in Subsection 7.3.4, we present the general formula.

7.3.2 The special case with a trivial space

Lemma 7.3.1. *Let $\underline{d}_1$ and $\underline{d}_2$ be vectors of the form given in* (7.3.1)*, such that* $d_{V'} = 0$*. Then*

$$c(\underline{d}_1, \underline{d}_2) = 1$$

if $\underline{d}_1$ satisfies BDP1 with $d_U = d_{UW}$, $d_{UV} = d_{UVW}$ and $\underline{d}_2 = (0, d_U, 0, d_U, d_{VU}, 0, 0, d_{VU}, 0, 0)$; and $c(\underline{d}_1, \underline{d}_2) = 0$ otherwise.

Proof. If $\underline{d}_1$ does not satisfy BDP1 then $c(\underline{d}_1, \underline{d}_2) = 0$ by definition.

Let $\underline{d}_1$ be a vector satisfying BDP1 and let V, U, W be fixed spaces whose dimensions correspond to the values in $\underline{d}_1$, such that $V + W = V + U$.

Since $c(\underline{d}_1, \underline{d}_2)$ gives the number of pairs (V', W') with $V' \subseteq V$, $W' \subseteq W$ and $W' + V' = U + V'$, if $d_{V'} = 0$ then $V' = \{0\}$ and $c(\underline{d}_1, \underline{d}_2)$ is simply the number of spaces $W' \subseteq W$ such that $W' = U$. Hence there is precisely one choice for W', namely $W' = U$. This is possible if and only if $U \subseteq W$, which is the case if and only if $d_U = d_{UW}$. Note that $U \subseteq W$ also forces $d_{UV} = d_{UVW}$. Once we set $V' = \{0\}$ and $W' = U$ all the entries in $\underline{d}_2$ are fixed as in the statement of the lemma (note that $\underline{d}_2$ satisfies BDP2, hence $(\underline{d}_1, \underline{d}_2)$ satisfy BDP). The result follows. □

7.3.3 The special case with trivial intersection

Lemma 7.3.2. *Let $\underline{d}_1$ and $\underline{d}_2$ be vectors of the form given in* (7.3.1)*, such that* $d_{V'W'} = 0$*. Then*

$$c(\underline{d}_1, \underline{d}_2) = q^{d_{W'} d_{V'}} l(d_V, d_{UV}, d_{WV}, d_{V'}, d_{UWV}, d_{UV'}, d_{WV'}, d_{UWV'})$$

(where l is as defined in Lemma 7.2.4) if $(\underline{d}_1, \underline{d}_2)$ satisfies BDP with

$$d_U - d_{UV'} = d_{UW} - d_{UWV'} = d_{W'} = d_{UW'} \text{ and}$$
$$d_{UV} - d_{UV'} = d_{UVW} - d_{UWV'} = d_{VW'} = d_{UVW'}; \qquad (7.3.2)$$

and $c(\underline{d}_1, \underline{d}_2) = 0$ *otherwise.*

Proof. If $(\underline{d}_1, \underline{d}_2)$ do not satisfy BDP then $c(\underline{d}_1, \underline{d}_2) = 0$ by definition.

Let $(\underline{d}_1, \underline{d}_2)$ be a pair of vectors satisfying BDP with $d_{V'W'} = 0$, and let V, U, W be fixed spaces whose dimensions correspond to the values in $\underline{d}_1$, such that $V + W = V + U$. Since $d_{V'W'} = 0$, we wish to count pairs (V', W') with trivial intersection.

First we pick V' to be a $d_{V'}$-dimensional subspace of V with $\dim(U \cap V') = d_{UV'}$, $\dim(W \cap V') = d_{WV'}$ and $\dim(U \cap W \cap V') = d_{UWV'}$.

Note that, since $V' \subseteq V$ we have $U \cap V' = (U \cap V) \cap V'$, $W \cap V' = (W \cap V) \cap V'$ and $U \cap W \cap V' = (U \cap V) \cap (W \cap V) \cap V'$. Therefore, it is equivalent to count $d_{V'}$-dimensional subspaces of V with $\dim((U \cap V) \cap V') = d_{UV'}$, $\dim((W \cap V) \cap V') = d_{WV'}$ and $\dim((U \cap V) \cap (W \cap V) \cap V') = d_{UWV'}$, where $U \cap V$ and $W \cap V$ are subspaces of V and $\dim((U \cap V) \cap (W \cap V)) = d_{UVW}$. This value is given in Lemma 7.2.4 to be

$$l(d_V, d_{UV}, d_{WV}, d_{V'}, d_{UWV}, d_{UV'}, d_{WV'}, d_{UWV'}).$$

Fix one such subspace V'. We must now count $d_{W'}$-dimensional spaces $W' \subseteq W$ such that $U + V' = W' + V'$.

Consider the quotient space $\mathbb{F}_q^m / V'$. For a subspace $A \subseteq \mathbb{F}_q^m$, let $[A]$ denote the image of A in the quotient $\mathbb{F}_q^m / V'$, then $d_{[A]} = d_A - d_{AV'}$.

Let

$$[\underline{d}_1] = (d_{[U]}, d_{[V]}, d_{[W]}, d_{[UV]}, d_{[UW]}, d_{[VW]}, d_{[UVW]})$$

and let

$$[\underline{d}_2] = (d_{[V']}, d_{[W']}, d_{[UV']}, d_{[UW']}, d_{[VW']}, d_{[WV']}, d_{[V'W']}, d_{[UVW']}, d_{[UWV']}, d_{[UV'W']})$$

It follows that

$$\begin{aligned}[\underline{d}_1] = (d_U - d_{UV'}, d_V - d_{V'}, d_W - d_{WV'}, \\ d_{UV} - d_{UV'}, d_{UW} - d_{UWV'}, d_{VW} - d_{WV'}, d_{UVW} - d_{UWV'})\end{aligned}$$

and

$$[\underline{d}_2] = (0, d_{W'}, 0, d_{UW'}, d_{VW'}, 0, 0, d_{UVW'}, 0, 0).$$

Thus in the quotient by V', the dimensions in $\underline{d}_1$ and $\underline{d}_2$ drop by either $d_{V'}, d_{UV'}, d_{WV'}$ or $d_{UWV'}$. Therefore, we are able to quotient in this way with any choice of V' with these fixed values. Note that, since $(\underline{d}_1, \underline{d}_2)$ satisfies BDP, it follows that $([\underline{d}_1], [\underline{d}_2])$ also satisfies BDP.

Now, by Lemma 7.3.1, $c([\underline{d}_1], [\underline{d}_2]) = 1$ if $d_{[U]} = d_{[UW]} = d_{[W']} = d_{[UW']}$ and $d_{[UV]} = d_{[UVW]} = d_{[VW']} = d_{[UVW']}$; and $c([\underline{d}_1], [\underline{d}_2]) = 0$ otherwise. Hence there is precisely one possible choice for the space $[W'] \subseteq \mathbb{F}_q^m/V'$ and it is necessary that $d_{[U]} = d_{[UW]} = d_{[W']} = d_{[UW']}$ and $d_{[UV]} = d_{[UVW]} = d_{[VW']} = d_{[UVW']}$, that is

$$\begin{aligned} d_U - d_{UV'} = d_{UW} - d_{UWV'} = d_{W'} = d_{UW'} \text{ and} \\ d_{UV} - d_{UV'} = d_{UVW} - d_{UWV'} = d_{VW'} = d_{UVW'}. \end{aligned} \tag{7.3.3}$$

By Lemma 3.2.7 there are precisely $q^{(d_{W'}-0)(d_{V'}-0)} = q^{d_{W'}d_{V'}}$ spaces W' in $\mathbb{F}_q^m$, with $W' \cap V'$ trivial, whose image in $\mathbb{F}_q^m/V'$ is $[W']$. Thus for the fixed choice of V', there are $q^{d_{W'}d_{V'}}$ possible choices for W' if (7.3.3) holds, and zero possible choices otherwise. Since this only depends on the fixed dimensions of V', multiplying the number of possible choices for W' by the number of possible choices for V' gives the result. □

7.3.4 The general case

Lemma 7.3.3. *The function $c(\underline{d}_1, \underline{d}_2)$ is well defined and is given by*

$$c(\underline{d}_1, \underline{d}_2) = q^{(d_{V'W'}-d_{UV'W'})(d_{UVW}-d_{UV'W'})} \cdot \begin{bmatrix} d_{VW} - d_{UVW} \\ d_{V'W'} - d_{UV'W'} \end{bmatrix} \begin{bmatrix} d_{UVW} \\ d_{UV'W'} \end{bmatrix} c([\underline{d}_1], [\underline{d}_2]), \quad (7.3.4)$$

where

$$[\underline{d}_1] = (d_U - d_{UV'W'}, d_V - d_{V'W'}, d_W - d_{V'W'}, d_{UV} - d_{UV'W'}, d_{UW} - d_{UV'W'}, d_{VW} - d_{V'W'}, d_{UVW} - d_{UV'W'})$$

and

$$[\underline{d}_2] = (d_{V'} - d_{V'W'}, d_{W'} - d_{V'W'}, d_{UV'} - d_{UV'W'}, d_{UW'} - d_{UV'W'}, d_{VW'} - d_{V'W'}, d_{WV'} - d_{V'W'}, 0, d_{UVW'} - d_{UV'W'}, d_{UWV'} - d_{UV'W'}, 0)$$

and the value $c([\underline{d}_1], [\underline{d}_2])$ is given in Lemma 7.3.2.

Proof. If $(\underline{d}_1, \underline{d}_2)$ do not satisfy BDP then $c(\underline{d}_1, \underline{d}_2) = 0$ by definition.

Let $(\underline{d}_1, \underline{d}_2)$ be a pair of vectors satisfying BDP, and let V, U, W be fixed spaces whose dimensions correspond to the values in $\underline{d}_1$, such that $V + W = V + U$.

We begin by calculating the number of pairs (V', W') when $V' \cap W'$ is some fixed $d_{V'W'}$-dimensional space. We will show that this count does not depend on the specific choice of $V' \cap W'$, but only on the dimensions $d_{V'W'}$ and $d_{UV'W'}$, thus in order to calculate $c(\underline{d}_1, \underline{d}_2)$ we can multiply this count by the number of possibilities for the space $V' \cap W'$, with fixed values of $d_{V'W'}$ and $d_{UV'W'}$.

Let S_1 be a fixed $d_{V'W'}$ dimensional subspace of $V \cap W$, with $\dim(U \cap S_1) = d_{UV'W'}$. We will count spaces W', V' such that $V' \cap W' = S_1$. Consider the quotient space $\mathbb{F}_q^m/S_1$, this is a space of dimension $m - d_{V'W'}$. For a subspace $A \subseteq \mathbb{F}_q^m$, let $[A]$ denote the image of A in the quotient $\mathbb{F}_q^m/S_1$, then $d_{[A]} = d_A - d_{AS_1}$. Since $S_1 = V' \cap W'$, the various dimensions in $\underline{d}_1$ and $\underline{d}_2$ will drop in the quotient by either $d_{V'W'}$ or $d_{UV'W'}$, giving

$$(d_{[U]}, d_{[V]}, d_{[W]}, d_{[UV]}, d_{[UW]}, d_{[VW]}, d_{[UVW]}) = [\underline{d}_1]$$

and

$$(d_{[V']}, d_{[W']}, d_{[UV']}, d_{[UW']}, d_{[VW']}, d_{[WV']}, d_{[V'W']}, d_{[UVW']}, d_{[UWV']}, d_{[UV'W']}) = [\underline{d}_2]$$

where $[\underline{d}_1]$ and $[\underline{d}_2]$ are as in the statement of the lemma.

