Universidade Estadual de Campinas

Instituto de Matemática, Estatística
e Computação Científica

Jerry Anderson Pinheiro

FORMA CANÔNICA PARA CÓDIGOS POSET E ESQUEMAS DE CODIFICAÇÃO-DECODIFICAÇÃO PARA PERDA ESPERADA

CANONICAL FORM FOR POSET CODES AND CODING-DECODING SCHEMES FOR EXPECTED LOSS

Campinas

2016

Jerry Anderson Pinheiro

Canonical Form for Poset Codes and Coding-Decoding Schemes for Expected Loss

Forma Canônica para Códigos Poset e Esquemas de Codificação-Decodificação para Perda Esperada

Tese apresentada ao Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Doutor em Matemática.

Thesis presented to the Institute of Mathematics, Statistics and Scientific Computing of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor in Mathematics.

Orientador: Marcelo Firer

Este exemplar corresponde à versão final da tese defendida pelo aluno Jerry Anderson Pinheiro, e orientada pelo Prof. Dr. Marcelo Firer.

Campinas

2016

Agência(s) de fomento e nº(s) de processo(s): CNPq, 141586/2014-1; CAPES

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Ana Regina Machado - CRB 8/5467

P655c

Pinheiro, Jerry Anderson, 1985-
Canonical form for poset codes and coding-decoding schemes for expected loss / Jerry Anderson Pinheiro. – Campinas, SP : [s.n.], 2016.

Orientador: Marcelo Firer.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Métricas sobre ordens parciais. 2. Códigos corretores de erros (Teoria da informação). I. Firer, Marcelo,1961-. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Forma canônica para códigos poset e esquemas de codificação-decodificação para perda esperada
Palavras-chave em inglês:
Poset metrics
Correcting codes (Information theory)
Área de concentração: Matemática
Titulação: Doutor em Matemática
Banca examinadora:
Marcelo Firer [Orientador]
Reginaldo Palazzo Junior
Marcelo Muniz Silva Alves
Marcelo da Silva Pinho
Emerson Luiz do Monte Carmelo
Data de defesa: 25-04-2016
Programa de Pós-Graduação: Matemática

Tese de Doutorado defendida em 25 de abril de 2016 e aprovada pela Banca Examinadora composta pelos Profs. Drs.

Prof. Dr. MARCELO FIRER

Prof. Dr. REGINALDO PALAZZO JUNIOR

Prof. Dr. MARCELO MUNIZ SILVA ALVES

Prof. Dr. MARCELO DA SILVA PINHO

Prof. Dr. EMERSON LUIZ DO MONTE CARMELO

A Ata da defesa com as respectivas assinaturas dos membros encontra-se no processo de vida acadêmica do aluno.

Agradecimentos

Durante esses pouco mais de 4 anos de trabalho, inúmeras foram as pessoas e instituições que de alguma forma me ajudaram nessa caminhada. Para algumas, uma página de agradecimento seria pouco, no entanto, corri o risco, fiz uma síntese e omiti algumas, porém não as esqueci.

Agradeço aos meus pais, Zilton e Adiles, pelo incondicional apoio. Apesar da distância que alimenta a saudade, vocês são os principais culpados desta conquista.

Agradeço aos professores Marcelo Firer e Marcus Greferath; o primeiro, não apenas pela orientação do doutorado na Unicamp mas também pela confiança depositada e pelos ensinamentos; e o segundo, pela orientação durante o período de um ano que passei na University College Dublin.

Para que um projeto seja iniciado, sempre há quem dê uma motivação e semeie uma ideia, um dos principais responsáveis para que eu me aventurasse pela pós graduação na Unicamp, e a quem deixo aqui meus agradecimentos, foi o professor Luciano Panek.

Uma pessoa não consegue passar tanto tempo em um lugar sem cultivar amigos, cultivei muitos, alguns listarei aqui, porém muito mais ficarão na memória. Agradeço aos companheiros de pós graduação Cleber, Felix, João (Jhon); aos companheiros de laboratório, Campello, Ana, Bruno, Elen, Akemi, Julianna (Ju), Cintya, Giselle, Eleonésio, Alessandro, Adriana e Cláudio; aos meus “irmãos” de trabalho “filhos” do mesmo orientador: Marcos (Marquinhos), Roberto, Christiane (Chris), Rafael e Luciano, este último, parceiro nas “pint” de Guinness nos pubs de Dublin; e aos amigos que cultivei na Irlanda durante meu doutorado sanduíche, Jens, Oliver, Cornelia, Ana e Carolina (Carol), as duas últimas, parceiras em nossa “trip” pela Irlanda.

Agradeço ao CNPq pelo suporte financeiro.

Resumo

No contexto de códigos corretores de erros, métricas são utilizadas para definir decodificadores de máxima proximidade, uma alternativa aos decodificadores de máxima verossimilhança. A família de métricas poset tem sido extensivamente estudada no contexto de teoria de códigos. Considerando a estrutura do grupo de isometrias lineares, é obtida uma forma canônica para matrizes geradoras de códigos lineares. Esta forma canônica permite obter expressões e limitantes analíticos para alguns invariantes clássicos da teoria: raio de empacotamento e complexidade de síndrome. Ainda, substituindo a probabilidade de erro pela perda esperada definida pelo desvio médio quadrático (entre a informação original e a informação decodificada), definimos uma proposta de codificação com ordem lexicográfica que, em algumas situações é ótima e em outras, as simulações feitas sugerem um desempenho ao menos subótimo. Finalmente, relacionamos a medida de perda esperada com proteção desigual de erros, fornecendo uma construção de códigos com dois níveis de proteção desigual de erros e com perda esperada menor que a obtida pelo produto de dois códigos ótimos, que separam as informações que são protegidas de modo diferenciado.

Palavras-chave: Métricas Sobre Ordens Parciais, Códigos Corretores de Erros (Teoria da Informação)

Abstract

In the context of error-correcting codes, metrics are used to define minimum distance decoders, an alternative to maximum likelihood decoders. The family of poset metrics has been extensively studied in the context of coding theory. Considering the structure of the group of linear isometries, we obtain a canonical form for generator matrices of linear codes. The canonical form allows to obtain analytics expressions and bounds for classical invariants of the theory: packing radius and syndrome complexity. By substituting the error probability by the expected loss defined by the mean square deviation (between the original information and the decoded information), we propose an encoder scheme which, in some situations is optimal, and in others the simulations suggest a performance at least sub-optimal. Finally, we relate the expected loss measure with unequal error protection, providing a construction of codes with two levels of unequal error protection and expected loss smaller than the one obtained by the product of two optimal codes, which divide the information that is protected differently.

Keywords: Poset Metrics, Error Correcting Codes (Information Theory).

Contents

Introduction

Metrics are mathematical structures of interest in coding theory. Several works are devoted to the study of metrics in the context of coding theory, and the best known and investigated metric in this context is the Hamming metric. It was suggested by R. W. Hamming in [25] when describing a geometric model of a code. Later, in ([28],1958) and ([44],1957), the Lee metric was defined, and became an interesting alternative when non-binary alphabets are used. Up to our knowledge, the first work considering metrics in a general approach is a short communication of S. W. Golomb [22] in 1969. In that work, it was described a family of additive metrics (metrics defined over an alphabet and extended additively to a set of words with a fixed length) which are still being investigated nowadays, as we can see in [41]. Recently, the interest on different and larger families of metrics in coding theory has been increased, as can be seen, for example, in [6], [2], [20] and [10]. This is partially motivated by the fact that metrics provide a decoding scheme using minimum distance, which in some cases (when metrics and channels are matched), is an alternative to Maximum a Posteriori (MAP) decoders. Also, minimum distance decoders may be used to add manageability to the decoding process, what can be achieved by the use of a Syndrome decoding algorithm, which is the most general and efficient decoding algorithm presented in this dissertation.

One of those large families of metrics (with interest in coding theory) is the family of poset metrics, they were introduced by Brualdi et al in ([6],1991) as a generalization of metrics obtained by Niederreiter in [35] and [36]. A poset metric on an n-dimensional vector space is determined by the choice of a partial order on the set $\{1, 2, \dots, n\}$. To describe the classical parameters of coding theory for such metrics is, in general, a difficult problem. For example, in [12], it was proved that to determining the packing radius of an one-dimensional code is an NP-hard problem. However, as we can see in [30], the family of hierarchical poset metrics, which is determined by the sub-family of hierarchical posets, is a natural generalization of the Hamming metric (which belongs to this sub-family) and, classical coding invariants for those metrics are “easy” (as easy as in the Hamming case) to obtain by using the canonical-systematic form for linear codes, determined in [15].

Canonical forms are obtained by using the group of linear isometries and determine a standard, and relatively clean representation of codes, see [15] and [1]. Besides the hierarchical case, the only known attempt to generalize it was made in [1], where a standard form for a particular case (Niederreiter-Rosenbloom-Tsfasman (NRT), or orders consisting of multiple disjoint chains of the same order) is presented. In that work, one can see that the standard form is not unique (in any possible sense). In a matter of fact, as we will see later, unicity of such a decomposition is a characteristic of hierarchical posets.

Here, considering general posets, we construct a standard form, decomposing a code as a direct sum of smaller codes, and this form is canonical with respect to the length and dimension of the smaller codes. The choice of a particular representation for codes (the canonical form), is motivated by the fact that it can be used to easily determine some code parameters. For the general case, there are no closed and general expressions for coding invariants, but using the canonical decomposition and comparing to hierarchical posets, we obtain bounds for two important invariants: the packing radius of a code and the complexity of syndrome decoding.

Assuming that the minimum (poset) distance decoding is a relevant decoding criterion, despite the fact we do not explore it, we are actually assuming some underlying situation (possibly given by a channel model). If the specific situation of interest in the coding-decoding process was not made explicit when studying the canonical form of poset metrics, the model of the channel and a model for evaluating the errors is the core of the last part of this work, where we propose some alternatives to unequal error protection.

In the classical coding theory, decoders are constructed in order to minimize both the error and refusal probabilities. In many communications scenarios, it is more efficient to better protect only a crucial part of the information. In order to achieve this goal, in [32], it was proposed the framework of unequal error protection (UEP). To evaluate the performance of a coding system with an unequal error protection, Masnik and Wolf introduced a measure called Average Error Cost (AEC). In this context, it is assigned, for each information bit, a value corresponding the protection level of the bit (information bits with equal assigned value have the same importance). Then, considering these levels of protection, a value is assigned to each bit and this determines a cost for an error in a specific bit. The values on each bit position define a way to measure the cost of an error: the sum of the errors in each bit, each of those weighted by the value of the bit position. The AEC is then defined as the average of this quantity, the average being taken over all transmitted and received codeword.

In [17], it was presented a broader framework for AEC, considering the expected loss function (ELF) of a coding-decoding scheme. Here, to each pair of codewords it is assigned a value, representing the value of the error occurred when one information is exchanged by the other. In this sense, the AEC may be seen as a particular case of measure, which occurs in the instance of a ELF that is invariant by translations.

As noted by Masnik and Wolf in [32], the main difference among the classical coding theory and the one with unequal error protection, is that the encoder is a relevant part of systems with UEP, while in the classical theory, for a given code, the error probability depends only on the code, not on the encoder. In this context, we explore some possibilities for encoding and decoding separately. First of all, we propose a lexicographic encoder that in some simple situations is proved to be optimal, with some experimental evidences of very good performance in more general situations. We also explore a possibility of choice of a code and encoding process that performs a two-levels UEP, showing also some experimental measurements for its performance. Some relevant conjectures concerning those proposals are left open.

This work is organized as follows:

In Chapter 1 we describe the most common decoding schemes: maximum likelihood, maximum a posteriori probability and minimum distance decoders. We start with a very general definition of a model for a decoder as a stochastic map. With the goal of minimizing both the error and refusal probabilities simultaneously, we prove that we can restrict the attention to deterministic decoders, which is one of the most common definition of decoders found in the literature. In addition to pointing out the relevance of metrics in coding theory, this chapter also makes a connection between the concept of error probability and expected loss, which is introduced in the last chapter.

In Chapter 2, the basics of coding theory are introduced: linear codes, code equivalence, packing radius and group of linear isometries. We also introduce the basic concepts, definitions and properties concerning posets and poset metrics.

The original contributions are concentrated in the last two chapters.

Chapter 3 is devoted to the canonical form for a generator matrix of poset codes. The decomposition of a code according to the poset metric and its canonical form are defined and constructed. Using the maximal decomposition, we obtain bounds for the packing radius and the complexity of syndrome decoding.

Chapter 4 starts exploring the framework of expected loss in the same generality used in the first chapter to introduce the error and refusal probabilities. We show that, considering the expected loss, it is possible to exchange a probabilistic decoder by a deterministic one and, on the encoding side, the problem of minimizing the ELF may be translated into a problem of minimizing the trace of some matrices. The ELF of an average encoder is determined as the ELF concerning an equal valued system of information (the most usual instance in coding theory) and comparison to this average encoder can be used as a performance measure of an encoding scheme. The last two sections are devoted to particular cases, one regarding encoders and the other regarding decoders. In these sections, some conjectures are presented and a practical example is described.

Chapter 1

Metrics in Coding Theory

This chapter is a brief introduction on decoders and the role of metrics in coding theory. As a complementary reading we cite the recent survey of Gabidulin [20]. Even though we introduce the concepts of information theory in a slightly different way from Gabidulin's survey, the survey can also be used as a supplementary reading.