The value of $c([\underline{d}_1], [\underline{d}_2])$ is given in Lemma 7.3.2; this gives the number of pairs $([V'], [W'])$ in the quotient space $\mathbb{F}_q^m/S_1$. There is a direct correspondence between pairs (V', W') containing S_1 and their image in the quotient space $\mathbb{F}_q^m/S_1$. Therefore the number of pairs (V', W') with $V' \cap W' = S_1$ is equal to $c([\underline{d}_1], [\underline{d}_2])$.

Note that, $c([\underline{d}_1], [\underline{d}_2])$ does not depend on the specific space S_1, but only on $\dim(S_1) = d_{V'W'}$ and $\dim(S_1 \cap U) = d_{UV'W'}$. Therefore if we multiply $c([\underline{d}_1], [\underline{d}_2])$ by the number of possibilities for $S_1 \subseteq V \cap W$ with $\dim(S_1) = d_{V'W'}$ and $\dim(S_1 \cap U) = d_{UV'W'}$ we will obtain the number of pairs (V', W') of the required form with $\dim(V' \cap W') = d_{V'W'}$ and $\dim(U \cap V' \cap W') = d_{UV'W'}$, that is, we will obtain the value $c(\underline{d}_1, \underline{d}_2)$.

In order to count the number of possible choices for S_1 we begin by fixing a $d_{UV'W'}$-dimensional subspace $S_2 \subseteq U \cap V \cap W$ and will count spaces S_1 with $S_1 \cap U = S_2$. There are

$$\begin{bmatrix} d_{UVW} \\ d_{UV'W'} \end{bmatrix} \tag{7.3.5}$$

possible choices for S_2. Fix one. Now, by Corollary 3.2.8 there are

$$q^{(d_{V'W'} - d_{UV'W'})(d_{UVW} - d_{UV'W'})} \begin{bmatrix} d_{VW} - d_{UVW} \\ d_{V'W'} - d_{UV'W'} \end{bmatrix} \tag{7.3.6}$$

spaces S_1 with $S_1 \cap U = S_2$. Since this doesn't depend on the specific space S_2, only its dimension, multiplying together (7.3.5) and (7.3.6) gives the total number of possible choices for the space S_1. Multiplying the number of choices for S_1 by $c([\underline{d}_1], [\underline{d}_2])$ gives the result. □

7.3.5 The final counting argument

Recall that the aim of this section was to count the number of pairs (V', W') such that $V' \subseteq V$, $W' \subseteq W$, $U + V' = W' + V'$ and the values $d_{V'}, d_{W'}, d_{V'W'}$ are fixed. The following lemma gives this result.

Lemma 7.3.4. *Let $V, U, W \subseteq \mathbb{F}_q^m$ be some fixed spaces, such that $V + W = V + U$. The number of pairs (V', W') such that $V' \subseteq V$, $W' \subseteq W$ and $V' + W' = V' + U$ with $\dim V' = d_{V'}$, $\dim W' = d_{W'}$ and $\dim V' \cap W' = d_{V'W'}$ depends only on the dimensions $d_U, d_V, d_W, d_{UV}, d_{UW}, d_{VW}, d_{UVW}, d_{V'}, d_{W'}, d_{V'W'}$ and is given by*

$$c'(d_U, d_V, d_W, d_{UV}, d_{UW}, d_{VW}, d_{UVW}; d_{V'}, d_{W'}, d_{V'W'}) = \sum_{i,j,k,p,q,s,t=0}^{m} c(\underline{d}_1, \underline{d}_2) \tag{7.3.7}$$

where $\underline{d}_1 = (d_U, d_V, d_W, d_{UV}, d_{UW}, d_{VW}, d_{UVW})$ and $\underline{d}_2 = (d_{V'}, d_{W'}, i, j, k, p, d_{V'W'}, q, s, t)$ and the value $c(\underline{d}_1, \underline{d}_2)$ is given in Lemma 7.3.3.

Proof. Since the entries of $\underline{d}_1$ correspond to the fixed spaces U, V and W, BDP1 is satisfied. Then given some fixed i, j, k, p, q, s, t, the value $c(\underline{d}_1, \underline{d}_2)$ gives the number of pairs (V', W') of the required form with $d_{UV'} = i$, $d_{UW'} = j$, $d_{VW'} = k$, $d_{WV'} = p$, $d_{UVW'} = q$, $d_{UWV'} = s$ and $d_{UV'W'} = t$. Note that some choices of i, j, k, p, q, s, t will result in no such spaces (V', W'), in this case BDP2 will not be satisfied, and the value $c(\underline{d}_1, \underline{d}_2)$ is zero by definition. Therefore, summing over all possible values of i, j, k, p, q, s, t gives the result. □

7.4 Calculating the matrix functions

We have now established the necessary preliminary results in order to calculate the values of the matrix functions f_0, f_1, f_2 discussed at the start of this chapter. This section calculates these functions. Subsection 7.4.1 calculates f_0, Subsection 7.4.2 calculates f_1 and finally Subsection 7.4.3 calculates f_2.

7.4.1 Calculating f_0

The following lemma gives the number of matrices with a fixed rowspace.

Lemma 7.4.1. *Let U be a subspace of $\mathbb{F}_q^m$ of dimension u. The number $f_0(u)$ of matrices $M \in \mathbb{F}_q^{n\times m}$ such that $\mathrm{Row}(M) = U$ can be efficiently computed; it depends only on q, n, m and u. For $0 \le u \le \min\{n, m\}$,*

$$f_0(u) = \sum_{v=0}^{u} (-1)^{u-v} q^{nv+\binom{u-v}{2}} \begin{bmatrix} u \\ v \end{bmatrix}. \tag{7.4.1}$$

Proof. Let

$$f(V) = |\{M : M \in \mathbb{F}_q^{n\times m}, \mathrm{Row}(M) = V\}|,$$

and let

$$g(V) = |\{M : M \in \mathbb{F}_q^{n\times m}, \mathrm{Row}(M) \subseteq V\}|.$$

Then

$$g(V) = \sum_{U\subseteq V} f(U).$$

Therefore by Lemma 3.3.3

$$f(U) = \sum_{V\subseteq U} (-1)^{\dim(U)-\dim(V)} q^{\binom{\dim(U)-\dim(V)}{2}} g(V). \tag{7.4.2}$$

Now, $g(V)$ is the number of $n \times m$ matrices whose rowspace is contained in V. For any $n \times m$ matrix M, $\mathrm{Row}(M) \subseteq V$ if and only if each row of M is an element of V. Therefore there are $q^{\dim(V)}$ choices for each row, hence $q^{n\dim(V)}$ possible matrices. Thus

$$g(V) = q^{n\dim(V)}. \tag{7.4.3}$$

Substituting (7.4.3) into (7.4.2) gives

$$\begin{aligned}
f(U) &= \sum_{V\subseteq U} (-1)^{\dim(U)-\dim(V)} q^{\binom{\dim(U)-\dim(V)}{2}} q^{n\dim(V)} \\
&= \sum_{V\subseteq U} (-1)^{\dim(U)-\dim(V)} q^{n\dim(V)+\binom{\dim(U)-\dim(V)}{2}} \\
&= \sum_{v=0}^{\dim(U)} (-1)^{\dim(U)-v} q^{nv+\binom{\dim(U)-v}{2}} \begin{bmatrix} \dim(U) \\ v \end{bmatrix}, \qquad (7.4.4)
\end{aligned}$$

where (7.4.4) follows since there are $\left[\begin{smallmatrix}\dim(U)\\ v\end{smallmatrix}\right]$ subspaces $V \subseteq U$ of dimension v for $v = 0, \dots, \dim(U)$.

Note that $f(U)$ depends only on the dimension of the space U; therefore f_0 is well defined and (7.4.4) establishes (7.4.1). □

Remark. An expression for f_0 was already known to Gabidulin [14, Theorem 4], who showed that for $0 \le u \le \min\{n, m\}$

$$f_0(u) = \prod_{i=0}^{u-1} q^n - q^i. \tag{7.4.5}$$

This result is equivalent to Lemma 7.4.1, since (7.4.5) is equal to (7.4.1) by [14, Equation 13].

7.4.2 Calculating f_1

For a space $U \subseteq \mathbb{F}_q^m$, the following theorem gives the number of matrices B, of rank r, such that $\mathrm{Row}(M + B)$ is some fixed space $V \subseteq \mathbb{F}_q^m$, where M is any fixed matrix with $\mathrm{Row}(M) = U$.

Theorem 7.4.2. *Let U and V be subspaces of $\mathbb{F}_q^m$ of dimensions u and v respectively. Let $h = \dim((U + V)/V)$. Let $M \in \mathbb{F}_q^{n\times m}$ be a fixed matrix such* *that* $\mathrm{Row}(M) = U$*. Let* r *be a non-negative integer. The number* $f_1(u,v,h;r)$ *of matrices* $B \in \mathbb{F}_q^{n\times m,r}$ *such that* $\mathrm{Row}(M+B) = V$ *can be efficiently computed; it depends only on* q*,* n*,* m*,* r*,* u*,* v *and* h*. For* $u-h \leq v \leq \min\{n,m\}$ *and* $h \leq r \leq v+h$*,*

$$f_1(u,v,h;r) = \sum_{d_{UW}=0}^{\min\{u,r\}} \sum_{d_{UWV}=0}^{\min\{v-h,r-h,d_{UW}\}} l(u,v,r,v-h,d_{UW},r-h,d_{UVW})$$
$$\cdot \sum_{d_{W'}=0}^{r} \sum_{d_{V'}=0}^{v} \sum_{d_{V'W'}=0}^{\min\{d_{W'},d_{V'}\}} c'((u,v,r,v-h,d_{UW},r-h,d_{UVW};d_{V'},d_{W'},d_{V'W'}))$$
$$\cdot (-1)^{r-d_{W'}+v-d_{V'}} q^{\binom{r-d_{W'}}{2}+\binom{v-d_{V'}}{2}} q^{nd_{W'V'}}, \quad (7.4.6)$$

where the values of the functions l *and* c' *are given in Lemmas 7.2.4 and 7.3.4, respectively. When* $v < u-h$*,* $v > \min\{n,m\}$*,* $r < h$ *or* $r > v+h$*, we find that* $f_1(u,v,h;r) = 0$*.*

Proof. We begin by establishing the regions for which f_1 is zero.

Since $M, B \in \mathbb{F}_q^{n\times m}$, if $V = \mathrm{Row}(M+B)$ we must have $\dim(V) \leq \min\{n,m\}$; thus $f_1(v,u,h;r) = 0$ for $v > \min\{n,m\}$. Since $h = \dim((U+V)/V)$, by Lemma 3.2.6 it follows that $\dim(U\cap V) = u-h$. Clearly $\dim(V) \geq \dim(U\cap V)$, hence $f_1(u,v,h;r) = 0$ for $v < u-h$.

Consider the quotient space $\mathbb{F}_q^m/V$. For a subspace $S \subseteq \mathbb{F}_q^m$, let $[S]$ denote the image of S in $\mathbb{F}_q^m/V$. Then $[U] = (U+V)/V$, so $\dim([\mathrm{Row}(M)]) = \dim([U]) = h$.

Let B be an $n\times m$ matrix of rank r. For $i = 1,\ldots,n$ let m_i denote the i-th row of M and let b_i denote the i-th row of B. If $\mathrm{Row}(M+B) \subseteq V$ it follows that for each i, $m_i + b_i = v_i$ for some $v_i \in V$. Therefore

$$\begin{aligned} b_i + V &= -m_i + v_i + V \\ &= -m_i + V \qquad (7.4.7) \end{aligned}$$

for $i = 1, \dots, n$. Hence

$$\begin{aligned} [\mathrm{Row}(B)] &= \mathrm{Span}\{-m_1 + V, \dots, -m_n + V\} \\ &= [\mathrm{Row}(M)] \\ &= [U]. \end{aligned} \tag{7.4.8}$$

Therefore $[\mathrm{Row}(B)]$ is uniquely defined by M and V, and $\dim([\mathrm{Row}(B)]) = \dim([U]) = h$.

Note that $\mathrm{rk}(B) \geq \dim([\mathrm{Row}(B)]) = h$, hence $f_1(u, v, h; r) = 0$ for $r < h$. Also, $r - h = \dim(\mathrm{Row}(B) \cap V) \leq \dim(V)$, so $f_1(u, v, h; r) = 0$ for $r > v + h$.

We have established that f_1 is zero whenever $v < u - h$, $v > \min\{n, m\}$, $r < h$ or $r > v + h$.

We now go on the calculate f_1 for $u - h \leq v \leq \min\{n, m\}$ and $h \leq r \leq v + h$.

Let $S = [U]$, so that for any matrix B of the required form, we must have $[\mathrm{Row}(B)] = S$.

If $h \leq r \leq \dim(V) + h$, given any r-dimensional subspace W, of $\mathbb{F}_q^m$ with $[W] = S$, there exists a matrix B with $\mathrm{Row}(B) \subseteq W$ such that $\mathrm{Row}(M + B) \subseteq V$. To see this, since $[W] = S$, there exists vectors $v_1, \dots, v_n \in V$ such that $-m_i + v_i \in W$ for $i = 1, \dots, n$. Take the i-th row of the matrix B to be $-m_i + v_i$, then the matrix B has the required form. Note that, for any space $W \subseteq \mathbb{F}_q^m$, since $[W] = (W + V)/V$ and $S = (U + V)/V$, it follows that $[W] = S$ if and only if $W + V = U + V$.

For a fixed space W we count the number of matrices B, such that $\mathrm{Row}(B) \subseteq W$ and $\mathrm{Row}(M + B) \subseteq V$. We will then use Möbius inversion to count the number of matrices B, such that $\mathrm{Row}(B) = W$ and $\mathrm{Row}(M + B) = V$. Finally, we will sum over possible choices of W to get the total number of matrices B of the required form.

Let W be a fixed subspace of $\mathbb{F}_q^m$ of dimension r with $W + V = U + V$. Then we have shown that there exists a matrix B with $\mathrm{Row}(B) \subseteq W$ such that $\mathrm{Row}(M + B) \subseteq V$. Given one such matrix B and fixed vectors $v_1, \ldots, v_n \in V$, consider the matrix

$$B' = B + \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}. \tag{7.4.9}$$

Then since $v_i \in V$, $\mathrm{Row}(M + B') \subseteq V$. Note that adding anything to B not in V will result in a matrix B'' such that $\mathrm{Row}(M + B'') \not\subseteq V$, so we are only interested in adding rows in V. Now, if $v_i \in W \cap V$ for each i, then $\mathrm{Row}(B') \subseteq W$. However if for any i, $v_i \notin W$ then $\mathrm{Row}(B') \not\subseteq W$. Thus B' is an $n \times m$ matrix with $\mathrm{Row}(B') \subseteq W$ and $\mathrm{Row}(M + B') \subseteq V$ if and only if B' is of the form given in (7.4.9) with $v_1, \ldots, v_n \in W \cap V$. Therefore the number of matrices B of the required form is equal to the number of choices for $v_1, \ldots, v_n \in W \cap V$. There are $q^{\dim(W \cap V)}$ choices for each v_i, hence $q^{n \dim(W \cap V)}$ choices for the ordered set $\{v_1, \ldots, v_n\}$. By Lemma 3.2.6, $\dim(W \cap V) = r - h$, thus we have shown that the number of matrices B, with $\mathrm{Row}(B) \subseteq W$ and $\mathrm{Row}(M + B) \subseteq V$, is

$$q^{n \dim(W \cap V)} = q^{n(r-h)}.$$

Recall, $\mathrm{Po}(\mathbb{F}_q^m)$ is the poset of all subspaces of $\mathbb{F}_q^m$, ordered by containment. Consider the direct product $\mathrm{Po}(\mathbb{F}_q^m) \times \mathrm{Po}(\mathbb{F}_q^m)$. For $(W', V') \in \mathrm{Po}(\mathbb{F}_q^m) \times \mathrm{Po}(\mathbb{F}_q^m)$, let

$$f((W', V')) = |\{B \in \mathbb{F}_q^{n \times m} : \mathrm{Row}(B) = W', \mathrm{Row}(M + B) = V'\}|,$$

and let

$$g((W', V')) = |\{B \in \mathbb{F}_q^{n \times m} : \mathrm{Row}(B) \subseteq W', \mathrm{Row}(M + B) \subseteq V'\}|.$$

The above argument shows that

$$g((W',V')) = \begin{cases} q^{n\dim(W'\cap V')} & \text{if } W'+V'=U+V', \\ 0 & \text{otherwise.} \end{cases} \tag{7.4.10}$$

By the definition of f and g,

$$g((W,V)) = \sum_{(W',V')\leq(W,V)} f((W',V')).$$

Therefore, by Lemma 3.3.5,

$$\begin{aligned} &f((W,V)) \\ &= \sum_{(W',V')\leq(W,V)} (-1)^{r-\dim(W')+v-\dim(V')} q^{\binom{r-\dim(W')}{2}+\binom{v-\dim(V')}{2}} g((W',V')) \\ &= \sum_{\substack{(W',V')\leq(W,V)\\ W'+V'=U+V'}} (-1)^{r-\dim(W')+v-\dim(V')} q^{\binom{r-\dim(W')}{2}+\binom{v-\dim(V')}{2}} q^{n\dim(W'\cap V')}, \end{aligned} \tag{7.4.11}$$

where (7.4.11) follows from (7.4.10).

Suppose that for our fixed space W, $\dim(U\cap W) = d_{UW}$ and $\dim(U\cap V\cap W) = d_{UVW}$. Then by Lemma 7.3.4, the number of pairs (W',V'), with $(W',V')\leq(W,V)$ and $W'+V'=U+V'$, with $\dim(W')=d_{W'}$, $\dim V'=d_{V'}$ and $\dim(W'\cap V')=d_{W'V'}$ is

$$c'(\underline{d}) = c'((u,v,r,r,v-h,d_{UW},r-h,d_{UVW};d_{V'},d_{W'},d_{V'W'})), \tag{7.4.12}$$

where c' is as defined in Lemma 7.3.4.

Substituting (7.4.12) into (7.4.11) gives

$$\begin{aligned} f((W,V)) = \sum_{d_{W'}=0}^{r}\sum_{d_{V'}=0}^{v}\sum_{d_{V'W'}=0}^{\min\{d_{W'},d_{V'}\}} c'(\underline{d}) \\ \cdot (-1)^{r-d_{W'}+v-d_{V'}} q^{\binom{r-d_{W'}}{2}+\binom{v-d_{V'}}{2}} q^{nd_{W'V'}}. \end{aligned} \tag{7.4.13}$$

Thus we have calculated the number of matrices of the required form with a specific rowspace W.

Now, by Lemma 7.2.4 the number of spaces W with $\dim(W) = r$, $\dim(W \cap V) = r - h$, $\dim(U \cap W) = d_{UW}$ and $\dim(U \cap V \cap W) = d_{UVW}$ is $l(m, u, v, r, v - h, d_{UW}, r - h, d_{UVW})$, where l is as defined in Lemma 7.2.4.

Hence

$$\begin{aligned} f_1(u, v, h; r) &= \sum_{W} f((W, V)) \\ &= \sum_{d_{UW}=0}^{\min\{u,r\}} \sum_{d_{UWV}=0}^{\min\{v-h,r-h,d_{UW}\}} l(m, u, v, r, v - h, d_{UW}, r - h, d_{UVW}) \\ &\cdot \sum_{d_{W'}=0}^{r} \sum_{d_{V'}=0}^{v} \sum_{d_{V'W'}=0}^{\min\{d_{W'},d_{V'}\}} c'(\underline{d})(-1)^{r-d_{W'}+v-d_{V'}} q^{\binom{r-d_{W'}}{2}+\binom{v-d_{V'}}{2}} q^{nd_{W'V'}} \end{aligned}$$

as claimed. □

7.4.3 Calculating f_2

For a fixed matrix X of rank r_X, the function $f_2(r_X, r_B, r)$ gives the number of matrices B of rank r_B such that $\mathrm{rk}(X + B) = r$. This is equal to the number of matrices B' of rank r_B such that $\mathrm{rk}(X - B') = r$ (setting $B' = -B$). The *rank distance* is a metric defined for two matrices $M_1, M_2 \in \mathbb{F}_q^{n\times m}$ to be

$$d_R(M_1, M_2) = \mathrm{rk}(M_1 - M_2).$$

Therefore, the value $f_2(r_X, r_B, r)$ gives the number of matrices of rank r_B, that have rank distance r from some fixed matrix of rank r_X. Or equivalently, considering the space of all $n \times m$ matrices over $\mathbb{F}_q$, $f_2(r_X, r_B, r)$ is the volume of intersection of two spheres with rank radii r_X and r_W with centres at rank distance r. With this phrasing it becomes clear that the value of f_2 is an important property of rank metric codes, a class of codes based on the rank metric, which are of much interest due to their applications to data storage, public-key cryptosystems, space-time coding and random linear network coding (see e.g. [17], [38]). The analysis of the volume of intersection of spheres in the rank metric space can lead to the development of covering properties for rank metric codes, as explored by Gadouleau and Yan [18]. In [18, Lemma 1], the authors give an expression for the function f_2, showing that indeed it is efficiently computable. However, the expression given was developed using the theory of association schemes and the formula does not explain what it is counting. Our work was developed independently and the following theorem gives an expression for f_2 that is constructed using the counting arguments considered in this chapter, thus our new formula and proof give extra insight.