The stochastic map notation will be used in order to define channels and decoders. The stochastic maps approach simplifies the notations of channels and allows us to justify the usual definition for decoder. Thus, before we discuss information and coding theory, we shall define the two basic mathematical concepts that underlie this work:

Definition 1. (*Stochastic Map*) Let $\mathcal{P}_{\mathcal{Y}}$ be the set of all probability distributions over a finite set $\mathcal{Y}$. A *stochastic map* $\mathcal{P} : \mathcal{X} \to \mathcal{P}_{\mathcal{Y}}$ is a map from a finite set $\mathcal{X}$ to a probability distribution over $\mathcal{Y}$.

Given $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, the expression $y \sim \mathcal{P}(x)$ means that the element y was sampled from the probability distribution $\mathcal{P}(x)$. We will write $\mathcal{P}(y|x) = Pr(y = y'|y' \sim \mathcal{P}(x))$ to express the conditional probability for the occurrence of the event y given that x has occurred.

Definition 2. (*Metric*) A *metric* d over $\mathcal{X}$ is a function $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ satisfying the following conditions:

(a) $d(x, y) \geq 0$ and $d(x, y) = 0$ if, and only if, $x = y$;

(b) $d(x, y) = d(y, x)$;

(c) $d(x, y) \leq d(x, z) + d(z, y)$ for all $x, y, z \in \mathcal{X}$.

1.1 Decoders over Discrete Channels

The first geometric model of a code was suggested by R. W. Hamming in ([25], 1950). This model gave rise to the well-known Hamming metric, the most investigated metric in coding theory. Just after this Hamming contribution, motivated by the cyclic structure of non-binary alphabets, the Lee model was introduced in ([44], 1957) and ([28], 1958). The Hamming metric is important for two reasons: it matches with one of the most studied channels in information theory, the Binary Symmetric Channel (BSC); and it is simple enough to allow the design of "good" decoding algorithms for specific types of codes. New geometric models of codes have been studied in coding theory and, consequently, new families of metrics fitting these models have been proposed, see [40], [6], [2] and [10]. In order to establish the precise relation between metrics and coding theory, we will first give a brief description of a communication system.

The Shannon model of a point-to-point communication system, as shown in figure 1.1, breaks the process of communication down into a handful of components. It is a minimalist abstraction of the reality; actually, in the "real world", most of the communication systems are much more complex. Each component of the model has its importance in the transmission process which may vary according to the system application.

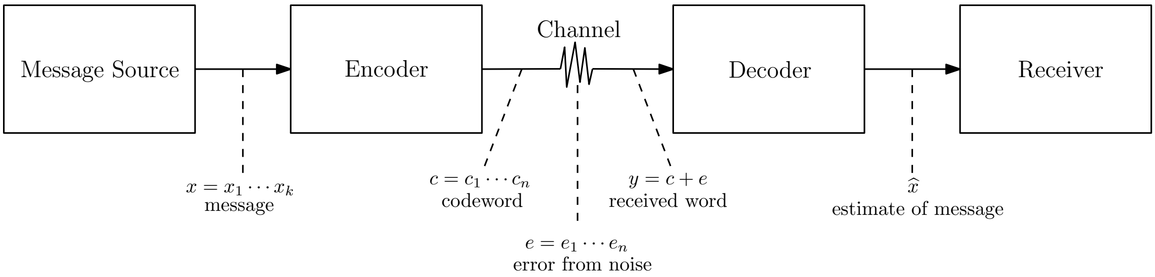


Figure 1.1: Communication System

Basically, the functioning of this communication system is as follows: The *Message Source* generates messages with length k in order to send them to the *Receiver*. This generator is modeled by a random variable and in most of the cases it is assumed to be uniformly distributed. The *Encoder* adds redundancy in each message following mathematical rules, increasing the block of data from length k to length n. This redundancy will provide structure in the ambient space in order to allow the *Decoder* to detect and correct some errors that eventually occur when a noisy channel is used. Each component described here will be formally defined since they are necessary to the comprehension of the main objects of this work: the decoders.

Definition 1.1.1. Let $\mathcal{X}$ and $\mathcal{Y}$ be finite sets. A *discrete channel* is a stochastic map $W : \mathcal{X} \to \mathcal{P}_{\mathcal{Y}}$ where $\mathcal{X}$ is called the input alphabet and $\mathcal{Y}$ the output alphabet of the channel.

We would like to stress that for a given channel W, the expression $W(y|x)$ denotes the probability to receive y given that x was sent. The transition probabilities (or conditional probabilities) of a channel are in general obtained from experimental data. If no information is known, the worst-case scenario, or an approximation to it, is assumed.

Given an alphabet $\mathcal{X}$, denote by $\mathcal{X}^n$ the set of all *words* with length n over this alphabet. A *block channel* with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ is a discrete channel $W_n : \mathcal{X}^n \to \mathcal{P}_{\mathcal{Y}^n}$. Note that block channels deal with words with a fixed length while the channels defined in 1.1.1 deal with alphabets. Given a block channel W_n, if there is a channel W such that for every $y = (y_1, \dots, y_n) \in \mathcal{Y}^n$ and $x = (x_1, \dots, x_n) \in \mathcal{X}^n$ with $y_i \in \mathcal{Y}$ and $x_i \in \mathcal{X}$ we have that

$$W_n(y|x) = \prod_{i=1}^{n} W(y_i|x_i),$$

the discrete block channel W_n will be denoted by W^n and called *memoryless*. We remark that W^n is obtained by extending the channel W to arrays. Due to this, it is often called the n-th extension of W. Only discrete channels are considered in this work, therefore the word discrete will be omitted. For simplicity, if $x = (x_1, \dots, x_n) \in \mathcal{X}^n$, sometimes the parenthesis or the commas of x will be suppressed. Also, if no confusion may arise, the index n is going to be omitted in the block channel notation.

Example 1.1.2. (*Binary Symmetric Channel*) A Binary Symmetric Channel $W : \mathbb{F}_2 \to \mathcal{P}_{\mathbb{F}_2}$ is a channel with input and output alphabets $\mathbb{F}_2$ (finite field with 2 elements) and conditional probabilities $W(1|1) = 1 - p = W(0|0)$ and $W(1|0) = p = W(0|1)$ where $0 \leq p \leq 1/2$. This channel and its n-th extension are the most studied channels in information theory. They are called symmetric because $W(x|y) = W(y|x)$ for all $x, y \in \mathbb{F}_2$. It is usual to represent this channel by the diagram in Figure 1.2.

Example 1.1.3. (*Binary Erasure Channel*) A Binary Erasure Channel $W : \mathbb{F}_2 \to \mathcal{P}_{\mathbb{F}_2 \cup \{?\}}$ is a channel with conditional probabilities $W(1|1) = 1 - p = W(0|0)$ and $W(?\,|0) = p =$

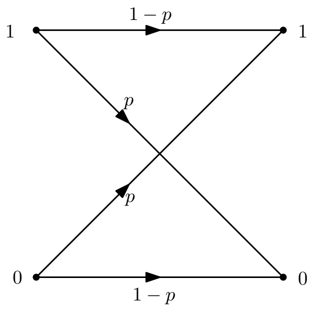


Figure 1.2: Binary Symmetric Channel

$W(?|1)$ where $0 \leq p \leq 1/2$. The symbol ? means that the bit sent was erased. This channel is frequently used in information theory because it is one of the simplest channels to analyze: whenever an error occurs, it tells you about the existence and position of the error in the array. It is represented by the diagram 1.3.

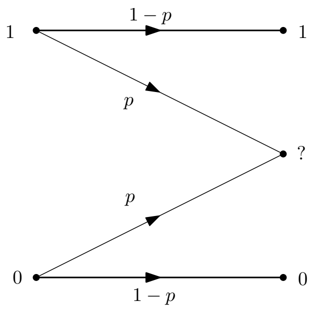


Figure 1.3: Binary Erasure Channel

Example 1.1.4. (*q-ary Symmetric Channels*) A memoryless symmetric channel with input and output alphabets $\mathcal{X}$ with $q = |\mathcal{X}|$, and crossover probability p, where $0 \leq p \leq 1/2$ is a channel defined by the following conditional probabilities: $W(y|y) = 1 - p$ and $W(y|x) = p/(q-1)$ if $x \neq y$. A particular case of these channels is the binary symmetric channel defined in Example 1.1.2.

Definition 1.1.5. Consider the set $\mathcal{X}^n$ of all words with length n over the alphabet $\mathcal{X}$. A *code* $\mathcal{C}$ is any subset of $\mathcal{X}^n$. The elements of the code $\mathcal{C}$ are called *codewords*.

An (n, M, q)-code is a code $\mathcal{C} \subset \mathcal{X}^n$ where $|\mathcal{C}|= M$ and $|\mathcal{X}|= q$. From now on, for the sake of simplicity, we will assume that the input alphabet $\mathcal{X}$ is contained in the output alphabet $\mathcal{Y}$. This restriction ensures that $\mathcal{C} \subset \mathcal{X}^n \subset \mathcal{Y}^n$ for every code $\mathcal{C}$ over $\mathcal{X}^n$. It facilitates the definition of the next structures (decoders and encoders).

An *information set* is any set with cardinality q^k where q and k are positive integers. We remark that the set of all messages can be identified with the set of all words with length k over a finite alphabet $\mathcal{X}$ with q elements, which is denoted by $\mathcal{X}^k$. Therefore, the notation (n, M, q) is going to be exchanged by $(n, k)_q$ since $M = q^k$.

Definition 1.1.6. Given an information set $\mathcal{X}^k$ and an integer $n \geq k$, an encoder is an injective map $f : \mathcal{X}^k \to \mathcal{X}^n$, its image is an $(n, k)_q$-code $\mathcal{C}$.

Definition 1.1.7. Let $\mathcal{C} \subset \mathcal{X}^n$ be an $(n, k)_q$-code. A *decoder* of $\mathcal{C}$ is a decision criteria for each $y \in \mathcal{Y}^n$ modeled by a stochastic map $D : \mathcal{Y}^n \to \mathcal{P}_{\mathcal{C}\cup\{\infty\}}$ such that $D(\infty|y) \in \{0, 1\}$ for every $y \in \mathcal{Y}^n$ and $D(c|c) = 1$ for every $c \in \mathcal{C}$. The set of all decoders of $\mathcal{C}$ will be denoted by $\mathcal{D}_{\mathcal{Y}}(\mathcal{C})$.

We remark that if $D \in \mathcal{D}_{\mathcal{Y}}(\mathcal{C})$, each $y \in \mathcal{Y}^n$ determines a (probabilistic) decision criteria given by the probability distribution $D(y)$. Given $x \in \mathcal{C} \cup \{\infty\}$ and $y \in \mathcal{Y}^n$, the value $D(x|y)$ denotes the probability to the decoder outputs x given that the word y was received. If a codeword $c \in \mathcal{C}$ is received, the assumption $D(c|c) = 1$ ensures that the decoder will assume the codeword c was sent. The symbol ∞ denotes that an error has occurred and the decoder could not solve the decision problem, refusing the received word. This means that the information must be sent again (or forgotten). Since $D(\infty|y) \in \{0, 1\}$, to decide when the received word $y \in \mathcal{Y}^n$ is going to be refused or not it is a deterministic criteria.

Given a decoder D and an encoder f for the code $\mathcal{C}$, a *full decoder* is a stochastic map $D' : \mathcal{Y}^n \to \mathcal{P}_{\mathcal{X}^k\cup\{\infty\}}$ defined by $D'(c|y) = D(f(c)|y)$ and $D'(\infty|y) = D(\infty|y)$. Note that $f^{-1} : \mathcal{C} \to \mathcal{X}^k$ is a bijection, thus the stochastic map D' is well defined. The difference between full decoders and decoders is that a decoder outputs an error ∞ or an element $c \in \mathcal{C}$, and a full decoder outputs an error or an information $f^{-1}(c) \in \mathcal{X}^k$.

Definition 1.1.8. A $(\mathcal{C}, f, D)$ encoding-decoding scheme for the channel $W : \mathcal{X}^n \to \mathcal{P}_{\mathcal{Y}^n}$ consists of

1 - An information set $\mathcal{X}^k$ with cardinality q^k where $q = |\mathcal{X}|$ and $k \leq n$;

2 - An encoder $f : \mathcal{X}^k \to \mathcal{X}^n$ where $f(\mathcal{X}^k) = \mathcal{C}$ is an $(n, k)_q$-code;

3 - A decoder (stochastic map) $D : \mathcal{Y}^n \to \mathcal{P}_{\mathcal{C}\cup\{\infty\}}$.

The main purpose of coding theory is to ensure reliable transmission of information through a noisy channel, reliability being determined by an appropriate measure.

Among the most important measures, there are the error probability and the refusal probability of a code. They express the amount of expected errors and refusals.