Theorem 7.4.3. *Let r, r_B and r_X be non-negative integers. Let X be a fixed matrix such that $\mathrm{rk}(X) = r_X$. The number $f_2(r, r_X, r_B)$ of matrices $B \in \mathbb{F}_q^{n\times m, r_B}$ such that $\mathrm{rk}(X + B) = r$ can be efficiently computed; it depends only on q, n, m, r, r_B and r_X. Indeed,*

$$f_2(r, r_X, r_B) = \sum_{h=0}^{\min\{r,r_X\}} q^{(r-h)(r_X-h)} \begin{bmatrix} m - r_X \\ r - h \end{bmatrix} \begin{bmatrix} r_X \\ h \end{bmatrix} f_1(r_X, r, h; r_B).$$

Proof. Using the definition of f_1 given in Theorem 7.4.2, we see that

$$\begin{aligned} & f_2(r, r_X, r_B) \\ &= \sum_{V\subseteq \mathbb{F}_q^m:\dim(V)=r} f_1(r_X, r, \dim(V \cap \mathrm{Row}(X)); r_B) && (7.4.14) \\ &= \sum_{h=0}^{\min\{r,r_X\}} |\{V \subseteq \mathbb{F}_q^m : \dim(V) = r, \dim(V \cap \mathrm{Row}(X)) = h\}| f_1(r_X, r, h; r_B) && (7.4.15) \end{aligned}$$

where (7.4.14) follows since the number of matrices B with $\mathrm{rk}(X + B) = r$ is equal to the number of matrices B with $\mathrm{Row}(X + B) = V$, summed over all spaces $V \subseteq \mathbb{F}_q^m$ with $\dim(V) = r$.

By Corollary 3.2.10, the number of r-dimensional subspaces $V \subseteq \mathbb{F}_q^m$, with $\dim(V \cap \mathrm{Row}(X)) = h$ is

$$q^{(r-h)(r_X-h)} \begin{bmatrix} m - r_X \\ r - h \end{bmatrix} \begin{bmatrix} r_X \\ h \end{bmatrix}. \qquad (7.4.16)$$

Substituting (7.4.16) into (7.4.15) gives the result. □

Chapter 8

The Gamma Channel

8.1 Overview

In this chapter we consider the Gamma channel, defined in Section 1.6, a channel used to model random linear network coding (see Section 2.4.3).

Chapters 5 and 6 determine the behaviour of the MMC and AMC channel capacities respectively, leading to natural classes of optimal input distributions. However, as discussed in Section 6.6, the techniques from these chapters are not enough to fully determine the behaviour of the Gamma channel capacity. This chapter takes an alternative approach to studying the Gamma channel capacity and optimal input distributions.

We show that a capacity-achieving input distribution can always be taken to have a very restricted form (the distribution should be uniform given the rank of the input matrix). A corollary of this result is that the Gamma channel capacity may be expressed as a maximisation over probability distributions on the set of possible ranks of input matrices (rather than all possible input matrices): a set of linear rather than exponential size. This gives an efficient way to compute the exact channel capacity and find an optimal input distribution for any channel parameters.

The chapter is organised as follows. Section 8.2 shows that the distribution on the rank of the output of the Gamma channel depends only on the distribution of the rank of the input and the channel parameters. Recall that a UGR distribution is a distribution that picks matrices uniformly once their rank is determined. Section 8.3 shows that (Theorem 8.3.5) a UGR input distribution can achieve channel capacity. Section 8.4 uses the result of Theorem 8.3.5 to determine the channel capacity as a maximisation over distributions on the set of possible input ranks (Corollary 8.4.2).

8.2 Input and output rank distributions

Recall the definition (Definition 1.6.1) of the Gamma channel:

Definition. Let $\mathcal{R}$ be a probability distribution on the set $\{0, 1, \ldots, \min\{m, n\}\}$ of possible ranks of matrices $M \in \mathbb{F}_q^{n\times m}$. The *Generalised Additive Multiplicative MAtrix Channel with rank error distribution* $\mathcal{R}$ *(the Gamma channel* $\Gamma(\mathcal{R})$*)* has input set $\mathcal{X}$ and output set $\mathcal{Y}$, where $\mathcal{X} = \mathcal{Y} = \mathbb{F}_q^{n\times m}$. The channel law is

$$\boldsymbol{Y} = \boldsymbol{A}(\boldsymbol{X} + \boldsymbol{B})$$

where $\boldsymbol{A} \in \mathrm{GL}(n, q)$ is chosen uniformly, where $\boldsymbol{B} \in \mathbb{F}_q^{n\times m}$ is UGR with rank distribution $\mathcal{R}$, and where $\boldsymbol{A}$ and $\boldsymbol{B}$ are chosen independently.

A distribution $\mathcal{P}_{\boldsymbol{X}}$ on the input set $\mathcal{X}$ of the Gamma channel induces a distribution (the *input rank distribution*) $\mathcal{R}_{\boldsymbol{X}}$ on the set of possible ranks of input matrices. Let $\mathcal{R}_{\boldsymbol{Y}}$ be the corresponding *output rank distribution*, induced from the distribution on the output set of the Gamma channel. A key result (Lemma 8.2.2) is that $\mathcal{R}_{\boldsymbol{Y}}$ depends on only the channel parameters and $\mathcal{R}_{\boldsymbol{X}}$ (rather than on $\mathcal{P}_{\boldsymbol{X}}$ itself). This section aims to prove this result: it will play a vital role in the proof of Theorem 8.3.5 below.

Definition 8.2.1. Let $r, r_X, r_B \in \{0, \dots, \min\{n, m\}\}$. Define

$$\rho(r; r_X, r_B) = \frac{f_2(r, r_X, r_B)}{|\mathbb{F}_q^{n\times m, r_B}|},$$

where f_2 is as defined in Theorem 7.4.3. For any fixed matrix $X \in \mathbb{F}_q^{n\times m, r_X}$, $\rho(r; r_X, r_B)$ gives the proportion of matrices $B \in \mathbb{F}_q^{n\times m, r_B}$ with $\mathrm{rk}(X + B) = r$. Let $\mathcal{R}$ be a probability distribution on the set $\{0, 1, \dots, \min\{n, m\}\}$ of possible ranks of $n \times m$ matrices. Define

$$\rho(r; r_X) = \sum_{r_B=0}^{\min\{n,m\}} \mathcal{R}(r_B)\rho(r; r_X, r_B),$$

so that $\rho(r; r_X)$ gives the weighted average of this proportion over the possible ranks of matrices B.

Lemma 8.2.1. *Let $\boldsymbol{X}$ be an $n \times m$ matrix, sampled from some distribution $\mathcal{P}_{\boldsymbol{X}}$ on $\mathbb{F}_q^{n\times m}$. Let $\boldsymbol{B}$ be an $n \times m$ matrix sampled from a UGR distribution with rank distribution $\mathcal{R}$, where $\boldsymbol{X}$ and $\boldsymbol{B}$ are chosen independently. Let $r, r_X, r_B \in \{0, \dots, \min\{n, m\}\}$. Then*

$$\rho(r; r_X, r_B) = \Pr(\mathrm{rk}(\boldsymbol{X} + \boldsymbol{B}) = r |\, \mathrm{rk}(\boldsymbol{X}) = r_X \textit{ and } \mathrm{rk}(\boldsymbol{B}) = r_B), \tag{8.2.1}$$

and

$$\rho(r; r_X) = \Pr(\mathrm{rk}(\boldsymbol{X} + \boldsymbol{B}) = r |\, \mathrm{rk}(\boldsymbol{X}) = r_X). \tag{8.2.2}$$

Proof. Let X be a fixed $n \times m$ matrix of rank r_X. Then, since $\boldsymbol{B}$ has a UGR distribution,

$$\begin{aligned}
&\Pr(\mathrm{rk}(X + \boldsymbol{B}) = r |\, \mathrm{rk}(\boldsymbol{B}) = r_B) \\
&\quad= \frac{|\{B \in \mathbb{F}_q^{n\times m, r_B} : \mathrm{rk}(X + B) = r\}|}{|\mathbb{F}_q^{n\times m, r_B}|} \\
&\quad= \frac{f_2(r, r_X, r_B)}{|\mathbb{F}_q^{n\times m, r_B}|} \\
&\quad= \rho(r; r_X, r_B).
\end{aligned} \tag{8.2.3}$$

Note that (8.2.3) only depends on $\operatorname{rk}(X)$, not X itself. Hence

$$\begin{aligned}
&\Pr(\operatorname{rk}(\boldsymbol{X}+\boldsymbol{B})=r|\operatorname{rk}(\boldsymbol{X})=r_X,\operatorname{rk}(\boldsymbol{B})=r_B)\\
&=\sum_X \Pr(\boldsymbol{X}=X)\Pr(\operatorname{rk}(X+\boldsymbol{B})=r|\operatorname{rk}(\boldsymbol{B})=r_B)\\
&=\sum_X \Pr(\boldsymbol{X}=X)\rho(r;r_X,r_B)\\
&=\rho(r;r_X,r_B),
\end{aligned}$$

where the sums are over matrices $X\in\mathbb{F}_q^{n\times m,r_X}$. Thus (8.2.1) holds. Also

$$\begin{aligned}
&\Pr(\operatorname{rk}(\boldsymbol{X}+\boldsymbol{B})=r|\operatorname{rk}(\boldsymbol{X})=r_X)\\
&\quad=\sum_{r_B=0}^{\min\{n,m\}}\mathcal{R}(r_B)\rho(r;r_X,r_B)\text{ (by (8.2.1))}\\
&\quad=\rho(r;r_X).
\end{aligned}$$

Thus (8.2.2) holds, and so the lemma follows. □

Lemma 8.2.2. *For the Gamma channel $\Gamma(\mathcal{R})$ with input rank distribution $\mathcal{R}_{\boldsymbol{X}}$, the output rank distribution is given by*

$$\mathcal{R}_{\boldsymbol{Y}}(r)=\sum_{r_X,r_B=0}^{\min\{n,m\}}\mathcal{R}_{\boldsymbol{X}}(r_X)\mathcal{R}(r_B)\frac{f_2(r,r_X,r_B)}{|\mathbb{F}_q^{n\times m,r_B}|}$$

for $r=1,\ldots,\min\{n,m\}$. In particular, $\mathcal{R}_{\boldsymbol{Y}}$ depends only on the input rank distribution (and the channel parameters), not on the input distribution itself.