Given a decoder $D \in \mathcal{D}_{\mathcal{Y}}(\mathcal{C})$ and a channel W, the refusal probability for a codeword $c \in \mathcal{C}$ is the probability that the decoder refuses a vector given that c was sent, i.e.,

$$P_{ref}^{D}(c) = \sum_{y \in \mathcal{Y}^n} W(y|c)D(\infty|y).$$

We remark that the probabilities $W(y|c)$ and $D(y|c)$ are determined by the channel and the decoder respectively. The *refusal probability* of the code $\mathcal{C}$ is the mean

$$P_{ref}^{D}(\mathcal{C}) = \sum_{c \in \mathcal{C}} P_{ref}^{D}(c)P(c)$$

where $P(c)$ is the probability of the codeword $c \in \mathcal{C}$ to be sent through the channel. The error probability can be defined similarly to the refusal probability. The *error probability* of a codeword $c \in \mathcal{C}$ is defined by

$$P_{e}^{D}(c) := \sum_{y \in \mathcal{Y}^n} W(y|c)(1 - D(\infty|y) - D(c|y))$$

where $1-D(\infty|y)-D(c|y)$ is the probability that the decoder outputs a codeword different to the sent one c. The error probability of the code is defined by

$$P_{e}^{D}(\mathcal{C}) := \sum_{c \in \mathcal{C}} P_{e}^{D}(c)P(c). \tag{1.1}$$

It is clear that the error and refusal probabilities do not depend on the encoder. This is not true when considering error value functions and unequal error protection, as we can see in [16]. Assuming those probabilities as measures of reliability, we define optimal decoder:

Definition 1.1.9. For a given code $\mathcal{C}$, an *optimal decoder* D^* is a decoder satisfying

$$P_{ref}^{D^*}(\mathcal{C}) + P_{e}^{D^*}(\mathcal{C}) = \min_{D \in \mathcal{D}_{\mathcal{Y}}(\mathcal{C})} \left(P_{ref}^{D}(\mathcal{C}) + P_{e}^{D}(\mathcal{C}) \right).$$

Proposition 1.1.10. Given a decoder $D \in \mathcal{D}_{\mathcal{Y}}(\mathcal{C})$, there exists a decoder $\widetilde{D} \in \mathcal{D}_{\mathcal{Y}}(\mathcal{C})$ such that

$$P_{e}^{\widetilde{D}}(\mathcal{C}) \leq P_{ref}^{D}(\mathcal{C}) + P_{e}^{D}(\mathcal{C})$$

and $P^{\widetilde{D}}_{ref}(\mathcal{C}) = 0$.

Proof. Given a decoder $D \in \mathcal{D}_{\mathcal{Y}}(\mathcal{C})$, suppose there exists $y_0 \in \mathcal{Y}^n$ such that $D(\infty|y_0) = 1$. Define a new decoder $D^* \in \mathcal{D}_{\mathcal{Y}}(\mathcal{C})$ satisfying $D^*(y) = D(y)$ for all $y \neq y_0$ but $D^*(y_0)$ is a new distribution which does not refuse y_0, so $D^*(\infty|y_0) = 0$. Our goal here is to show that

$$P^{D^*}_{ref}(\mathcal{C}) + P^{D^*}_{e}(\mathcal{C}) \leq P^{D}_{ref}(\mathcal{C}) + P^{D}_{e}(\mathcal{C}). \tag{1.2}$$

Initially, note that since $D^*(\infty|y_0) = 0$ and $D^*(\infty|y) = D(\infty|y)$ for all $y \neq y_0$, hence

$$P^{D^*}_{ref}(\mathcal{C}) = \sum_{c\in\mathcal{C}} \sum_{y\in\mathcal{Y}^n\setminus\{y_0\}} W(y|c)D(\infty|y)P(c).$$

Therefore,

$$P^{D}_{ref}(\mathcal{C}) = P^{D^*}_{ref}(\mathcal{C}) + \sum_{c\in\mathcal{C}} W(y_0|c)P(c) \tag{1.3}$$

because $D(\infty|y_0) = 1$. On the other hand, since $D(c|y_0) = 0$ for all $c \in \mathcal{C}$, by definition of D^*,

$$P^{D}_{e}(\mathcal{C}) = \sum_{c\in\mathcal{C}} \sum_{y\in\mathcal{Y}^n\setminus\{y_0\}} W(y|c)(1 - D^*(\infty|y) - D^*(c|y))P(c),$$

thus

$$P^{D^*}_{e}(\mathcal{C}) = P^{D}_{e}(\mathcal{C}) + \sum_{c\in\mathcal{C}} W(y_0|c)(1 - D^*(c|y_0))P(c). \tag{1.4}$$

It is straightforward that

$$\sum_{c\in\mathcal{C}} W(y_0|c)(1 - D^*(c|y_0))P(c) \leq \sum_{c\in\mathcal{C}} W(y_0|c)P(c),$$

therefore,

$$P^{D^*}_{ref}(\mathcal{C}) + P^{D}_{e}(\mathcal{C}) + \sum_{c\in\mathcal{C}} P(y_0|c)(1 - D^*(c|y_0))P(c) \leq P^{D^*}_{ref}(\mathcal{C}) + P^{D}_{e}(\mathcal{C}) + \sum_{c\in\mathcal{C}} W(y_0|c)P(c).$$

Together with Identities (1.3) and (1.4), we obtain the Inequality (1.2). The decoder $\widetilde{D}$ is constructed by following this procedure until we run out of refused elements. □

Corollary 1.1.11. Given a code $\mathcal{C}$, there exists an optimal decoder $D^* \in \mathcal{D}_{\mathcal{Y}}(\mathcal{C})$ satisfying $P^{D^*}_{ref}(\mathcal{C}) = 0$.

Since we are not concerned with specific applications, and in this level of generality we are assuming only the error and refusal measures, Corollary 1.1.11 ensures that we do not lose generality by assuming the refusal probability to be zero. Therefore, we will restrict ourselves to the goal in minimizing the error probability. Thus, from now on, the decoder model will be considered as stochastic maps $D : \mathcal{Y}^n \to \mathcal{P}_{\mathcal{C}}$ (instead of $D : \mathcal{Y}^n \to \mathcal{P}_{\mathcal{C}\cup\{\infty\}}$).

1.2 Decoding Schemes

As seen in the previous section, we can consider only *complete decoders*: decoders with no refusal option. In this situation, optimal decoders are the ones minimizing the error probability. We can basically divide the problem of searching good decoders using two criteria, the usefulness and the manageability of the decoder. Due to the generality of this work, we will not deal with the manageability criteria of the decoders when dealing with metrics decoders, the manageability of metric decoders will be justified later by the possibility to use syndrome decoding, a general procedure for decoding linear codes. Therefore, from our point of view, good decoders are not necessarily practical, since they can be hard to deal with. We will now define a series of abstract decoders and add some mathematical structures on $\mathcal{X}^n$, structures that are, in most of the cases, sufficient conditions to add manageability to the decoder.

Given a channel W, by the Bayes' rule, $W(y|c)P(c) = W(c|y)P(y)$, where $P(c)$ denotes the probability of the codeword $c \in \mathcal{C}$ to be sent and $P(y)$ is the probability that $y \in \mathcal{Y}^n$ is received. We remark that we are assuming complete decoders, so if $D \in \mathcal{D}_{\mathcal{Y}}(\mathcal{C})$, the error probability can be alternatively written in the following way:

$$\begin{aligned} P_e^D(\mathcal{C}) &= \sum_{c\in\mathcal{C}} \sum_{y\in\mathcal{Y}^n} W(y|c)\left(1 - D(c|y)\right) P(c) \\ &= \sum_{c\in\mathcal{C}} \sum_{y\in\mathcal{Y}^n} W(c|y)\left(1 - D(c|y)\right) P(y) \\ &= 1 - \sum_{c\in\mathcal{C}} \sum_{y\in\mathcal{Y}^n} W(c|y)P(y)D(c|y). \end{aligned} \tag{1.5}$$

Definition 1.2.1. Given a channel W, a *Maximum a Posteriori Probability (MAP) decoder* is a decoder $D : \mathcal{Y}^n \to \mathcal{P}_{\mathcal{C}}$ such that $D(y)$ is a conditional distribution over $\mathcal{C}$ satisfying

$$W(c' \sim D(y)|y) = \max\{W(c|y) : c \in \mathcal{C}\}$$

for every $y \in \mathcal{Y}^n$.

As one should expect, decoders that minimize error probability are those that maximize the conditional probabilities $W(c|y)$, as explained below.

Lemma 1.2.2. Given a sequence of n non-negative integers $a_1 \geq \cdots \geq a_n$ such that $\sum_i a_i = 1$. Suppose $a_1 = \ldots = a_k > a_{k+1}$ for some $1 \leq k \leq n-1$. Consider the maximization problem

$$\sum_{i=1}^{n} a_i b_i' = \max_{\{b_1,\ldots,b_n\}} \sum_{i=1}^{n} a_i b_i$$

where the maximum is over the sets $\{b_1, \ldots, b_n\}$ satisfying $\sum_i b_i = 1$. Then every sequence of non-negative integers $b_1', \ldots, b_k'$ satisfying $\sum_{i=1}^{k} b_i' = 1$ is a solution for this problem.

Proof. Note that

$$\sum_{i=1}^{n} a_i b_i = a_1 \left(b_1 + \frac{a_2}{a_1} b_2 + \cdots + \frac{a_n}{a_1} b_n \right).$$

Since $a_1 \geq a_i$ for every i, it follows that $b_1 + \frac{a_2}{a_1} b_2 + \cdots + \frac{a_n}{a_1} b_n \leq 1$, therefore,

$$a_1 \left(b_1 + \frac{a_2}{a_1} b_2 + \cdots + \frac{a_n}{a_1} b_n \right) \leq a_1.$$

Then $b_1 = 1$ and $b_i = 0$ for all $i > 1$ is a solution for the maximization problem. Because $a_1 = \ldots = a_k$, every sequence of non-negative integers $b_1, \ldots, b_k$ satisfying $\sum_{i=1}^{k} b_i = 1$ is also a solution. □

Theorem 1.2.3. If D is a MAP decoder, then D is optimal.

Proof. By Equation 1.5, for any code $\mathcal{C}$,

$$P_e^D(\mathcal{C}) = 1 - \sum_{y \in \mathcal{Y}^n} \left(\sum_{c \in \mathcal{C}} W(c|y) D(c|y) \right) P(y).$$

Note that to minimize the error probability is equivalent to maximize the expression

$$\sum_{c \in \mathcal{C}} W(c|y) D(c|y)$$

for each choice of $y \in \mathcal{Y}$. Note that if $c \sim D(y)$ and $c' \sim D(y)$, by the definition of a MAP decoder, $W(c|y) = W(c'|y)$. Also, if c cannot be sampled from $D(y)$, it means that $D(c|y) = 0$. Thus, the conditional probabilities $D(c|y)$ satisfy the conditions of

Lemma 1.2.2 and are solutions for that maximization problem. Therefore, MAP decoders minimizes the error probability. □

A decoder $D \in \mathcal{D}_{\mathcal{Y}}(\mathcal{C})$ is said to be *deterministic*[1] if for every $y \in \mathcal{Y}^n$, there is an element $c \in \mathcal{C}$ such that $D(c|y) = 1$. Therefore, a deterministic decoder is a surjective map $D : \mathcal{Y}^n \to \mathcal{C}$. These decoders will be denoted by g. The next proposition ensures that we lose no generality if we consider only deterministic decoders. The proof follows directly from the definition of a MAP decoder and from Lemma 1.2.2.

Proposition 1.2.4. Given a channel W, there is a deterministic decoder $g : \mathcal{Y}^n \to \mathcal{C}$ which is a MAP decoder.

Due to Proposition 1.2.4, from now on we will consider only deterministic decoders. We can reformulate the definition of MAP decoder as follows:

Definition 1.2.5. (*MAP Decoders - Revisited*) A *Maximum a Posteriori Probability (MAP) decoder* is a decoder $g : \mathcal{Y}^n \to \mathcal{C}$ satisfying the condition

$$W(g(y)|y) = \max\{W(c|y) : c \in \mathcal{C}\}.$$

Consequently, the error probability of a code $\mathcal{C}$ can be rewritten as

$$P_e^g(\mathcal{C}) = \sum_{c\in\mathcal{C}} \sum_{y\notin g^{-1}(c)} W(y|c)P(c).$$

Instead of defining a decoder according to a posteriori probabilities $W(c|y)$, we can define it by using a priori probabilities $W(y|c)$: the probability to receive y if c is sent.

Definition 1.2.6. A *Maximum Likelihood (ML) decoder* is a decoder $g : \mathcal{Y}^n \to \mathcal{C}$ satisfying the condition

$$W(y|g(y)) = \max\{W(y|c) : c \in \mathcal{C}\}.$$

The proof of the next well-known proposition follows straight from the Bayes' rule.

Proposition 1.2.7. If the distribution of the code $P(c)$ (the probability to send c) is uniform, then a decoder is an ML decoder if and only if it is a MAP decoder.

[1] The word "deterministic" is used since for those decoders, given a particular received array, the decoder will always produce the same output.

From now on we will assume that $P(c)$ is uniformly distributed. It is clear that when dealing with decoders, we are playing with a sort of measure. One of the mathematical measures that can be used in order to decrease the computational complexity for the decoding algorithms are metrics. In order to use metrics, we will use the assumption that the input and output alphabets of a channel are the same, i.e., $\mathcal{X} = \mathcal{Y}$.

Definition 1.2.8. Given a metric d over $\mathcal{X}^n$, a *Minimum Distance (MD) decoder* is a decoder $g : \mathcal{X}^n \to \mathcal{C}$ satisfying the condition

$$d(y, g(y)) = \min\{d(y, c) : c \in \mathcal{C}\}.$$

If g is a MD decoder according to d, we will say that g is a d-MD decoder.

The Hamming metric was constructed in [25] when exploring the geometric representation (Hamming cube) of a code and it is the most studied metric in coding theory.

Example 1.2.9. (*Hamming Distance*) The function $d_H : \mathcal{X}^n \times \mathcal{X}^n \to \mathbb{R}_+$ defined by

$$d_H(x, y) = |\{i : x_i \neq y_i\}|$$

where $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ is called *Hamming distance.*

Proposition 1.2.10. For any memoryless symmetric channel with crossover probability $p \leq 1/2$, the MD decoder determined by the Hamming metric is also an ML decoder.

Proof. Let $\mathcal{X}$ be the alphabet of the channel. Given $x = (x_1, \ldots, x_n) \in \mathcal{X}^n$ and $c = (c_1, \ldots, c_n) \in \mathcal{C}$, then

$$\begin{aligned} W(y|c) &= W(y_1|c_1) \cdot \ldots \cdot W(y_n|c_n) \\ &= (1-p)^{|\{i:y_i=c_i\}|} \left(\frac{p}{q-1}\right)^{|\{i:y_i\neq c_i\}|} \\ &= (1-p)^{n-d_H(y,c)} \left(\frac{p}{q-1}\right)^{d_H(y,c)} \\ &= (1-p)^n \left(\frac{p}{(1-p)(q-1)}\right)^{d_H(y,c)}. \end{aligned}$$

Since $p/[(1-p)(q-1)] < 1$,

$$W(y|c) \geq W(y|c') \text{ if, and only if } d_H(y,c) \leq d_H(y,c')$$

for every $c, c' \in \mathcal{C}$. □

We remark that we are considering $P(c)$ to be uniformly distributed. Thus, from Propositions 1.2.7 and 1.2.10 we have the following result.

Theorem 1.2.11. In a q-ary symmetric channel with crossover probability $p \leq 1/2$, MAP decoders, ML decoders and MD decoders according to the Hamming metric are optimal decoders.

We now add some mathematical structure to $\mathcal{X}$ and $\mathcal{X}^n$ in order to develop tools that can be helpful in handling the problems that may arise when dealing with large finite sets. The most common ones are the structures of finite fields and vector spaces. We consider the alphabet $\mathcal{X}$ to be a finite field $\mathbb{F}_q$ where $q = p^r$ (p prime), so that, $\mathcal{X}^n$ is an n-dimensional vector space, namely $\mathbb{F}_q^n$.

Definition 1.2.12. A metric $d : \mathbb{F}_q^n \times \mathbb{F}_q^n \to \mathbb{R}$ is said to be *invariant by translations* if

$$d(x+z, y+z) = d(x,y)$$

for every $x, y, z \in \mathbb{F}_q^n$.

Metrics can be defined by using norm and weight functions.

Definition 1.2.13. A function $w : \mathbb{F}_q^n \to \mathbb{R}$ is a *weight* if it satisfies the following axioms:

- $w(x) \geq 0$, for every x;
- $w(x) = 0$ if, and only if, $x = 0$;

It is clear that if we define the function d by $d(x,y) = w(x-y)$, then d is a *semimetric* (it has all properties of a metric but the triangular inequality). Moreover, given a semmimetric d, the function $w(x) = d(x, \mathbf{0})$ is a weight.

The family of weight functions that we are interested in are the ones preserving the support. A particular but important family (the poset metrics) may be obtained by generalizing the Hamming metric. This particular family will be introduced in the next chapter and will be the main object of this work. The *support* of an element $x \in \mathbb{F}_q^n$ is the set

$$supp(x) = \{i \ : \ x_i \neq 0\}.$$

Definition 1.2.14. A weight function $w : \mathbb{F}_q^n \to \mathbb{R}$ is said to *preserve the support* if

$$supp(x) \subset supp(y) \Rightarrow w(x) \leq w(y).$$

Example 1.2.15. (*Combinatorial Metrics*, [21]) Let $[n] := \{1, \dots, n\}$. Consider $T = \{T_0 = \emptyset, T_1, \dots, T_s\}$ to be a family of subsets covering $[n]$, i.e., $\cup_{i=0}^{s} T_i = [n]$. The T-weight is defined by

- $w_T(x) = 0 \iff supp(x) = \emptyset$;
- $w_T(x) = 1 \iff supp(x) \subset T_i$ for some i;
- $w_T(x) = k \iff$

  $$supp(x) \subset \{a\ union\ of\ exactly\ k\ subsets\ from\ T\}$$

  but

  $$supp(x) \not\subset \{a\ union\ of\ k-1\ or\ less\ subsets\ from\ T\}.$$

The *Combinatorial Metric* d_T is obtained by taking

$$d_T(x, y) = w_T(x - y).$$

It is well-known [20] that d_T is a metric and that preserves support. Several particular cases of those metrics are well-studied in the context of coding theory, see [20] for more details.

Since d_T is also semimetric, it follows that w_T is a weight. Moreover, it is a norm in the following sense:

Definition 1.2.16. A function $N : \mathbb{F}_q^n \to \mathbb{R}$ is a *norm* if it is a weight and satisfies

$$N(x + y) \leq N(x) + N(y) \text{ for all } x, y \in \mathbb{F}_q^n.$$

Concerning metrics and semimetrics invariant by translations, we have the following:

(1) If d is a metric invariant by translations, then the function $N_d(x) = d(x, \mathbf{0})$ is a norm. Moreover, if N is a norm, the function d_N defined by $d_N(x, y) = N(x - y)$ is a metric invariant by translations.

(2) If d is a semimetric invariant by translations, then the function $w_d(x) = d(x, \mathbf{0})$ is a weight. Moreover, if w is a weight, the function d_w defined by $d_w(x, y) = w(x - y)$ is a semimetric invariant by translations.

Example 1.2.17. Not every weight preserving support is a norm. Indeed, define the weight w over $\mathbb{F}_2^2$ by setting $w(00) = 0$, $w(01) = w(10) = 1$ and $w(11) = 3$. This is a weight that preserves the support but does not satisfies the third condition of norms (triangular inequality) because $3 = w(11) > w(10) + w(01) = 2$. Then the function $d(x, y) := w(x-y)$ is a *semimetric* (satisfies all the properties of metrics but the triangular inequality).

The example below provides a norm which does not preserve support. This extension of the Lee metrics are induced by the ℓ_p metric in $\mathbb{Z}^n$, as we can see in [7].

Example 1.2.18. The Lee weight over $\mathbb{Z}_l$ is defined by

$$w_L(x) = \min\{x \pmod{l}, -x \pmod{l}\}.$$

The p extension of this norm is the p-Lee norm over $\mathbb{Z}_l^n$ where if $x = (x_1, \ldots, x_n) \in \mathbb{Z}_l^n$, then

$$w_L^p(x) = \left(\sum_{i=1}^{n} w_L(x_i)^p \right)^{1/p}.$$

Take $l > 3$, then if $supp(x) = 1 = supp(y)$ with $x_1 = 1$ and $x_2 = 2$, thus $w_L^p(x) = 1$ and $w_L^p(y) = 2$. If p-Lee weights preserve support, then we should have $w_L^p(x) = w_L^p(y)$ since $supp(x) = supp(y)$.

It is clear that if we define the function d by $d(x, y) = N(x - y)$, then d is a metric. Norm and weight functions are also related by their induced metric and semimetric. Indeed, if the semimetric satisfies the triangular inequality, then the weight function is a norm. We will see that regarding matching metrics and channels, the triangular inequality can be easily obtained by using semimetrics. The next two propositions are well-known results that we will not proof here.

Proposition 1.2.19. If a metric d is invariant by translations, then d is induced by a norm.

There is another family of weight functions with interest in coding theory, the *invariant weights*, which are weight functions satisfying $w(tx) = w(x)$ for every $t \in \mathbb{F}_q \backslash \{0\}$ and $x \in \mathbb{F}_q^n$. In [23], a characterization of invariant weights satisfying MacWilliams Extension property was given. The invariant weights and the weights preserving support are related by the next proposition, which proof follows directly from the fact that $supp(tx) = supp(x)$ for all $t \neq 0$.

Proposition 1.2.20. A weight preserving support is an invariant weight.

Even though the metrics are induced by norms, we will use the weight notation since it is commonly used in coding theory. In the next section we will present the syndrome decoding algorithm, an algorithm that works only with invariant by translations metrics (or semimetrics).

1.2.1 Syndrome Decoding

We are interested in MD decoding according to a metric (or semimetric). Assuming the metric is invariant by translations, we can use the well-known *Syndrome Decoding* algorithm as an alternative to MD decoding. This is the most powerful and general decoder presented in this thesis, justifying the use of metrics in coding theory. Our main goal is to prove that in the invariant by translation case, syndrome decoding is a minimum distance decoder.

Let $\mathcal{C}$ be a linear code over $\mathbb{F}_q^n$ ($\mathcal{C}$ is a subspace of $\mathbb{F}_q^n$). Then, $\mathcal{C}$ is the kernel of some linear transformation, therefore there is an $(n-k) \times n$ matrix H, the *parity check matrix*, satisfying

$$\mathcal{C} = \{x \in \mathbb{F}_q^n \ : \ Hx^T = 0\}.$$

The vector $Hx^T \in \mathbb{F}_q^k$ is called the *syndrome of* x, therefore it will be denoted by $Syn(x)$. Two elements belong to the same coset of $\mathcal{C}$ if, and only if, they have the same syndrome, i.e.,

$$x + \mathcal{C} = y + \mathcal{C} \iff Syn(x) = Syn(y).$$

Syndrome Decoding Algorithm.

Precomputation: For each coset $x+\mathcal{C}$, we choose a coset leader $\overline{x}$ such that $w(\overline{x}) = \min\{w(z) \ : \ z \in x+\mathcal{C}\}$.

Input: $y \in \mathbb{F}_q^n$

1. Find the coset leader $\overline{x}$ such that $Syn(\overline{x}) = Syn(y)$.

Output: $y - \overline{x}$.

Given that y was received, all the possible errors have the same syndrome of y, hence it is obvious that the error to be found is an element of a particular coset, the one having the syndrome of y. In order to conclude that syndrome decoding is a minimum distance decoding, we need to proof that the coset leader with smaller weight is the appropriate choice. The next example shows that the hypothesis that d is invariant by translations is essential.

Example 1.2.21. Let $d : \mathbb{F}_2^2 \times \mathbb{F}_2^2 \to \mathbb{R}_+$ be a metric defined by the distance table below:

d(x,y)	00	01	10	11
00	0	2	2	3
01	2	0	2	4
10	2	2	0	1
11	3	4	1	0

It is clear that d is indeed a metric over $\mathbb{F}_2^2$, furthermore, d is not invariant by translation since

$$2 = d(01,10), \quad 3 = d(00,11) \ \text{ and } \ d(01+01,10+01) = d(00,11).$$

Let $\mathcal{C} = \{00,01\}$ be an 1-dimensional linear code and suppose we want to decode $y = 11$. If the decoder used is a d-MD, then the closest codeword to y is 00 since $3 = d(00,11) < d(01,11) = 4$. Suppose now we will use syndrome decoding according to d. The parity check matrix of $\mathcal{C}$ is given by $H = [1 \ 0]$. Note that $Syn(10) = Syn(11) = 1$ and that $Syn(00) = Syn(01) = 0$. Since $Syn(y) = 1$, the possible errors are 10 and 11. Note that $2 = d(10,00) < d(11,00) = 3$. Due to this, syndrome decoding assumes 10 as being the error, so it outputs $y - 10 = 01$, but as we saw, MD decoders output 00, therefore syndrome is not an MD decoder.

Theorem 1.2.22. If d is invariant by translations, then syndrome decoding is an MD decoder.

Proof. Suppose y is the vector to be decoded and that c' was obtained by decoding y using syndrome, then $y = e + c'$ with $Syn(e) = Syn(y)$. This representation is not unique since $y = e_2 + (c' - c_1)$ for every $c_1 \in C$ where $e_2 = e + c_1$. Assume that e is a coset leader. Because $y - c' = e$ and $y - (c' - c_1) = e_2$,

$$d(y, c') = d(y - c', \mathbf{0}) = d(e, \mathbf{0}) \leq d(e_2, \mathbf{0}) = d(y - (c' - c_1), \mathbf{0}) = d(y, c' - c_1),$$

therefore,

$$d(y, c') = \min_{c \in \mathcal{C}} d(y, c).$$

□

Since the algorithm for syndrome decoding allows precomputations to choose the coset leader in advance, during the decoding process, we only need to find the right coset, so the search ambient is reduced from q^k elements (the number of codewords) to q^{n-k} elements (the number of cosets). We remark that good codes are expected to have high rates, so $n - k$ tends to be smaller than k. Therefore, syndrome decoding provide, in many cases, an improvement in the performance of an MD decoder. It is not true in general that every MD decoder is obtained by the syndrome decoding algorithm, but any metric (or semimetric) determines some MD decoder that admits a syndrome algorithm.

Example 1.2.23. Let $\mathcal{C}$ be the unidimensional code in $\mathbb{F}_2^2$ generated by 11, so $\mathcal{C} = \{00, 11\}$. The parity check matrix of $\mathcal{C}$ is $H = [1\ 1]$. The vectors 00 and 11 have syndrome equal to 0 and the others have syndrome equal to 1. Let $g : \mathbb{F}_2^2 \to \mathcal{C}$ be a d_H-MD decoder defined by $g(00) = 00$, $g(11) = 11$, $g(10) = 11$ and $g(01) = 11$. There are only two syndrome decoders, one is given by electing 10 as a coset leader and the other is defined by electing 01. These decoders are as follows:

$$g_1'(00) = 00 \quad g_1'(11) = 11 \quad \text{and} \quad g_1'(10) = 00 \quad g_1'(01) = 11$$

and

$$g_2'(00) = 00 \quad g_2'(11) = 11 \quad g_2'(10) = 11 \quad \text{and} \quad g_2'(01) = 00.$$

Neither g_1' nor g_2' coincides with g.

1.3 Matching Metrics and Channels

There are many ways to define a matching between channels and metrics, as we can see in [20], but their purpose are all the same: to give characterizations of metrics which are as close as possible to (optimal) MAP decoders. As seen in Proposition 1.2.10, a decoder determined by the Hamming metric d_H and an ML decoder determined by a q-ary symmetric channel W are matched in the sense that

$$d_H(x, y) \leq d_H(x, z) \iff W(y|x) \geq W(z|x).$$

Unfortunately, cases like this do not always occur, and even when it is possible to match channels and metrics, the metric constructed may be so complex that it is useless for practical purposes.