Proof. We have that $\Pr(\operatorname{rk}(\boldsymbol{X})=r_X)=\mathcal{R}_{\boldsymbol{X}}(r_X)$ and $\Pr(\operatorname{rk}(\boldsymbol{B})=r_B)=\mathcal{R}(r_B)$. Hence, by (8.2.1),

$$\begin{aligned}
\mathcal{R}_{\boldsymbol{Y}}(r)&=\Pr(\operatorname{rk}(\boldsymbol{Y})=r)\\
&=\sum_{r_X,r_B=0}^{\min\{n,m\}}\mathcal{R}_{\boldsymbol{X}}(r_X)\mathcal{R}(r_B)\rho\left(r_Y;r_X,r_B\right)\\
&=\sum_{r_X,r_B=0}^{\min\{n,m\}}\mathcal{R}_{\boldsymbol{X}}(r_X)\mathcal{R}(r_B)\frac{f_2(r,r_X,r_B)}{|\mathbb{F}_q^{n\times m,r_B}|}.
\end{aligned}$$ □

8.3 A UGR input distribution achieves capacity

This section shows (Theorem 8.3.5) that there exists a UGR input distribution to the Gamma channel that achieves capacity.

Lemma 8.3.1. *Let M and M' be fixed $n \times m$ matrices of the same rank. Let $\boldsymbol{B}$ be an $n \times m$ matrix picked from a UGR distribution, and let $\boldsymbol{A}$ be an $n \times n$ matrix picked uniformly from $\mathrm{GL}(n,q)$, with $\boldsymbol{B}$ and $\boldsymbol{A}$ picked independently. Let $\boldsymbol{Y} = \boldsymbol{A}(M+\boldsymbol{B})$ and let $\boldsymbol{Y'} = \boldsymbol{A}(M'+\boldsymbol{B})$. Then*

$$H(\boldsymbol{Y}) = H(\boldsymbol{Y'}).$$

Proof. Let A be a fixed $n \times n$ invertible matrix. Since the matrices AM and AM' have the same rank, there exist invertible matrices R and C such that $AM' = RAMC$. Consider the linear transformation $\varphi : \mathbb{F}_q^{n\times m} \to \mathbb{F}_q^{n\times m}$ defined by $\varphi(\boldsymbol{Y}) = R\boldsymbol{Y}C$. It is simple to check that φ is well defined and a bijection. Note that

$$\begin{aligned}\varphi(A(M+\boldsymbol{B})) &= RAMC + RA\boldsymbol{B}C\\ &= A(M' + A^{-1}RA\boldsymbol{B}C).\end{aligned}$$

Since $\boldsymbol{B}$ is picked uniformly once its rank is determined, pre- and post-multiplying $\boldsymbol{B}$ by fixed invertible matrices gives a uniform matrix of the same rank, therefore $\boldsymbol{B}$ and $A^{-1}RA\boldsymbol{B}C$ have the same distribution. Now

$$\begin{aligned}
&\Pr(\boldsymbol{Y} = Y|\boldsymbol{A} = A)\\
&\quad = \Pr(A(M + \boldsymbol{B}) = Y)\\
&\quad = \Pr(\varphi(A(M + \boldsymbol{B})) = \varphi(Y))\\
&\quad = \Pr\big(A(M' + A^{-1}RA\boldsymbol{B}C) = \varphi(Y)\big)\\
&\quad = \Pr(A(M' + \boldsymbol{B}) = \varphi(Y)) \qquad (8.3.1)\\
&\quad = \Pr(\boldsymbol{Y'} = \varphi(Y)|\boldsymbol{A} = A), \qquad (8.3.2)
\end{aligned}$$

where (8.3.1) holds since the distributions of $\boldsymbol{B}$ and $A^{-1}RA\boldsymbol{B}C$ are the same. Since (8.3.2) is true for any fixed matrix A, it follows that

$$\begin{aligned}
\Pr(\boldsymbol{Y} = Y) &= \sum_{A\in \mathrm{GL}(n,q)} \Pr(\boldsymbol{A} = A)\Pr(\boldsymbol{Y} = Y|\boldsymbol{A} = A)\\
&= \sum_{A\in \mathrm{GL}(n,q)} \Pr(\boldsymbol{A} = A)\Pr(\boldsymbol{Y'} = \varphi(Y)|\boldsymbol{A} = A)\\
&= \Pr(\boldsymbol{Y'} = \varphi(Y)). \qquad (8.3.3)
\end{aligned}$$

Thus $\boldsymbol{Y}$ and $\boldsymbol{Y'}$ have the same distribution, up to relabeling by φ. In particular, we find that $H(\boldsymbol{Y}) = H(\boldsymbol{Y'})$. □

Definition 8.3.1. Let M be any $n \times m$ matrix of rank r. Let $\boldsymbol{A}$ be an $n \times n$ invertible matrix chosen uniformly from $\mathrm{GL}(n, q)$. Let $\boldsymbol{B}$ be an $n \times m$ matrix chosen from a UGR distribution with rank distribution $\mathcal{R}$, where $\boldsymbol{A}$ and $\boldsymbol{B}$ are picked independently. We define

$$h_r = H(\boldsymbol{A}(M + \boldsymbol{B})).$$

Lemma 8.3.1 implies that the value h_r does not depend on M, only on the rank r and the channel parameters q, n, m and $\mathcal{R}$. The exact value of h_r will be calculated later in Theorem 8.4.1.

Lemma 8.3.2. *Consider the Gamma channel* $\Gamma(\mathcal{R})$*. Let the input matrix* $\boldsymbol{X}$ *be sampled from a distribution* $\mathcal{P}_{\boldsymbol{X}}$ *with associated rank distribution* $\mathcal{R}_{\boldsymbol{X}}$*, and let* $\boldsymbol{Y}$ *be the corresponding output matrix. Then*

$$H(\boldsymbol{Y}|\boldsymbol{X}) = \sum_{r=0}^{\min\{n,m\}} \mathcal{R}_{\boldsymbol{X}}(r) h_r.$$

In particular, $H(\boldsymbol{Y}|\boldsymbol{X})$ *depends only on the associated input rank distribution* $\mathcal{R}_{\boldsymbol{X}}$ *and the channel parameters.*

Proof. Choosing $\boldsymbol{A}$ and $\boldsymbol{B}$ as in the definition of the Gamma channel, we see that

$$\begin{aligned} H(\boldsymbol{Y}|\boldsymbol{X}) &= \sum_{X \in \mathcal{X}} P(\boldsymbol{X} = X) H(\boldsymbol{A}(X + \boldsymbol{B})) \\ &= \sum_{X \in \mathcal{X}} P(\boldsymbol{X} = X) h_{\mathrm{rk}(X)} \\ &= \sum_{r=0}^{\min\{n,m\}} \mathcal{R}_{\boldsymbol{X}}(r) h_r, \end{aligned}$$

which establishes the first assertion of the lemma. The second assertion follows since h_r depends only on r and the channel parameters. □

Lemma 8.3.3. *Let* $\boldsymbol{Y_1}$ *and* $\boldsymbol{Y_2}$ *be two random* $n \times m$ *matrices, sampled from distributions with the same associated rank distribution* $\mathcal{R}_{\boldsymbol{Y}}$*. If the distribution of* $\boldsymbol{Y_2}$ *is UGR then* $H(\boldsymbol{Y_2}) \geq H(\boldsymbol{Y_1})$*.*

Proof. For $i = 1, 2$, since $\mathrm{rk}(\boldsymbol{Y_i})$ is fully determined by $\boldsymbol{Y_i}$ it follows that $H(\boldsymbol{Y_i}, \mathrm{rk}(\boldsymbol{Y_i})) = H(\boldsymbol{Y_i})$. Therefore by the chain rule for entropy (Lemma 3.4.3),

$$\begin{aligned} H(\boldsymbol{Y_i}) &= H(\boldsymbol{Y_i}, \mathrm{rk}(\boldsymbol{Y_i})) \\ &= H(\boldsymbol{Y_i}|\, \mathrm{rk}(\boldsymbol{Y_i})) + H(\mathrm{rk}(\boldsymbol{Y_i})). \end{aligned} \tag{8.3.4}$$

Since $\boldsymbol{Y_2}$ is distributed uniformly once its rank is determined, $H(\boldsymbol{Y_2}|\, \mathrm{rk}(\boldsymbol{Y_2}) = r)$ is maximal (Lemma 3.4.2), hence

$$H(\boldsymbol{Y_2}|\, \mathrm{rk}(\boldsymbol{Y_2}) = r) \geq H(\boldsymbol{Y_1}|\, \mathrm{rk}(\boldsymbol{Y_1}) = r). \tag{8.3.5}$$

Thus, using (8.3.4)

$$\begin{aligned}
H(\boldsymbol{Y_2}) &= H(\boldsymbol{Y_2}|\operatorname{rk}(\boldsymbol{Y_2})) + H(\operatorname{rk}(\boldsymbol{Y_2})) \\
&= \sum_{r=0}^{\min\{n,m\}} (\mathcal{R}_{\boldsymbol{Y}}(r) H(\boldsymbol{Y_2}|\operatorname{rk}(\boldsymbol{Y_2}) = r)) + H(\operatorname{rk}(\boldsymbol{Y_2})) \\
&\geq \sum_{r=0}^{\min\{n,m\}} (\mathcal{R}_{\boldsymbol{Y}}(r) H(\boldsymbol{Y_1}|\operatorname{rk}(\boldsymbol{Y_1}) = r)) + H(\operatorname{rk}(\boldsymbol{Y_2})) \qquad (8.3.6) \\
&= H(\boldsymbol{Y_1}|\operatorname{rk}(\boldsymbol{Y_1})) + H(\operatorname{rk}(\boldsymbol{Y_2})) \\
&= H(\boldsymbol{Y_1}|\operatorname{rk}(\boldsymbol{Y_1})) + H(\operatorname{rk}(\boldsymbol{Y_1})) \qquad (8.3.7) \\
&= H(\boldsymbol{Y_1}),
\end{aligned}$$

where (8.3.6) follows from (8.3.5), and (8.3.7) follows since the rank distributions of $\boldsymbol{Y_1}$ and $\boldsymbol{Y_2}$ are the same. □

Lemma 8.3.4. *Consider the Gamma channel $\Gamma(\mathcal{R})$. If the input distribution $\mathcal{P}_{\boldsymbol{X}}$ is UGR then the induced output distribution $\mathcal{P}_{\boldsymbol{Y}}$ is also UGR.*