One of the first papers relating metrics and channels is actually a course note given by Massey [33] apud [28]. Later, Séguin in [42], obtained necessary and sufficient conditions for a discrete memoryless channel to admit an additive metric (metrics determined over the alphabet which are additively extended to vectors) matching to it. The relations between metrics and channels (including the matching problem) were set aside for many years, until renewed interest arose due to new applications, as we can see in [43], [40] and [20]. Since the 1990s, many different families of metrics started to be studied in the context of coding theory, despite the fact that their role, in connection to a channel (or a general communication scheme), is not properly understood. A particular family of such metrics will be explored later. The definition of matching we will adopt is the one given by Séguin in [42].

Definition 1.3.1. Given a discrete channel $W : \mathbb{F}_q^n \to \mathcal{P}_{\mathbb{F}_q^n}$ and a metric $d : \mathbb{F}_q^n \times \mathbb{F}_q^n \to \mathbb{R}$, we say that W and d are *matched* if, for every code $\mathcal{C} \subset \mathbb{F}_q^n$ and every vector $x \in \mathbb{F}_q^n$,

$$\arg\max_{y \in \mathcal{C}} W(x|y) = \arg\min_{y \in \mathcal{C}} d(x, y),$$

or equivalently,

$$d(x, y) < d(x, z) \text{ if and only if } W(x|y) > W(x|z).$$

for every $x, y, z \in \mathbb{F}_q^n$.

To say that W and d are matched, it means there is a decoder that is simultaneously an MD and an ML decoder, i.e., both the probabilistic and metric criteria coincide.

In [19], Firer and Walker proved that the Z channel[2], which is a particular case of an asymmetric channel, also has a metric matched to it, however this is a completely theoretical metric and a priori does not provide a better (less complex) model of decoding. Firer and Walker conjectured the possibility to find metrics matching to every asymmetric channel, but for this general case, the matching problem remains unsolved until now. The reciprocal is always true, as we can see in the next proposition.

Proposition 1.3.2. Given a metric d, there is a discrete channel matching to d.

Proof. Given a metric d, for every $x, y \in \mathbb{F}_q^n$, define a channel $W : \mathbb{F}_q^n \to \mathcal{P}_{\mathbb{F}_q^n}$ by setting

$$W(x|y) := \frac{1/d(x,y)}{M + \sum_{y \neq x} 1/d(x,y)}$$

if $x \neq y$ and

$$W(x|x) := \frac{M}{M + \sum_{y \neq x} 1/d(x,y)}$$

where M is any constant satisfying $M > 1/d(x,y)$ for every $x, y \in \mathbb{F}_q^n$. It is straightforward to conclude that d and W are matched. □

The matching relation is not bijective since it is possible to construct channels which do not admit metrics matching to them. Indeed, if W is a channel such that $W(x|y) < W(x|z)$, $W(z|x) < W(z|y)$ and $W(y|z) < W(y|x)$ for some $x, y, z \in \mathbb{F}_q^n$, then if d is a metric matching to W, due to the symmetry of d, it should satisfy the following relations:

$$d(x,y) > d(x,z) = d(z,x) > d(z,y) = d(y,z) > d(y,x) = d(x,y),$$

which is a contradiction. We stress that if x, y and z are distinct elements, $W(x)$, $W(y)$ and $W(z)$ are probability distributions which are independently defined.

[2] The *asymmetric metric* is commonly used as an MD decoder for the Z channel. Despite the fact that it is not matched to the Z channel, it is used to perform decoding due to its simplicity, details in [9].

For channels that do not admit a metric matching to them, Gabidulin in [20] presented a series of definitions by weakening the definition of matched metrics and channel. One of them introduces an asymptotic definition which characterizes minimum distance decoders that are asymptotically as good as maximum likelihood decoders. Since our goal concerning matching metrics and channels is only to explain the relevance of metrics in this field, we will not go deeper into this subject. To prove the existence of metrics matching to some channels, semimetrics can be used, see [10].

Proposition 1.3.3. [19] If a channel W and a semimetric d' are matched, then, there is a metric d such that d and W are matched.

Chapter 2

Metrics Induced by Partially Ordered Sets

In this chapter, we present the main object of this dissertation: the family of poset metrics. Because these metrics are defined by partially ordered sets, some basic properties of partially ordered sets will be explored in order to state some notations. The first section will be devoted to linear codes and properties, which were superficially sketched in the first chapter when describing the syndrome decoding algorithm. In the following, all metrics will be considered to be defined by a norm and hence to be invariant by translations. As supplementary readings we suggest the books [27] and [34] and the papers [6], [39] and [12].

2.1 Linear Codes

Linear codes over $\mathbb{F}_q$ (finite field with q elements) are the most common and studied type of code in the literature.

Definition 2.1.1. An $[n,k]_q$ *linear code* $\mathcal{C}$ over $\mathbb{F}_q$ is a k-dimensional linear subspace $\mathcal{C} \subset \mathbb{F}_q^n$. The elements of $\mathcal{C}$ will be called codewords.

From here on, we will consider only linear codes, except when otherwise stated. For simplicity in the notations, from now on we will assume that metrics take only natural values, i.e., $d(x,y) \in \mathbb{N}$. Since the space $\mathbb{F}_q^n$ will always be endowed with a metric d, codes may be called d-codes in order to avoid confusion.

Definition 2.1.2. Given a metric d over $\mathbb{F}_q^n$, the *minimum distance* of an $[n,k]_q$ code $\mathcal{C}$ is

$$\delta = \min\{d(x,y) \ : \ x,y \in \mathcal{C} \text{ and } x \neq y\}.$$

We say that $\mathcal{C}$ is an $[n,k,\delta]_q$ code.

Since d is invariant by translations, by Proposition 1.2.19, the minimum distance may be alternatively written as

$$\delta = \min\{w(x) \ : \ x \in \mathcal{C} \setminus \{\mathbf{0}\}\}$$

where $w(x) := d(x,\mathbf{0})$ is a norm function and $\mathbf{0}$ is the null vector.

An $[n,k,\delta]_q$ linear code $\mathcal{C}$ can be represented as the image of an injective linear map $T : \mathbb{F}_q^n \to \mathbb{F}_q^k$ or the kernel of a surjective map $S : \mathbb{F}_q^n \to \mathbb{F}_q^{n-k}$. Considering the canonical basis, we will name G^T and H as the matrices of T and S, respectively. Then, G and H are known respectively as the *generator matrix* and the *parity check matrix* of $\mathcal{C}$. In other words:

Definition 2.1.3. A *generator matrix* G of an $[n,k,\delta]$ code $\mathcal{C}$ is a $k \times n$ matrix, of which the k rows form a basis of $\mathcal{C}$, then

$$\mathcal{C} = \{aG \ : \ a \in \mathbb{F}_q^k\}.$$

Definition 2.1.4. A *parity check matrix* H of an $[n,k,\delta]$ code $\mathcal{C}$ is an $(n-k) \times n$ matrix, satisfying

$$c \in \mathcal{C} \iff Hc^T = \mathbf{0}^T,$$

where $\mathbf{0}$ represents the null vector and x^T is the transpose of the vector x.

Example 2.1.5. (*Hamming Code*) Given k and $n = (q^k-1)/(q-1)$, the $[n, n-k, 3]$ Hamming code over $\mathbb{F}_q$ is a code defined by the parity check matrix that has columns that are pairwise linearly independent (over $\mathbb{F}_q$), i.e., the set of columns is a maximal set of pairwise linearly independent vectors in $\mathbb{F}_q^{n-k}$.

The kernel of an $(n-k) \times n$ parity check matrix is a k-dimensional subspace, therefore the rank of H is $(n-k)$ and all its rows are linearly independent, so it can be seen as a generator matrix of a code.

Definition 2.1.6. If $\mathcal{C}$ is an $[n,k,\delta]$ code with parity check matrix H, the $[n,n-k,\delta_2]$ code generated by H is denoted by $\mathcal{C}^{\perp}$ and called the *dual code* of $\mathcal{C}$. If $\mathcal{C}=\mathcal{C}^{\perp}$, $\mathcal{C}$ is said to be a *self-dual code.*

The *rate* of an $[n,k,\delta]$ linear code is the quotient k/n, and it represents the amount of information contained in each symbol (coordinate). Considering the Hamming metric (Definition 1.2.9), the fundamental problem of coding theory as suggested by Hall in [24], is

Fundamental Problem: *find practical codes with reasonable large rate and minimum distance.*

For a general metric over $\mathbb{F}_q$, the fundamental problem needs to be reformulated by substituting "minimum distance" by "packing radius", the true object to be maximized, as we will see. Given a metric d, the (closed) *d-ball* centered at x with radius r is the set of all elements with a distance of at most r from x,

$$B_d(x,r) := \{y \in \mathbb{F}_q^n \ : \ d(x,y) \leq r\}.$$

Definition 2.1.7. Given a metric d, the *packing radius* of a linear code $\mathcal{C}$ is the maximal integer $\mathcal{R}_d(\mathcal{C})$ satisfying

$$B_d(c_1,\mathcal{R}_d(\mathcal{C})) \cap B_d(c_2,\mathcal{R}_d(\mathcal{C})) = \emptyset$$

for every $c_1, c_2 \in \mathcal{C}$ with $c_1 \neq c_2$. For simplicity, we may suppress the explicit dependence on d in the notation $\mathcal{R}_d(\mathcal{C})$.

The following is a well-known standard result in coding theory.

Proposition 2.1.8. Let d_H be the Hamming metric. The packing radius of a linear code $\mathcal{C}$ is given by

$$\mathcal{R}_{d_H}(\mathcal{C}) = \left\lfloor \frac{\delta-1}{2} \right\rfloor$$

where $\lfloor a \rfloor$ denotes the integer part of the real number a and δ is the minimum distance of $\mathcal{C}$.

The importance of the packing radius in coding theory is due to the fact that it determines the code *error-correcting capability*, indeed, considering that a codeword c

is sent over a noisy channel and y is received. If $d(c,y) \leq \mathcal{R}_d(\mathcal{C})$, then the received vector y will still be closer to c than to any other codeword, therefore a minimum distance decoder will output c. On the other hand, if $d(c,y) > \mathcal{R}_d(\mathcal{C})$ it is not guaranteed that a minimum distance decoder will output c. For the Hamming metric, the packing radius of a code is determined by its minimum distance (Proposition 2.1.8), therefore the minimum distance also determines the code error-correcting capability, justifying the description of the fundamental problem of coding given before. As we will see later, there are metrics for which the minimum distance of a code does not determine its packing radius. In order to include this kind of metric, the fundamental problem is restated as follows:

General Fundamental Problem: *find practical codes with reasonable large rate and packing radius.*

The *error-detection capability* of an $[n,k,\delta]$ code is $\delta - 1$ since for every received element y, if $d(c,y) \leq \delta - 1$, by definition of the minimum distance, y will never be a codeword different from c. On the other hand, if $d(c,y) > \delta - 1$ it may happen that $y \in \mathcal{C}$ and is undetectable. Even when the minimum distance of a code does not determine its packing radius, it is relevant in coding theory since it always determines the code error-detection capability and it provides bounds for the packing radius, as we will see in the next proposition.

Proposition 2.1.9. Let $\mathcal{C}$ be an $[n,k,\delta]$ d-code over $\mathbb{F}_q$, then

$$\left\lfloor \frac{\delta - 1}{2} \right\rfloor \leq \mathcal{R}_d(\mathcal{C}) \leq \delta - 1.$$

Proof. The minimum distance definition ensures that $\mathcal{R}_d(\mathcal{C}) \leq \delta - 1$. Denote $t = \lfloor (\delta-1)/2 \rfloor$, we just need to prove that $B_d(u,t) \cap B_d(v,t) = \emptyset$ for all $u, v \in \mathcal{C}$ with $u \neq v$. Suppose $x \in B_d(u,t) \cap B_d(v,t)$. By the triangular inequality,

$$d(u,v) \leq d(u,x) + d(x,v) \leq 2t \leq \delta - 1,$$

a contradiction since $d(u,v) \geq \delta$. □

As we will see in the poset metrics section, it is possible to construct metrics and codes such that their packing radius reaches the extremal values of the bounds obtained in Proposition 2.1.9. Furthermore, in some cases, those bounds may be attained by codes with the same minimum distance.

The packing radius $\mathcal{R}_d(c)$ of a non-null codeword $c \in \mathcal{C}$ is defined by being the packing radius of the (not necessary linear) code $\{\mathbf{0}, c\}$. The packing radius of a code $\mathcal{C}$ can be alternatively defined as the minimal packing radius of its non-null codewords, i.e.,

$$\mathcal{R}_d(\mathcal{C}) = \min_{c \in \mathcal{C}^*} \mathcal{R}_d(c), \tag{2.1}$$

where $\mathcal{C}^* = \mathcal{C} \setminus \{\mathbf{0}\}$. A codeword with minimum packing radius is called a *packing vector*. We stress that the problem to find the packing radius of a single codeword for general metrics is an NP-hard problem, see [12].

2.1.1 Code Equivalence

In order to investigate the fundamental problem of coding, codes can be gathered in classes such that elements in the same class are “geometrically equivalent”. In particular, we are interested in classes of codes having the same error-correcting capability. A simple way to construct these classes is by using linear isometries. As an example, in the binary Hamming case, two codes belong to the same class if one is a permutation of the other, see [26].