Proof. Suppose the input distribution is UGR, with rank distribution $\mathcal{R}_{\boldsymbol{X}}$. We start by showing that the distribution of $\boldsymbol{X} + \boldsymbol{B}$ is UGR. Let D be any $n \times m$ matrix. Then

$$\begin{aligned}
&\Pr(\boldsymbol{X} + \boldsymbol{B} = D) \\
&= \sum_{X \in \mathbb{F}_q^{n \times m}} \Pr(\boldsymbol{X} = X) \Pr(\boldsymbol{X} + \boldsymbol{B} = D | \boldsymbol{X} = X) \\
&= \sum_{X \in \mathbb{F}_q^{n \times m}} \frac{\mathcal{R}_{\boldsymbol{X}}(\operatorname{rk}(X))}{|\mathbb{F}_q^{n \times m, \operatorname{rk}(X)}|} \Pr(X + \boldsymbol{B} = D),
\end{aligned}$$

since $\boldsymbol{X}$ is sampled from a UGR distribution. Hence

$$\begin{aligned}
&\Pr(\boldsymbol{X}+\boldsymbol{B}=D)\\
&=\sum_{r=0}^{\min\{n,m\}} \frac{\mathcal{R}_{\boldsymbol{X}}(r)}{|\mathbb{F}_q^{n\times m,r}|} \sum_{X\in\mathbb{F}_q^{n\times m,r}} \Pr(\boldsymbol{B}=D-X)\\
&=\sum_{r=0}^{\min\{n,m\}} \frac{\mathcal{R}_{\boldsymbol{X}}(r)}{|\mathbb{F}_q^{n\times m,r}|} \sum_{X\in\mathbb{F}_q^{n\times m,r}} \frac{\mathcal{R}(\mathrm{rk}(D-X))}{|\mathbb{F}_q^{n\times m,\mathrm{rk}(D-X)}|},
\end{aligned}$$

since $\boldsymbol{X}$ and $\boldsymbol{B}$ are independent, and since $\boldsymbol{B}$ has a UGR distribution with rank distribution $\mathcal{R}$. Now

$$\begin{aligned}
&\sum_{X\in\mathbb{F}_q^{n\times m,r}} \frac{\mathcal{R}(\mathrm{rk}(D-X))}{|\mathbb{F}_q^{n\times m,\mathrm{rk}(D-X)}|}\\
&=\sum_{r_B=0}^{\min\{n,m\}} |\{X\in\mathbb{F}_q^{n\times m,r} : \mathrm{rk}(D-X)=r_B\}| \frac{\mathcal{R}(r_B)}{|\mathbb{F}_q^{n\times m,r_B}|}\\
&=\sum_{r_B=0}^{\min\{n,m\}} f_2(r_B,\mathrm{rk}(D),r)\frac{\mathcal{R}(r_B)}{|\mathbb{F}_q^{n\times m,r_B}|}
\end{aligned}$$

and so

$$\begin{aligned}
&\Pr(\boldsymbol{X}+\boldsymbol{B}=D)\\
&=\sum_{r=0}^{\min\{n,m\}} \frac{\mathcal{R}_{\boldsymbol{X}}(r)}{|\mathbb{F}_q^{n\times m,r}|} \sum_{r_B=0}^{\min\{n,m\}} f_2(r_B,\mathrm{rk}(D),r)\frac{\mathcal{R}(r_B)}{|\mathbb{F}_q^{n\times m,r_B}|}.
\end{aligned}$$

So $\Pr(\boldsymbol{X}+\boldsymbol{B}=D)$ does not depend on the specific matrix D, only its rank. Therefore, given any two $n\times m$ matrices D_1, D_2 of the same rank,

$$\Pr(\boldsymbol{X}+\boldsymbol{B}=D_1)=\Pr(\boldsymbol{X}+\boldsymbol{B}=D_2).$$

Hence $\boldsymbol{X}+\boldsymbol{B}$ has a UGR distribution.

Let A be a fixed $n\times n$ invertible matrix. Since $\boldsymbol{X}+\boldsymbol{B}$ is picked uniformly once its rank is determined, multiplying $\boldsymbol{X}+\boldsymbol{B}$ by the invertible matrix A will give a uniform matrix of the same rank, therefore $A(\boldsymbol{X}+\boldsymbol{B})$ has a UGR distribution. So, defining $\boldsymbol{Y} = \boldsymbol{A}(\boldsymbol{X} + \boldsymbol{B})$ to be the output of the Gamma channel, we see that for any $n \times m$ matrix Y

$$\begin{aligned}\Pr(\boldsymbol{Y} = Y | \boldsymbol{A} = A) &= \Pr(A(\boldsymbol{X} + \boldsymbol{B}) = Y) \\ &= \frac{\Pr\left(\operatorname{rk}(A(\boldsymbol{X} + \boldsymbol{B})) = \operatorname{rk}(Y)\right)}{|\mathbb{F}_q^{n\times m, \operatorname{rk}(Y)}|} \\ &= \frac{\Pr\left(\operatorname{rk}(\boldsymbol{Y}) = \operatorname{rk}(Y) | \boldsymbol{A} = A\right)}{|\mathbb{F}_q^{n\times m, \operatorname{rk}(Y)}|},\end{aligned}$$

where the second equality follows since $A(\boldsymbol{X} + \boldsymbol{B})$ has a UGR distribution. Thus

$$\begin{aligned}\Pr(\boldsymbol{Y} = Y) &= \sum_{A \in \operatorname{GL}(n,q)} \Pr(\boldsymbol{A} = A) \Pr(\boldsymbol{Y} = Y | \boldsymbol{A} = A) \\ &= \sum_{A \in \operatorname{GL}(n,q)} \Pr(\boldsymbol{A} = A) \frac{\Pr(\operatorname{rk}(\boldsymbol{Y}) = \operatorname{rk}(Y) | \boldsymbol{A} = A)}{|\mathbb{F}_q^{n\times m, \operatorname{rk}(Y)}|} \\ &= \frac{1}{|\mathbb{F}_q^{n\times m, \operatorname{rk}(Y)}|} \Pr(\operatorname{rk}(\boldsymbol{Y}) = \operatorname{rk}(Y)). \qquad (8.3.8)\end{aligned}$$

Since (8.3.8) holds for all $Y \in \mathbb{F}_q^{n\times m}$ it follows that $\boldsymbol{Y}$ has a UGR distribution.

□

Theorem 8.3.5. *For the Gamma channel $\Gamma(\mathcal{R})$, there exists a UGR input distribution that achieves channel capacity. Moreover, given any input distribution $\mathcal{P}_{\boldsymbol{X}}$ with associated rank distribution $\mathcal{R}_{\boldsymbol{X}}$, if $\mathcal{P}_{\boldsymbol{X}}$ achieves capacity then the UGR distribution with rank distribution $\mathcal{R}_{\boldsymbol{X}}$ achieves capacity.*

Proof. Let $\boldsymbol{X_1}$ be a channel input, with output $\boldsymbol{Y_1}$ such that $\mathcal{P}_{\boldsymbol{X_1}}$ is a capacity achieving input distribution. That is $\max_{\mathcal{P}_{\boldsymbol{X}}}\{I(\boldsymbol{X}, \boldsymbol{Y})\} = I(\boldsymbol{X_1}, \boldsymbol{Y_1})$. Then define the input $\boldsymbol{X_2}$ with output $\boldsymbol{Y_2}$ to be distributed such that $\mathcal{P}_{\boldsymbol{X_2}}$ is the UGR distribution with $\mathcal{R}_{\boldsymbol{X_2}} = \mathcal{R}_{\boldsymbol{X_1}}$. To prove the theorem it suffices to show $I(\boldsymbol{X_2}, \boldsymbol{Y_2}) \geq I(\boldsymbol{X_1}, \boldsymbol{Y_1})$.

By Lemma 8.2.2, $\mathcal{R}_{\mathbf{Y_2}} = \mathcal{R}_{\mathbf{Y_1}}$ and by Lemma 8.3.4, $\boldsymbol{Y_2}$ has a UGR distribution. Therefore, by Lemma 8.3.3,

$$H(\mathbf{Y_2}) \geq H(\boldsymbol{Y_1}). \tag{8.3.9}$$

Also, since $\mathcal{R}_{\boldsymbol{X_2}} = \mathcal{R}_{\boldsymbol{X_1}}$, Lemma 8.3.2 implies that

$$H(\mathbf{Y_2}|\boldsymbol{X_2}) = H(\boldsymbol{Y_1}|\boldsymbol{X_1}). \tag{8.3.10}$$

Using (8.3.9) and (8.3.10), it follows that

$$\begin{aligned} I(\boldsymbol{X_2}, \mathbf{Y_2}) &= H(\mathbf{Y_2}) - H(\mathbf{Y_2}|\boldsymbol{X_2}) \\ &\geq H(\mathbf{Y_1}) - H(\mathbf{Y_2}|\boldsymbol{X_2}) \\ &= H(\mathbf{Y_1}) - H(\boldsymbol{Y_1}|\boldsymbol{X_1}) \\ &= I(\boldsymbol{X_1}, \boldsymbol{Y_1}). \qquad \square \end{aligned}$$

8.4 Optimal input distributions and channel capacity

Recall that the channel capacity is defined to be the maximum mutual information between channel input and output over all possible input distributions (Definition 3.4.7). Theorem 8.3.5 reduces the problem of computing the Gamma channel capacity to a maximisation over a set of variables of linear rather than exponential size, since the UGR distribution is determined by the distribution $\mathcal{R}_{\boldsymbol{X}}$ on a set of size $\min\{n, m\} + 1$. In this section we give an expression for this maximisation problem in terms of the channel parameters and the efficiently computable functions f_0, f_1 and f_2 defined in Chapter 7. Since the mutual information is concave when considered as a function over possible input distributions (Lemma 3.4.6), this is a concave maximisation problem and hence efficiently computable (see e.g. [7]). Therefore the expression obtained provides a means for efficiently computing the exact channel capacity, and determining an optimal input rank distribution.

We begin by computing the value of h_r, as defined in Definition 8.3.1. This is needed in the proof of the expression given for the maximisation problem in Corollary 8.4.2.

Theorem 8.4.1. *The value h_r, as defined in Definition 8.3.1, is given by*

$$h_r = \sum_{v=0}^{\min\{n,m\}} \sum_{h=0}^{\min\{r,v\}} q^{(v-r+h)h} \begin{bmatrix} m-r \\ v-r+h \end{bmatrix} \begin{bmatrix} r \\ r-h \end{bmatrix}$$
$$\cdot \left(\sum_{r_B=h}^{\min\{n,m,v+h\}} \mathcal{R}(r_B) \frac{f_1(r,v,h;r_B)}{|\mathbb{F}_q^{n\times m, r_B}|} \right) \log \left(\frac{f_0(v)}{\sum_{r_B=h}^{\min\{n,m,v+h\}} \mathcal{R}(r_B) \frac{f_1(r,v,h;r_B)}{|\mathbb{F}_q^{n\times m, r_B}|}} \right).$$

where f_0 is as defined in Lemma 7.4.1 and f_1 is as defined in Theorem 7.4.2.