Definition 2.1.10. If $\mathbb{F}_q^n$ is a metric space endowed with the metric d, a *linear isometry* (or d-linear isometry) T is a linear transformation $T : \mathbb{F}_q^n \to \mathbb{F}_q^n$ preserving distance, i.e., for every $x, y \in \mathbb{F}_q^n$,

$$d(T(x), T(y)) = d(x, y).$$

Since we are assuming d is invariant by translations, $d(x, y) = w(x - y)$, then a linear transformation T is an isometry if, and only if, $w(T(x)) = w(x)$ for every $x \in \mathbb{F}_q^n$. We denote by $GL_d(\mathbb{F}_q^n)$ the group of all linear isometries of $\mathbb{F}_q^n$.

Example 2.1.11. [31] The linear isometry group of $\mathbb{F}_q^n$ when $\mathbb{F}_q^n$ is endowed with the Hamming metric d_H is the group of all monomial maps, i.e.,

$$GL_{d_H}(\mathbb{F}_q^n) \simeq \mathbb{F}_q^{*n} \rtimes \mathcal{S}_n$$

where:

- $\mathbb{F}_q^{*n}$ is isomorphic to the the group of all $n \times n$ invertible diagonal matrices. It acts on $\mathbb{F}_q^n$ by multiplying each coordinate by a non-zero constant.
- $\mathcal{S}_n$ is isomorphic to the group of all permutation matrices and corresponds to the permutations of coordinates in the space.
- the product is semi-direct.

Definition 2.1.12. Two linear codes $\mathcal{C}_1$ and $\mathcal{C}_2$ are said to be *equivalent* if there is a linear isometry $T : \mathbb{F}_q^n \to \mathbb{F}_q^n$ such that $T(\mathcal{C}_1) = \mathcal{C}_2$. We denote by $\mathcal{C}_1 \sim_d \mathcal{C}_2$.

With this definition, considering Example 2.1.11 we obtain, for the Hamming metric case, the usual definition of code equivalence. Equivalence of codes defines an equivalence relation which gather linear codes in classes of codes having same geometrical properties, in particular having the same error-correction capability, however it is not true that codes with the same weight distribution (same number of codewords having weight i, for every i) belong to the same class.

Example 2.1.13. [27] Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be binary codes with generator matrices

$$G_1 = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \quad \text{and} \quad G_2 = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix},$$

respectively. Suppose $\mathbb{F}_q^n$ is endowed with the Hamming metric. If $A_i(\mathcal{C}_j)$ is the number of codewords in $\mathcal{C}_j$ with weight i, we have that $A_0(\mathcal{C}_j) = A_6(\mathcal{C}_j) = 1$ and $A_2(\mathcal{C}_j) = A_4(\mathcal{C}_j) = 3$ for $j \in \{1, 2\}$. Thus, the codes $\mathcal{C}_1$ and $\mathcal{C}_2$ have the same weight distribution consequently, the same minimum distance $\delta = 2$ and the same packing radius $\mathcal{R}_{d_H}(\mathcal{C}_1) = \mathcal{R}_{d_H}(\mathcal{C}_2) = 0$. Since the basis elements of $\mathcal{C}_1$ are also elements of $\mathcal{C}_1^\perp$ and these two codes have same dimension, $\mathcal{C}_1$ is self-dual. But $\mathcal{C}_2$ is not self-dual since the vector 000011 is a codeword of $\mathcal{C}_2^\perp$ but not of $\mathcal{C}_2$. If $\mathcal{C}_1$ and $\mathcal{C}_2$ were equivalent, there should exist a permutation matrix P satisfying $\mathcal{C}_1 P = \mathcal{C}_2$, but this would imply that $\mathcal{C}_1^\perp P = \mathcal{C}_2^\perp$ and since $\mathcal{C}_1 = \mathcal{C}_1^\perp$, $\mathcal{C}_2$ would be self-dual. Therefore, $\mathcal{C}_1$ and $\mathcal{C}_2$ are not equivalent.

Given a linear code $\mathcal{C}$, we denote by $GL_d(\mathcal{C})$ its orbit under $GL_d(\mathbb{F}_q^n)$. Since $GL_d(\mathbb{F}_q^n)$ is a group, their orbits are equivalence classes, hence $\mathcal{C} \sim_d \mathcal{C}'$ if, and only if, $GL_d(\mathcal{C}) = GL_d(\mathcal{C}')$. The representatives of each class may be chosen according to codes having a generator matrix with a particular form, which is, in general, as simple as possible in order to provide some information about the code.

Proposition 2.1.14. Every $[n, k, \delta]$ linear d_H-code $\mathcal{C}$ is equivalent to a linear d_H-code with the same parameters having a generator matrix of the form $G = [I_k \mid A]$, where I_k is the identity matrix with order k and A is a $k \times (n-k)$ matrix.

This form is possible since permutations and non-null scalar multiplications in the columns of the generator matrix are performed by linear isometries of the Hamming metric. Due to Proposition 2.1.14, it is common to find in the literature the assumption that every code has a generator matrix in the *standard form* $G = [I_k \mid A]$. As consequence of the standard form, we get an easy way to obtain the parity check matrix of a d_H-code.

Theorem 2.1.15. If $G = [I_k \mid A]$ is a generator matrix for an $[n, k, \delta]$ code $\mathcal{C}$, then $H = [-A^T \mid I_{n-k}]$ is a parity check matrix for $\mathcal{C}$.

2.2 Partially Ordered Sets

Partial orders will be the main mathematical structures used in this work to define poset metrics. The definitions in this section are mainly to fix notations for the development of the next sections. As complementary readings, the book [34] and the papers [5] and [14] are indicated. Let X and Y be finite non-empty sets. A binary relation over X and Y is any subset R of the product $X \times Y$. If $(x, y) \in R$, we write xRy. If $X = Y$, we say that R is a binary relation over X.

Definition 2.2.1. A *partial order relation* in a set X is a binary relation, usually denoted by $\leqslant$, satisfying, for every $x, y, z \in X$, the following conditions:

(a) $x \leqslant x$ (reflexivity);

(b) If $x \leqslant y$ and $y \leqslant x$, then $x = y$ (anti-symmetry);

(c) If $x \leqslant y$ and $y \leqslant z$, then $x \leqslant z$ (transitivity).

If $\leqslant$ is a partial order relation over X, the pair $P = (X, \leqslant)$ is called *poset*.

Eventually, the binary relation $\leqslant$ of the poset $P = (X, \leqslant)$ is denoted by $\leqslant_P$. By an abuse of notation, the poset P will be identified with X. Therefore, elements of

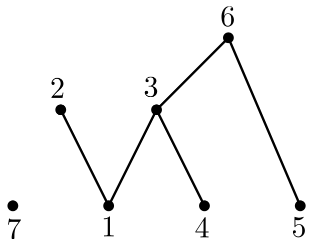
X will be considered to be also elements of P and the order relation over X will also be regarded to be an order relation over P.

If any two elements of a partial order are comparable, the order is named *total* and the poset is called a *chain*. An *anti-chain* is a poset where distinct elements are never comparable. We say that y *covers* x if $x \leqslant y$ and there is no extra element z such that $x \leqslant z$ and $z \leqslant y$. If y covers x, the pair (x, y) is said a *covering pair*. In order to geometrically describe a poset, the Hasse diagram is used.

Definition 2.2.2. The *Hasse diagram* of a poset $P = (X, \leqslant)$ is the directed graph in which the vertex set is X and whose arcs are the covering pairs (x, y) in the poset.

We usually draw the Hasse diagram of a poset in the plane in such a way that, if y covers x, then the point representing y is higher than the point representing x. No arrows are required in the drawing, since the directions of the arrows are implicit downward.

Example 2.2.3. Let $X = \{1, 2, 3, 4, 5, 6, 7\}$ and consider the following partial order relation: $1 \leqslant 2$, $1 \leqslant 3$, $4 \leqslant 3$, $3 \leqslant 6$ and $5 \leqslant 6$. The Hasse diagram of $P = (X, \leqslant)$ is given by



Definition 2.2.4. An *ideal* in a poset P is a nonempty subset $I \subset X$ such that, for $i \in I$ and $j \in X$, if $j \leqslant_P i$ then $j \in I$.

Given $A \subset X$, we denote by $\langle A \rangle_P$ the smaller ideal of P containing A. If $A = \{i\}$, we denote by $\langle i \rangle_P$ the ideal $\langle \{i\} \rangle_P$. The set $Max_P(A)$ is the set of all maximal elements of A in P, equivalently,

$$Max_P(A) = \{i \in A \ : \ i \nleqslant_P j \text{ for all } j \in A \setminus \{i\}\}.$$

The *rank* of an element $j \in X$, denoted by $h_P(j)$, is the maximal cardinality of a chain contained in $\langle j \rangle$,

$$h_P(j) = max\{|C| : \ C \subset \langle j \rangle_P \text{ and } C \text{ is a chain}\}.$$

The *height* $h(P)$ *of* P is the maximal rank among the elements of X. The *i-level* of P is the set of all elements with rank i, $\Gamma_P^i := \{j \in X : h_P(j) = i\}$. The level distribution of a poset P is the vector $(\Gamma_P^1, \ldots, \Gamma_P^{h(P)})$. This distribution defines a partition of X in the sense that $X = \bigsqcup \Gamma_P^i$. Also, since the levels are disjoint, if $|X|= n$ and $|\Gamma_P^i|= n_i$, then $n = n_1 + \cdots + n_{h(P)}$. The *level enumerator* of a poset is the array $(|\Gamma_P^1|, \ldots, |\Gamma_P^{h(P)}|)$.

Example 2.2.5. The level distribution of the poset defined in Example 2.2.3 is the vector with length 3 whose coordinates are given by

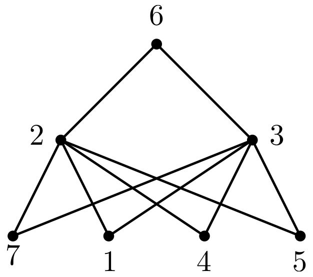
$$\Gamma_P^1 = \{1, 4, 5, 7\}, \ \ \Gamma_P^2 = \{2, 3\} \ \text{ and } \ \Gamma_P^3 = \{6\}.$$

Furthermore, $\langle 6 \rangle = \{6, 3, 1, 4, 5\}$ and $\langle \{2, 7\} \rangle = \{2, 7, 1\}$.

In coding theory, an important family of posets is the family of Hierarchical posets, as we shall see, the metrics obtained by these posets can be considered as a "true" generalization of the Hamming metric, see [30] for more details concerning this relation.

Definition 2.2.6. Given a poset P, if $x \leqslant_P y$ for every $x \in \Gamma_P^i$ and $y \in \Gamma_P^j$ with $i < j$, then P is called a *hierarchical poset*.

Example 2.2.7. The poset P constructed in Example 2.2.3 is not hierarchical since 7 and 2 belong to different levels and are not comparable to each other. By suitably adding relations until we can obtain a hierarchical poset. The resulting poset has the following Hasse diagram:



Posets can be characterized by their Hasse diagrams. This characterization is obtained by using the equivalence relation defined by isomorphism of posets. This relation ensures that posets with the same unlabeled Hasse diagram are essentially the same poset.

Definition 2.2.8. Let $P = (X, \leqslant_P)$ and $Q = (Y, \leqslant_Q)$ be two partially ordered sets. A map $f : X \to Y$ is called *order-preserving* if $x \leqslant_P y$ implies $f(x) \leqslant_Q f(y)$.

Definition 2.2.9. Let $P = (X, \leqslant_P)$ and $Q = (Y, \leqslant_Q)$ be two posets such that $Y \subset X$. The set Q is a *subposet* of P if the identity map $I : Y \to X$ is an order-preserving map.

The example below shows that a bijective and order-preserving map not always has an inverse which also preserve order.

Example 2.2.10. Let P be an anti-chain over $[n]$ and Q any other poset over $[n]$. Then, any bijective map from P to Q is an order-preserving map with an inverse that does not preserve order.

Definition 2.2.11. A one-to-one order-preserving map f from a poset $(X, \leqslant_P)$ onto a poset $(Y, \leqslant_Q)$ is called an *isomorphism* if the inverse f^{-1} is also an order-preserving mapping. An isomorphism from a poset to itself is called an *automorphism.*

Whenever two posets are order isomorphic, they can be considered to be essentially the same in the sense that one of the orders can be obtained from the other just by renaming of elements.

Proposition 2.2.12. [34] Two hierarchical posets are isomorphic if, and only if, they have the same level enumerator.

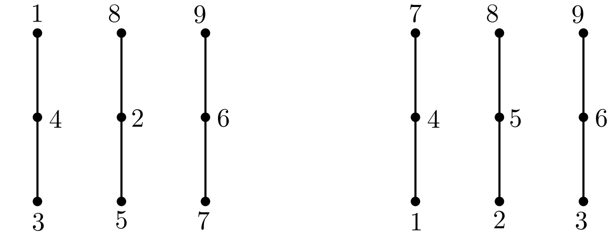
Due to Proposition 2.2.12, if P is hierarchical, it will be denoted by $(n : n_1, \ldots, n_s)$ where $n_j = |\Gamma_P^j|$ and $s = h(P)$. The set of all automorphisms of a poset P is a group which will be denoted by $Aut(P)$.