Proof. Let M be a fixed $n \times m$ matrix of rank r. Let $\boldsymbol{Y} = \boldsymbol{A}(M + \boldsymbol{B})$, where $\boldsymbol{A}$ is picked uniformly from $\mathrm{GL}(n,q)$ and $\boldsymbol{B}$ has a UGR distribution with rank distribution $\mathcal{R}$. Then

$$h_r = H(\boldsymbol{A}(M+\boldsymbol{B})|\,\mathrm{rk}(M) = r) = H(\boldsymbol{Y}).$$

Since $\mathrm{Row}(\boldsymbol{Y})$ is fully determined by $\boldsymbol{Y}$, it follows that $H(\boldsymbol{Y}, \mathrm{Row}(\boldsymbol{Y})) = H(\boldsymbol{Y})$. Therefore, using the chain rule for entropy (Lemma 3.4.3), we have

$$\begin{aligned} H(\boldsymbol{Y}) &= H(\boldsymbol{Y}, \mathrm{Row}(\boldsymbol{Y})) \\ &= H(\boldsymbol{Y}|\,\mathrm{Row}(\boldsymbol{Y})) + H(\mathrm{Row}(\boldsymbol{Y})). \qquad (8.4.1) \end{aligned}$$

Now, multiplying $(M + \boldsymbol{B})$ by a uniformly picked invertible matrix will result in a uniform matrix of the same rowspace as $(M+\boldsymbol{B})$. That is, the distribution of $\boldsymbol{Y}$ is uniform given the rowspace of $\boldsymbol{Y}$. Thus by Lemma 3.4.2

$$\begin{aligned} H(\boldsymbol{Y}|\,\mathrm{Row}(\boldsymbol{Y}) = V) &= \log\left(|\{Y' : Y' \in \mathbb{F}_q^{n\times m}, \mathrm{Row}(Y') = V\}|\right) \\ &= \log\left(f_0(\dim(V))\right), \qquad (8.4.2) \end{aligned}$$

where f_0 is as defined in Lemma 7.4.1. Therefore

$$\begin{aligned} H(\boldsymbol{Y}|\,\mathrm{Row}(\boldsymbol{Y})) &= \sum_{V\subseteq\mathbb{F}_q^m} \Pr(\mathrm{Row}(\boldsymbol{Y})=V)H(\boldsymbol{Y}|\,\mathrm{Row}(\boldsymbol{Y})=V) \\ &= \sum_{V\subseteq\mathbb{F}_q^m} \Pr(\mathrm{Row}(\boldsymbol{Y})=V)\log\left(f_0(\dim(V))\right). \qquad (8.4.3)\end{aligned}$$

Hence

$$\begin{aligned} h_r &= H(\boldsymbol{Y}) \\ &= H(\boldsymbol{Y}|\,\mathrm{Row}(\boldsymbol{Y})) + H(\mathrm{Row}(\boldsymbol{Y})) \\ &= \sum_{V\subseteq\mathbb{F}_q^m} \Pr(\mathrm{Row}(\boldsymbol{Y})=V)\log\left(f_0(\dim(V))\right) \\ &\qquad\qquad - \sum_{V\subseteq\mathbb{F}_q^m} \Pr(\mathrm{Row}(\boldsymbol{Y})=V)\log\left(\Pr(\mathrm{Row}(\boldsymbol{Y})=V)\right) \\ &= \sum_{V\subseteq\mathbb{F}_q^m} \Pr(\mathrm{Row}(\boldsymbol{Y})=V)\log\left(\frac{f_0(\dim(V))}{\Pr(\mathrm{Row}(\boldsymbol{Y})=V)}\right). \qquad (8.4.4)\end{aligned}$$

Now, we calculate the probability of $\boldsymbol{Y}$ having a particular rowspace V. For $V \subseteq \mathbb{F}_q^m$, let $d_{V'} = \dim((\mathrm{Row}(M)+V)/V)$. Using the function f_1, defined in Theorem 7.4.2, we obtain the following result.

$$\begin{aligned} &\Pr(\mathrm{Row}(\boldsymbol{Y})=V) \\ &\qquad = \Pr(\mathrm{Row}(M+\boldsymbol{B})=V) \\ &\qquad = \sum_{r_B=0}^{\min\{n,m\}} \Pr(\mathrm{rk}(\boldsymbol{B})=r_B)\Pr(\mathrm{Row}(M+\boldsymbol{B})=V|\,\mathrm{rk}(\boldsymbol{B})=r_B) \\ &\qquad = \sum_{r_B=0}^{\min\{n,m\}} \mathcal{R}(r_B)\frac{|\{B:\mathrm{rk}(B)=r_B,\mathrm{Row}(M+B)=V\}|}{|\mathbb{F}_q^{n\times m,r_B}|} \qquad (8.4.5) \\ &\qquad = \sum_{r_B=0}^{\min\{n,m\}} \mathcal{R}(r_B)\frac{f_1(r,\dim(V),d_{V'};r_B)}{|\mathbb{F}_q^{n\times m,r_B}|} \\ &\qquad = \sum_{r_B=d_{V'}}^{\min\{n,m,\dim(V)+d_{V'}\}} \mathcal{R}(r_B)\frac{f_1(r,\dim(V),d_{V'};r_B)}{|\mathbb{F}_q^{n\times m,r_B}|}, \qquad (8.4.6)\end{aligned}$$

where (8.4.5) follows since $\boldsymbol{B}$ has a UGR distribution and (8.4.6) follows since $f_1(r, \dim(V), d_{V'}; r_B) = 0$ for $r_B < d_{V'}$ and $r_B > \dim(V) + d_{V'}$.

Substituting (8.4.6) into (8.4.4) we get

$$h_r = \sum_{V \subseteq \mathbb{F}_q^m} \left(\sum_{r_B = d_{V'}}^{\min\{n,m,\dim(V)+d_{V'}\}} \mathcal{R}(r_B) \frac{f_1(r, \dim(V), d_{V'}; r_B)}{|\mathbb{F}_q^{n\times m, r_B}|} \right) \\ \cdot \log \left(\frac{f_0(\dim(V))}{\sum_{r_B=d_{V'}}^{\min\{n,m,\dim(V)+d_{V'}\}} \mathcal{R}(r_B) \frac{f_1(r,\dim(V),d_{V'};r_B)}{|\mathbb{F}_q^{n\times m, r_B}|}} \right) \quad (8.4.7)$$

In (8.4.7), for a given subspace $V \subseteq \mathbb{F}_q^m$, the corresponding term in the sum depends only on $\dim(V)$ and the value $d_{V'} = \dim(\mathrm{Row}(M) + V)/V)$. Therefore, we will count the number of spaces V with $\dim(V) = v$ and $\dim(\mathrm{Row}(M) + V)/V) = h$ for some v and h. We can then replace the sum over all V by a sum over the values v and h.

Since $f_1(r, \dim(V), h; r_B) = 0$ if $\dim(V) > n$ we can restrict our attention to subspaces V of $\mathbb{F}_q^m$ with $\dim(V) \leq n$. Given some integers v, h with $0 \leq v \leq \min\{n, m\}$ and $0 \leq h \leq \min\{r, v\}$, by Corollary 3.2.10 the number of v-dimensional subspaces $V \subseteq \mathbb{F}_q^m$ such that $\dim((\mathrm{Row}(M) + V)/V) = h$ is

$$q^{(v-(r-h))(r-(r-h))} \begin{bmatrix} m-r \\ v-(r-h) \end{bmatrix} \begin{bmatrix} r \\ r-h \end{bmatrix} \\ = q^{(v-r+h)h} \begin{bmatrix} m-r \\ v-r+h \end{bmatrix} \begin{bmatrix} r \\ r-h \end{bmatrix}. \quad (8.4.8)$$

Substituting (8.4.8) into (8.4.7) gives the result. □

Now we give the result of this section: an efficiently computable expression for the Gamma channel capacity as a maximisation over the set of possible input rank distributions.

Corollary 8.4.2. *The capacity of the Gamma channel* $\Gamma(\mathcal{R})$ *is given by*

$$C = \max_{\mathcal{R}_{\boldsymbol{X}}} \left\{ - \left(\sum_{r_Y=0}^{\min\{n,m\}} \left(\sum_{r_X, r_B=0}^{\min\{n,m\}} \mathcal{R}_{\boldsymbol{X}}(r_X)\mathcal{R}(r_B) \frac{f_2(r_Y, r_X, r_B)}{|\mathbb{F}_q^{n\times m, r_B}|} \right) \right. \right.$$
$$\cdot \log \left(\frac{1}{|\mathbb{F}_q^{n\times m, r_Y}|} \sum_{r_X, r_B=0}^{\min\{n,m\}} \mathcal{R}_{\boldsymbol{X}}(r_X)\mathcal{R}(r_B) \frac{f_2(r_Y, r_X, r_B)}{|\mathbb{F}_q^{n\times m, r_B}|} \right) \Bigg) \Bigg)$$
$$- \left(\sum_{r=0}^{\min\{n,m\}} \mathcal{R}_{\boldsymbol{X}}(r) \sum_{u=0}^{\min\{n,m\}} \sum_{h=0}^{\min\{r,u\}} \left(q^{(v-r+h)h} \begin{bmatrix} m-r \\ v-r+h \end{bmatrix} \begin{bmatrix} r \\ r-h \end{bmatrix} \right. \right.$$
$$\left. \left. \left. \left. \cdot \left(\sum_{r_B=0}^{\min\{n,m\}} \mathcal{R}(r_B) \frac{f_1(r,u,h;r_B)}{|\mathbb{F}_q^{n\times m, r_B}|} \right) \log \left(\frac{f_0(u)}{\sum_{r_B=0}^{\min\{n,m\}} \mathcal{R}(r_B) \frac{f_1(r,u,h;r_B)}{|\mathbb{F}_q^{n\times m, r_B}|}} \right) \right) \right) \right) \right\},$$

where f_0, f_1 *and* f_2 *are as defined in Lemma 7.4.1, Theorem 7.4.2 and Theorem 7.4.3, respectively.*

Proof. The capacity of the channel is defined to be $C = \max_{\mathcal{P}_{\boldsymbol{X}}} I(\boldsymbol{X}; \boldsymbol{Y})$. By Theorem 8.3.5, to achieve capacity we can chose the input distribution $\mathcal{P}_{\boldsymbol{X}}$ to be UGR. By Lemma 8.3.4, the output distribution will also be UGR. Therefore the output distribution is given by

$$\mathcal{P}_{\boldsymbol{Y}}(Y) = \Pr(\boldsymbol{Y} = Y) = \frac{1}{|\mathbb{F}_q^{n\times m, \mathrm{rk}(Y)}|} \mathcal{R}_{\boldsymbol{Y}}(\mathrm{rk}(Y)) \tag{8.4.9}$$

for any $Y \in \mathbb{F}_q^{n\times m}$. Thus the entropy of $\boldsymbol{Y}$ is given by

$$\begin{aligned} H(\boldsymbol{Y}) &= - \sum_{Y \in \mathbb{F}_q^{n\times m}} \Pr(\boldsymbol{Y} = Y) \log \Pr(\boldsymbol{Y} = Y) \\ &= - \sum_{Y \in \mathbb{F}_q^{n\times m}} \left(\frac{1}{|\mathbb{F}_q^{n\times m, \mathrm{rk}(Y)}|} \mathcal{R}_{\boldsymbol{Y}}(\mathrm{rk}(Y)) \right) \log \left(\frac{1}{|\mathbb{F}_q^{n\times m, \mathrm{rk}(Y)}|} \mathcal{R}_{\boldsymbol{Y}}(\mathrm{rk}(Y)) \right) \\ &= - \sum_{r_Y=0}^{\min\{n,m\}} \mathcal{R}_{\boldsymbol{Y}}(r_Y) \log \left(\frac{1}{|\mathbb{F}_q^{n\times m, r_Y}|} \mathcal{R}_{\boldsymbol{Y}}(r_Y) \right). \end{aligned} \tag{8.4.10}$$

Therefore, using (8.4.10) and Lemma 8.3.2,

$$
\begin{aligned}
I(\boldsymbol{X};\boldsymbol{Y}) &= H(\boldsymbol{Y}) - H(\boldsymbol{Y}|\boldsymbol{X}) \\
&= -\sum_{r_Y=0}^{\min\{n,m\}} \mathcal{R}_{\boldsymbol{Y}}(r_Y) \log\left(\frac{1}{|\mathbb{F}_q^{n\times m,r_Y}|}\mathcal{R}_{\boldsymbol{Y}}(r_Y)\right) - \sum_{r_X=0}^{\min\{n,m\}} \mathcal{R}_{\boldsymbol{X}}(r_X) h_{r_X}.
\end{aligned}
$$

Substituting the expressions for $\mathcal{R}_{\boldsymbol{Y}}$ and h_r given in Lemma 8.2.2 and Theorem 8.4.1 respectively, and taking the maximum over all possible input rank distributions yields the result. □

Chapter 9

Conclusion

This thesis has considered several mathematical problems motivated by network coding. We began by considering partial decoding in random linear network coding and then analysed several finite field matrix channels that can be used to model random linear network coding in various situations.

Whereas previous literature computes the probability of complete recovery of network-coded messages, we computed the exact probability of recovering some fraction of the message and investigated the implication to network coding protocols and secure communication. The derived expressions can prove useful in network design and system-level optimisation, as discussed in Section 4.5. We focused on the case of random linear network coding, assuming the coding vectors form a uniform random matrix. This is a widely considered case due to the simplicity of implementation and its efficiency. However, in practice it may be useful to consider sparse linear network coding, where the receiver obtains sparse linear combinations of source packets. This is of interest in practical implementations because it can vastly reduce decoding complexity. Therefore, an interesting area for future research would be to compute and analyse the probability of decoding a fraction of the source message when the coding vectors form a sparse random matrix.

We considered several finite field matrix channels to model network coding. Building on the work of Silva, Kschischang and Kötter [39], we defined a generalisation of the channels the authors consider, allowing the modelling of network coding in a wider variety of cases, with different error patterns. We gave bounds on the capacities of the MMC and AMC channels that differ by small additive constants and are independent of all channel parameters. For the general Gamma channels we showed that an optimal input distribution can always be taken to have a very restricted form (the distribution should be uniform given the rank of the input matrix). We expressed the capacity for the Gamma channel as a maximisation over probability distributions on the set of possible ranks of input matrices: a set of linear rather than exponential size. Thus we gave an efficient method for computing the exact channel capacity and finding optimal input distributions for any channel parameters. The expressions obtained are complex and without further analysis give little intuition to the behaviour of the channel capacity. Therefore there are several possible areas for future research. It would be useful to use our formula to obtain some data for various channel parameters, to gain understanding of the behaviour of the Gamma channel capacity and optimal input rank distributions. Furthermore, using this data or otherwise, one could compare our exact expression of capacity to the previously known bounds for the special case of the Gamma channel with fixed error rank, considered by [39] (see Section 6.6). It is not clear whether the bounds in Section 6.6 can be improved further, if possible it would be very useful to give a neat bound that gives more intuition on the behaviour of the channel capacity for all parameter values.

To conclude this work has focused on questions motivated by random linear network coding. We have contributed to the field by identifying a generalised class of matrix channels to model network coding, giving a thorough analysis of the capacity of these channels and partial decoding in network coding. Several questions remain open giving plenty of scope for future research.

Bibliography

[1] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung. Network information flow. *IEEE Transactions on Information Theory*, 46(4):1204–1216, Jul 2000.

[2] T. M. Apostol. *Introduction to Analytic Number Theory, Undergraduate Texts in Mathematics.* Springer-Verlag, New York, 1976.

[3] R. Bassoli, H. Marques, J. Rodriguez, K. W. Shum, and R. Tafazolli. Network coding theory: A survey. *IEEE Communications Surveys & Tutorials*, 15(4):1950–1978, 2013.

[4] E. A. Bender and J. R. Goldman. On the applications of Möbius inversion in combinatorial analysis. *The American Mathematical Monthly*, 82(8):789–803, October 1975.

[5] E. R. Berlekamp. The technology of error-correcting codes. *Proceedings of the IEEE*, 68(5):564–593, May 1980.

[6] K. Bhattad and K. R. Narayanan. Weakly secure network coding. In *Proceedings 1st Workshop on Network Coding, Theory and Applications (NETCOD)*, volume 104, Riva Del Garda, Italy, April 2005.

[7] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge university press, 2004.

[8] N. Cai and R. W. Yeung. Secure network coding. In *Information Theory Proceedings (ISIT), 2002 IEEE International Symposium on*, page 323, Lausanne, Switzerland, June 2002.

[9] P. J. Cameron. *Combinatorics: Topics, techniques, algorithms.* Cambridge University Press, 1994.

[10] C.-C. Chao, C.-C. Chou, and H.-Y. Wei. Pseudo random network coding design for IEEE 802.16m enhanced multicast and broadcast service. In *Vehicular Technology Conference (VTC 2010-Spring), 2010 IEEE 71st*, Taipei, Taiwan, May 2010.

[11] P. A. Chou, Y. Wu, and K. Jain. Practical network coding. 2003.

[12] J. Claridge and I. Chatzigeorgiou. Probability of Partially Solving Random Linear Systems in Network Coding. *ArXiv e-prints*, July 2016.

[13] T. M. Cover and J. A. Thomas. *Elements of information theory.* John Wiley & Sons, 2012.

[14] È. M. Gabidulin. Theory of codes with maximum rank distance. *Problems of Information Transmission*, 21(1):1–12, January 1985.

[15] M. Gadouleau and A. Goupil. A matroid framework for noncoherent random network communications. *IEEE Transactions on Information Theory*, 57(2):1031–1045, 2011.

[16] M. Gadouleau and Z. Yan. On the decoder error probability of bounded rank-distance decoders for maximum rankdistance codes. *IEEE Transactions on Information Theory*, 54(7):3202–3206, 2008.

[17] M. Gadouleau and Z. Yan. Packing and covering properties of rank metric codes. *IEEE Transactions on Information Theory*, 54(9):3873–3883, 2008.

[18] M. Gadouleau and Z. Yan. Bounds on covering codes with the rank metric. *IEEE Communications Letters*, 13(9):691–693, 2009.

[19] M. Gadouleau and Z. Yan. Packing and covering properties of subspace codes for error control in random linear network coding. *IEEE Transactions on Information Theory*, 56(5):2097–2108, 2010.

[20] P. R. Halmos. *Finite-Dimensional Vector Spaces*. Undergraduate Texts in Mathematics. Springer, 1974.

[21] G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities*. Cambridge University Press, 1934.

[22] J. Heide, M. V. Pedersen, F. H. P. Fitzek, and M. Médard. On code parameters and coding vector representation for practical RLNC. In *Communications (ICC), 2011 IEEE International Conference on*, Kyoto, Japan, June 2011.

[23] T. Ho, M. Médard, R. Kötter, D. R. Karger, M. Effros, J. Shi, and B. Leong. A random linear network coding approach to multicast. *IEEE Transactions on Information Theory*, 52(10):4413–4430, October 2006.

[24] A. L. Jones, I. Chatzigeorgiou, and A. Tassi. Binary systematic network coding for progressive packet decoding. In *2015 IEEE International Conference on Communications (ICC)*, pages 4499–4504. IEEE, 2015.

[25] A. Khaleghi, D. Silva, and F. R. Kschischang. Subspace codes. In *IMA International Conference on Cryptography and Coding*, pages 1–21. Springer, 2009.

[26] R. Kötter and F. R. Kschischang. Coding for errors and erasures in random network coding. *IEEE Transactions on Information Theory*, 54(8):3579–3591, Aug 2008.

[27] R. Kötter and M. Médard. An algebraic approach to network coding. *IEEE/ACM Transactions on Networking (TON)*, 11(5):782–795, 2003.

[28] S.-Y. R. Li, R. W. Yeung, and N. Cai. Linear network coding. *IEEE Transactions on Information Theory*, 49(2):371–381, 2003.

[29] L. Lima, M. Médard, and J. Barros. Random linear network coding: A free cipher? In *Information Theory Proceedings (ISIT), 2007 IEEE International Symposium on*, Nice, France, June 2007.

[30] H.-T. Lin, Y.-Y. Lin, and H.-J. Kang. Adaptive network coding for broadband wireless access networks. *IEEE Transactions on Parallel and Distributed Systems*, 24(1):4–18, January 2013.

[31] D. E. Lucani, M. Médard, and M. Stojanovic. On coding for delay – Network coding for time-division duplexing. *IEEE Transactions on Information Theory*, 58(4):2330–2348, April 2012.

[32] A. Montanari and R. L. Urbanke. Iterative coding for network coding. *IEEE Transactions on Information Theory*, 59(3):1563–1572, 2013.

[33] R. W. Nobrega, D. Silva, and B. F. Uchoa-Filho. On the capacity of multiplicative finite-field matrix channels. *IEEE Transactions on Information Theory*, 59(8):4949–4960, 2013.

[34] G.-C. Rota. On the foundations of combinatorial theory I. Theory of Möbius functions. *Probability theory and related fields*, 2(4):340–368, January 1964.

[35] M. Sanna and E. Izquierdo. A survey of linear network coding and network error correction code constructions and algorithms. *International Journal of Digital Multimedia Broadcasting*, 2011:12, 2011.

[36] B. Shrader and N. M. Jones. Systematic wireless network coding. In *Military Communications Conference (MILCOM), 2009 IEEE*, Boston, MA, October 2009.

[37] M. J. Siavoshani, S. Mohajer, C. Fragouli, and S. N. Diggavi. On the capacity of noncoherent network coding. *IEEE Transactions on Information Theory*, 57(2):1046–1066, 2011.

[38] D. Silva, F. R. Kschischang, and R. Kötter. A rank-metric approach to error control in random network coding. *IEEE Transactions on Information Theory*, 54(9):3951–3967, 2008.

[39] D. Silva, F. R. Kschischang, and R. Kötter. Communication over finite-field matrix channels. *IEEE Transactions on Information Theory*, 56(3):1296–1305, March 2010.

[40] O. Trullols-Cruces, J. Barcelo-Ordinas, and M. Fiore. Exact decoding probability under random linear network coding. *IEEE Communications Letters*, 15(1):67–69, January 2011.

[41] J. H. van Lint and R. M. Wilson. *A Course In Combinatorics*. Cambridge University Press, 2 edition, 2001.

[42] Z. Yan, H. Xie, and B. W. Suter. Rank deficient decoding of linear network coding. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, Vancouver, BC, May 2013.

[43] Z. Zhang. Linear network error correction codes in packet networks. *IEEE Transactions on Information Theory*, 54(1):209–218, 2008.

[44] X. Zhao. Notes on "Exact decoding probability under random linear network coding". *IEEE Communications Letters*, 16(5):720–721, May 2012.