Let $\mathcal{P}_n^*$ the set of all posets over $[n]$. The set $\mathcal{P}_n^*$ has a natural partial order: given two posets $P, Q \in \mathcal{P}_n^*$, we say that P is finer (or smaller) than Q (and write $P \leq Q$) if $i \leqslant_P j$ implies $i \leqslant_Q j$. Equivalently, $P \leq Q$ if, and only if, the identity map is an order-preserving map from P to Q. With this relation, the set $\mathcal{P}_n^*$ is itself a partially ordered set. The trivial order (anti-chain: $i \leqslant j \iff i = j$) is the (unique) minimal element in $\mathcal{P}_n^*$ and the linear order (chain: $1 \leqslant 2 \leqslant \cdots \leqslant n$), as much as its $n!$ permutations, are the maximal elements in $\mathcal{P}_n^*$.

A poset $P \in \mathcal{P}_n^*$ is said to be *naturally labeled* if for every $i \in \Gamma_P^{r_1}$ and $j \in \Gamma_P^{r_2}$ with $r_1 < r_2$, then $i < j$ (where $<$ is the natural order over $\mathbb{N}$). Therefore, if the set $\mathcal{P}_n$ denote the set of all naturally labeled posets,

$$\mathcal{P}_n = \{P \in \mathcal{P}_n^* \ : \ P \text{ is naturally labeled}\}$$

$\mathcal{P}_n$ has a unique maximal element, the linear order defined by $1 \leqslant 2 \leqslant \cdots \leqslant n$.



non-natural labelling natural labelling

Since given a poset $P \in \mathcal{P}_n^*$ there is always a poset $Q \in \mathcal{P}_n$ order isomorphic to P, from now on, assume all the posets to be naturally labeled.

2.3 Poset Metrics

Classical coding theory may be considered as the study of $\mathbb{F}_q^n$ when it is endowed with the Hamming metric. To generalize the classical problems in coding theory, Niederreiter made the initial progress in [37], [35] and [36] by introducing non-Hamming metrics in $\mathbb{F}_q^n$. Generalizing the metric introduced by Niederreiter, in [6], Brualdi et al. introduced a new large family of metrics, the so-called *poset metrics*. In the following, consider P to be a poset over $[n]$.

Definition 2.3.1. The *P-weight* of a vector $x \in \mathbb{F}_q^n$ is defined as the cardinality of the smallest ideal of P containing *supp*(x), i.e.,

$$w_P(x) = |\langle supp(x)\rangle_P|.$$

It is clear that $w_P(x) \geq 0$ for every $x \in \mathbb{F}_q^n$ and $w_P(x) = 0$ if and only if $x = \mathbf{0}$. Also, the relations $supp(x + y) \subset supp(x) \cup supp(y)$ and $\langle A \cup B\rangle_P = \langle A\rangle_P \cup \langle B\rangle_P$ imply that $w_P(x + y) \leq w_P(x) + w_P(y)$. Therefore, w_P is a norm function over $\mathbb{F}_q^n$.

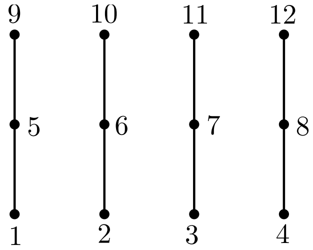
The support of a set $X \subset \mathbb{F}_q^n$ is the union of the supports of the elements of X,

$$supp(X) = \{i : i \in supp(x) \text{ for some } x \in X\}.$$

Definition 2.3.2. The *P-distance* in $\mathbb{F}_q^n$ is the (invariant by translation) metric induced by w_P,

$$d_P(x, y) = w_P(x - y).$$

Example 2.3.3 (*Niederreiter-Rosenbloom-Tsfasman metric - NRT*)**.** Given two integers n and r such that r divides n, an (n, r)-NRT metric is a poset metric induced by a poset formed by n/r disjoint chains, each chain having length r. Suppose $n = 12$ and $r = 3$. Consider the NRT weight w_{NRT} defined by the following Hasse diagram:



In particular, the (n, n)-NRT and the $(n, 1)$-NRT metrics are the chain and the anti-chain (or the Hamming) metrics.

2.3.1 Group of Linear Isometries for Poset Metrics

As seen in section 2.1.1, two linear codes are equivalent if, and only if, there is a linear isometry mapping one into the other. For a poset metric, the group of linear isometries was first determined for the Rosenbloom–Tsfasman space in [8] and for the crown space in [29]. The description of this group for a general poset metric was presented in [39]. We will describe it in some details since both the result and the approach used in the proof will be used in Chapter 3. The proof presented here is slightly different but follows the same idea of the proof given in [39]. Let $M_n(\mathbb{F}_q)$ be the set of all $n \times n$ matrices over $\mathbb{F}_q$ and

$$G_P := \{A = (a_{ij}) \in M_n(\mathbb{F}_q) \ : \ a_{ij} = 0 \text{ if } i \nleq j \text{ and } a_{ii} \neq 0\}. \tag{2.2}$$

Considering the usual basis $\beta = \{e_1, \ldots, e_n\}$ where $supp(e_i) = \{i\}$ and the i-th entry is 1. Each matrix $A \in G_P$ defines a linear map T_A. The set of such maps will be denoted by $\mathcal{G}_P$. By definition, $T \in \mathcal{G}_P$ if, and only if, $T(e_j) = \sum_{i \leqslant_P j} a_{ij} e_i$ with $a_{jj} \neq 0$ for every $j \in [n]$.

Note initially that an automorphism $\phi \in Aut(P)$ induces an isometry $T_\phi \in GL_P(\mathbb{F}_q^n)$, acting on $\mathbb{F}_q^n$ by permutation of the coordinates: $T_\phi(x_1, \ldots, x_n) = (x_{\phi(1)}, \ldots, x_{\phi(n)})$. The set of these isometries will be denoted by $\mathcal{A}ut(P)$.

Theorem 2.3.4. The group of isometries of $\mathbb{F}_q^n$ is the semi-direct product $GL_P(\mathbb{F}_q^n) = \mathcal{G}_P \rtimes \mathcal{A}ut(P)$.

The next two lemmas will be used in order to produce maximal decompositions in Chapter 3, a generalization of the canonical decomposition obtained in [15]. They are also used to prove Theorem 2.3.4, their proofs can be found in [39].

Lemma 2.3.5. If $T \in GL_P(\mathbb{F}_q^n)$, then the map $\phi_T : P \to P$ given by

$$\phi_T(i) = max\langle supp(T(e_i))\rangle_P$$

is an automorphism of P.

Lemma 2.3.6. The linear transformation T is an element of $GL_P(\mathbb{F}_q^n)$ if, and only if,

$$T(e_j) = \sum_{i \leqslant_P j} x_{ij} e_{\phi_T(i)}, \tag{2.3}$$

where ϕ_T is the automorphism associated with T as in Lemma 2.3.5 and x_{ij} are constants with $x_{jj} \neq 0$ for all $j \in [n]$.

Proof of Theorem 2.3.4. It is clear that $\mathcal{A}ut(P)$ is a subgroup of $GL_P(\mathbb{F}_q^n)$ and the characterization of $\mathcal{G}_P$ together with Lemma 2.3.6 ensures that $\mathcal{G}_P$ is also a subgroup of $GL_P(\mathbb{F}_q^n)$. Given $T \in GL_P(\mathbb{F}_q^n)$, by Lemma 2.3.6, $T(e_j) = \sum_{i \leqslant j} x_{ij} e_{\phi_T(i)}$ and $x_{jj} \neq 0$. Consider T' defined by $T'(e_i) = e_{\phi_T(i)}$, by Lemma 2.3.6, $T' \in GL_P(\mathbb{F}_q^n)$. Define T'' by setting $T''(e_j) = \sum_{i \leqslant j} x_{ij} e_i$, thus $T(e_j) = T'' \circ T'(e_j)$. Since T' is obtained by the automorphism ϕ_T, it follows that $T' \in \mathcal{A}ut(P)$, furthermore, by construction, $T'' \in \mathcal{G}_P$. Therefore, $GL_P(\mathbb{F}_q^n) = \mathcal{G}_P \cdot \mathcal{A}ut(P)$. In order to prove that $\mathcal{G}_P$ is a normal subgroup of $GL_P(\mathbb{F}_q^n)$, it is enough to show that $T' \circ T \circ T'^{-1} \in \mathcal{G}_P$ for every $T \in \mathcal{G}_P$ and $T' \in GL_P(\mathbb{F}_q^n)$. Since $T' = T'' \circ T_2$, where $T'' \in \mathcal{G}_P$ and $T_2 \in \mathcal{A}ut(P)$, it is sufficient to prove that $T_2 \circ T \circ T_2^{-1} \in \mathcal{G}_P$ for every $T_2 \in \mathcal{A}ut(P)$. First, note that $T_2 = T_\phi$ for some $\phi \in Aut(P)$

and that $T_{\phi}^{-1} = T_{\phi^{-1}}$, thus

$$T_{\phi} \circ T \circ T_{\phi^{-1}}(e_j) = T_{\phi} \circ T(e_{\phi^{-1}(j)}) = T_{\phi}\left(\sum_{i \leqslant \phi^{-1}(j)} x_{i\phi^{-1}(j)} e_i\right) = \sum_{i \leqslant \phi^{-1}(j)} x_{i\phi^{-1}(j)} e_{\phi(i)}.$$

Since $i \leqslant \phi^{-1}(j)$ implies $\phi(i) \leqslant j$, denoting $b_{\phi(i)j} = x_{i\phi^{-1}(j)}$, it follows that

$$T_{\phi} \circ T \circ T_{\phi^{-1}}(e_j) = \sum_{i \,:\, \phi(i) \leqslant j} b_{\phi(i)j} e_{\phi(i)}.$$

Therefore $T_{\phi} \circ T \circ T_{\phi^{-1}} \in \mathcal{G}_P$ and $\mathcal{G}_P$ is a normal subgroup of $GL_P(\mathbb{F}_q^n)$. By using the characterizations of $\mathcal{G}_P$ and $\mathcal{A}ut(P)$, it is straightforward to conclude that $\mathcal{G}_P \cap \mathcal{A}ut(P) = \{I\}$ where I is the identity map, therefore $GL_P(\mathbb{F}_q^n)$ is isomorphic to the semidirect product $\mathcal{G}_P \rtimes \mathcal{A}ut(P)$. □

Example 2.3.7. (*Hierarchical Case*) We remark that if $T \in \mathcal{A}ut(P)$, then T is induced by a permutation $\phi : P \to P$ and denoted by T_{ϕ}. If P is an $(n : n_1, \ldots, n_s)$ hierarchical poset, the image by ϕ of an element in the i-th level must also belong to this level. For each i, we denote the group $\mathcal{A}ut(\Gamma_P^i)$ by the group of all linear maps induced by permutations ϕ_i that permutes only elements of the i-th level of P, i.e., $\phi_i : P \to P$ is a bijection satisfying $\phi_i(j) = j$ if $j \notin \Gamma_P^i$. Since P is hierarchical, $\mathcal{A}ut(\Gamma_P^i) \subset \mathcal{A}ut(P)$. Hence, each ϕ_i induces an isometry T_{ϕ_i} thus $\phi = \phi_1 \circ \ldots \circ \phi_s$, i.e., $T_{\phi} = T_{\phi_1} \circ \ldots \circ T_{\phi_s}$. Therefore, the group $\mathcal{A}ut(P)$ is isomorphic to the product $\mathcal{A}ut(\Gamma_P^1) \times \ldots \times \mathcal{A}ut(\Gamma_P^s)$ and $T \in \mathcal{A}ut(P)$ if, and only if,

$$T_{\phi}(x_1, \ldots, x_n) = (x_{\phi_1(1)}, \ldots, x_{\phi_1(n_1)}) \times \cdots \times (x_{\phi_s(n_1+\ldots+n_{s-1}+1)}, \ldots, x_{\phi_s(n_1+\ldots+n_{s-1}+n_s)})$$

where $\phi_i \in \mathcal{A}ut(\Gamma_P^i)$ for every $i \in \{1, \ldots, s\}$.

2.3.2 Packing Radius of Poset Codes

Since poset metrics are defined by weights, they are invariant by translations and the *packing radius* of a poset code can be equivalently defined as being the largest integer $\mathcal{R}_P(\mathcal{C})$ satisfying

$$B_P(\mathbf{0}, \mathcal{R}_P(\mathcal{C})) \cap B_P(c, \mathcal{R}_P(\mathcal{C})) = \emptyset$$

for every $c \in \mathcal{C}$.

In order to simplify the notation, the set of maximal elements in the support of a codeword c, which is denoted by $Max_P(supp(c))$, will be denoted by $Max_P(c)$.

Proposition 2.3.8. Given an $[n,k,\delta]_q$ poset code $\mathcal{C}$ such that $|Max_P(c)|= 1$ for every $c \in \mathcal{C} \setminus \{\mathbf{0}\}$, then

$$\mathcal{R}_P(\mathcal{C}) = \delta - 1.$$

Proof. Suppose $Max_P(c) = \{i\}$, note that if $x \in B_P(c, w_P(c) - 1)$ then $x_i = c_i$. Since $i \in supp_P(x)$, it follows that $x \notin B_P(\mathbf{0}, w_P(c) - 1)$. Hence, $R_P(c) = w_P(c) - 1$. Taking c with minimum weight δ, the packing radius of this codeword is $\delta - 1$, by characterization 2.1, c is a packing vector, therefore, $\mathcal{R}_P(\mathcal{C}) = \delta - 1$. □

The proof of the next corollary follows straight from the fact that in a chain, every non-null vector has only one maximal element in its support.

Corollary 2.3.9. Let P be a chain. If $\mathcal{C}$ is an $[n,k,\delta]_q$ P-code, then

$$\mathcal{R}_P(\mathcal{C}) = \delta - 1.$$

Corollary 2.3.9 and Proposition 2.1.8 ensure that the bounds given in Proposition 2.1.9 are tight. The extremal values for the bound are obtained by extremal posets, one with no relations (anti-chain) and the other with the maximum number of relations (total order - chain). Both are hierarchical posets and the metrics induced by these posets are very well understood, as can be seen in [30]. A hierarchical poset can be obtained by summing up anti-chains, as follows:

Definition 2.3.10. Let $P = (X, \leqslant_P)$ and $Q = (Y, \leqslant_Q)$ be two posets with $X \cap Y = \emptyset$. The *ordinal sum* of P and Q is the poset $P \oplus Q$ with order relation given by

$$x \leqslant_{P\oplus Q} y \iff \begin{cases} x \leqslant_P y & \text{when } x, y \in X \\ x \leqslant_Q y & \text{when } x, y \in Y \\ x \in X \text{ and } y \in Y & \end{cases} .$$

Proposition 2.3.11. A hierarchical poset $(n : n_1, \ldots, n_l)$ is an ordinal sum of l anti-chains.

It is straightforward to conclude that the Hamming metric is a poset metric induced by an anti-chain. According to Proposition 2.3.11, hierarchical posets are characterized as ordinal sums of anti-chains, therefore, hierarchical metrics are closely related to the Hamming metric. We are particularly interested in the characterization of poset metrics for which the error-correction capability of a code is determined by its error-detection capability, i.e., poset metrics for which the packing radius of a code is determined by its minimum distance.

Proposition 2.3.12. If $\mathcal{C}$ is an $[n,k,\delta]_q$ P-code such that $\langle supp(\mathcal{C})\rangle_P$ is a hierarchical subposet of P, then

$$\mathcal{R}_P(\mathcal{C}) = n_1 + \cdots + n_{r-1} + \left\lfloor \frac{\delta - (n_1 + \cdots + n_{r-1}) - 1}{2} \right\rfloor$$

where r is the smallest level of P such that there exists $c \in \mathcal{C}$ with $Max_P(c) \subset \Gamma_P^r$ and $n_s = |\langle supp(\mathcal{C})\rangle_P \cap \Gamma_P^s|$.

Proof. Set $s_1 = 0$ and $s_i = n_1 + \cdots + n_{i-1}$ for all $i \in \{2, \ldots, h(P)\}$. Consider $h = s_r + \lfloor(\delta - s_r - 1)/2\rfloor$ where

$$r = \min\{i \ : \ Max_P(c) \subset \Gamma_P^i \text{ for some } c \in \mathcal{C}\}.$$

Suppose there exists $z \in B_P(\mathbf{0}, h) \cap B_P(c, h)$ for some $c \in \mathcal{C}$. Since $\langle supp(\mathcal{C})\rangle_P$ is hierarchical, $Max_P(c) \subset \Gamma_P^{j_0}$ for some level j_0 and by the minimality of r, it follows that $j_0 \geq r$. The weight of c is of the form $w_P(c) = s_{j_0} + t$ where $t = |Max_P(c)|$. It is straightforward that $Max_P(z-c) \subset \Gamma_P^{j_0}$. Thus, if

$$|Max_P(z-c)| > \left\lfloor \frac{t-1}{2} \right\rfloor,$$

then

$$d(z,c) = s_{j_0} + |Max_P(z-c)| > s_{j_0} + \left\lfloor \frac{t-1}{2} \right\rfloor \geq h,$$

i.e., $z \notin B_P(c,h)$. On the other hand, if

$$|Max_P(z-c)| \leq \left\lfloor \frac{t-1}{2} \right\rfloor,$$

then

$$2|Max_P(z-c)|< t = |Max(c)|.$$

Therefore, $|Max_P(z)|> |Max_P(c)|$, hence

$$d(z, \mathbf{0}) = s_{j_0} + |Max_P(z)|> s_{j_0} + |Max_P(c)|\geq h,$$

i.e., $z \notin B_P(\mathbf{0}, h)$. To conclude, we just need to prove that there exists $c \in \mathcal{C}$ such that $B_P(0, h+1) \cap B_P(c, h+1) \neq \emptyset$. Suppose $w_P(c) = \delta$, note that $|Max_P(c)|= \delta - s_r$, take a set $A \subset Max_P(c)$ such that $|A|= \lfloor(\delta - s_r - 1)/2\rfloor + 1$, define $z \in \mathbb{F}_q^n$ by the rule $z_i = c_i$ for every $i \in A$ and $z_i = 0$ otherwise, then

$$d_P(z, \mathbf{0}) = |A|+s_r = h+1.$$

Since $|Max_P(c)\backslash A|\leq |A|$, we also have that

$$d_P(z, c) = s_r + |Max_P(c)\backslash A|\leq s_r + |A|= h+1.$$

Thus, $z \in B_P(\mathbf{0}, h+1) \cap B_P(c, h+1)$, therefore, $\mathcal{R}_P(\mathcal{C}) = h$.

□

We stress that P is hierarchical if, and only if, every ideal of P is also hierarchical and Proposition 2.3.12 holds for every code.

Theorem 2.3.13. The poset P is a hierarchical poset if and only if any two P-codes with same minimum distance also have the same packing radius.

Proof. Proposition 2.3.12 ensures that if P is hierarchical any two codes having the same minimum distance also have the same packing radius. Conversely, suppose P is not hierarchical, then there exist $i_0 \in \Gamma_P^r$ and $j_0 \in \Gamma_P^{r+1}$ such that $i_0 \not\leq j_0$. Let r be minimal with this property. Consider the one-dimensional code $\mathcal{C}_1$ generated by the vector e_{j_0} and the one-dimensional code $\mathcal{C}_2$ generated by the vector v where $v_s = 1$ if $s \in A$, $v_{i_0} = 1$ and $v_s = 0$ for all $s \in [n] \setminus (A \cup \{i_0\})$ where $A = \Gamma_P^r \cap \langle j_0 \rangle$. By construction, $\mathcal{C}_1$ and $\mathcal{C}_2$ have the same minimum distance $\delta = |\langle j_0 \rangle_P|$. By Proposition 2.3.8, $\mathcal{R}_P(\mathcal{C}_1) = \delta - 1$ and by

Proposition 2.3.12,

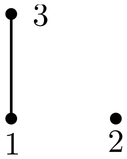
$$\mathcal{R}_P(\mathcal{C}_2) = n_1 + \cdots + n_{r-1} + \left\lfloor \frac{\delta - (n_1 + \cdots + n_{r-1}) - 1}{2} \right\rfloor ,$$

but

$$\begin{aligned} \mathcal{R}_P(\mathcal{C}_1) =& \delta - 1 = n_1 + \cdots + n_{r-1} + \delta - (n_1 + \cdots + n_{r-1}) - 1 \qquad (2.4) \\ >& n_1 + \cdots + n_{r-1} + \left\lfloor \frac{\delta - (n_1 + \cdots + n_{r-1}) - 1}{2} \right\rfloor = \mathcal{R}_P(\mathcal{C}_2). \qquad (2.5) \end{aligned}$$

□

Example 2.3.14. Consider the non-hierarchical poset P over $[3]$ with an order relation given by $1 \leqslant 3$. Its Hasse diagram is as follows:



Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be two one-dimensional linear codes over $\mathbb{F}_q^n$ generated by e_3 and $e_1 + e_2$, respectively. It is straightforward to conclude that both codes have minimum distance $(\delta = 2)$ and that their packing radius are different, namely, $\mathcal{R}_P(\mathcal{C}_1) = 1$ and $\mathcal{R}_P(\mathcal{C}_2) = 0$.

We conclude that among the poset metrics, hierarchical metrics are the only ones where the error-correcting capability of a code is determined by its error-detection capability. We remark that, as proved in [12], to determine the error-correction capability of a single pair of codewords is an NP-hard problem, so bounds for these invariants are very welcome.

Chapter 3

Canonical Form

It is well known that representation of a code by generator matrices is not unique. It depends on the choice of the code basis. The election of a particular representation is motivated by the fact that it can be useful for determining some code parameters. For example, the standard form for a d_H-code provides an easy way to obtain the parity check matrix of this code and allows the characterization of MDS codes and the singleton defect in general, see [45]. The form is called standard because every code has, up to equivalence, a unique such representation.

Analogous representations of the standard form were presented in [1] and [15] for codes over vector spaces endowed with NRT and hierarchical metrics, respectively. A standard form is obtained by the choice of an appropriate basis of the code and eventually by reordering the columns. The choice of a basis is obtained operating with the rows of a generating matrix. When considering poset metrics, it may be permitted to perform some operations (other than permutations) with the columns. Those permitted operations are determined by the group of linear isometries. In the referred cases of NRT and hierarchical posets, by operating with the columns we can get a generating matrix (or equivalently, a basis) that in many senses may be called canonical. The canonical matrix determines a canonical decomposition of a code into a direct sum of sub-codes with dimension and support uniquely prescribed by the code weight hierarchy. For a general poset, there is no such canonical decomposition, but there is a decomposition that is maximal in the number of components. In the sequence we will first present the canonical form originally presented in [15] and then proceed to define and determine what we call by a maximal P-decomposition. This decomposition is what allows to produce bounds for invariants that cannot be related or produced by explicit expressions.

3.1 Canonical Form for Hierarchical Metrics

In [15], it was proposed a canonical-systematic form for codes over spaces endowed with a hierarchical poset metric. In this section, we will present an alternative way to obtain the canonical-systematic form emphasizing the role of the basis of the code instead of the generating matrix. When using the generator matrix, the focus is on the form of the matrix, here, an particular form may be obtained in some cases by the choice of an appropriate basis which is determined by a decomposition. Consequently, a particular form for the matrix is not the main goal, the goal is to obtain a generator matrix as simple as possible.

Let P be a hierarchical poset over $[n]$. Given $w \in \mathbb{F}_q^n$, the *P-clean* of w is the vector $\widetilde{w}$ with the same value of w in the coordinates that are maximal in the support of w and zero in the remaining coordinates, namely: $\widetilde{w}_i = w_i$ if $i \in Max(w)$ and $\widetilde{w}_i = 0$ otherwise. Note that the P-clean of w has the same P-weight of w. The j-th projection of w is the vector $w^{(j)} \in \mathbb{F}_q^n$ obtained by the projection of w into the coordinates corresponding to the j-th level of P, i.e., ${w^{(j)}}_i = w_i$ if $i \in \Gamma_P^j$ and ${w^{(j)}}_i = 0$, otherwise. In the following lemma and theorem, if $w \in \mathbb{F}_q^n$, $\langle w \rangle$ denotes the one-dimensional space generated by w (not to be confused with the ideal generated by the support of w).

Lemma 3.1.1. Let P be a hierarchical poset with s levels over $[n]$ and $\mathcal{C}$ be a l-dimensional code satisfying $supp(\mathcal{C}) \subset \Gamma_P^{i_0}$ for some $i_0 \in \{1, \ldots, s\}$. Given $w \in \mathbb{F}_q^n$ such that $Max(w) \subset \Gamma_P^{i_0}$ and $\widetilde{w} \notin \mathcal{C}$, there exists a linear code $\mathcal{C}'$ such that $\mathcal{C}' \sim_P \mathcal{C} \oplus \langle w \rangle$ and $supp(\mathcal{C}') \subset \Gamma_P^{i_0}$.

Proof. Let $\{v^1, \ldots, v^l\}$ be a basis of $\mathcal{C}$. Given w as in the statements of the lemma, since $\widetilde{w} \notin \mathcal{C}$, the set $\beta = \{\widetilde{w}, v^1, \ldots, v^l\}$ is linearly independent. So, β is a basis for the linear code $\mathcal{C} \oplus \langle \widetilde{w} \rangle$. Let

$$\beta_1 = \{\widetilde{w}, v^1, \ldots, v^l, e_{j_1}, \ldots, e_{j_{n-l-1}}\}$$

be a basis of $\mathbb{F}_q^n$ obtained by extending the basis β using vectors of the canonical basis. Because $w = \widetilde{w} + \sum_{i=1}^{n-l-1} \alpha_i e_{j_i}$ for suitable scalars α_i, the set

$$\beta_2 = \{w, v^1, \ldots, v^l, e_{j_1}, \ldots, e_{j_{n-l-1}}\}$$

generates $\mathbb{F}_q^n$. Furthermore, note that β_2 contains a basis of $\mathcal{C} \oplus \langle w \rangle$. Define the linear map T by setting $T(w) = \widetilde{w}$, $T(v^i) = v^i$ for all $i \in \{1, \dots, l\}$ and $T(e_{j_r}) = e_{j_r}$ for all $r \in \{1, \dots, n-l-1\}$. Denoting by $\mathcal{C}'$ the code generated by β, we have that $supp(\mathcal{C}') \subset \Gamma_P^{i_0}$ and $T(\mathcal{C} \oplus \langle w \rangle) = \mathcal{C}'$. To conclude, we just need to prove that T is an isometry. Writing a canonical vector $e_j \in \mathbb{F}_q^n$ according to β_2 where $j \notin \{j_1, \dots, j_{n-l-1}\}$, we get that

$$e_j = \alpha w + \sum_{i=1}^{l} \gamma_i v^i + \sum_{i=1}^{n-l-1} \theta_i e_{j_i},$$

where $\alpha, \gamma_i, \theta_i \in \mathbb{F}_q$. Since $w = \widetilde{w} + \sum_{i=1}^{n-l-1} \alpha_i e_{j_i}$ and $\alpha_i = 0$ if $j_i \in \Gamma_P^r$ with $r \geq i_0$ ($\widetilde{w}$ is the P-clean of w), then

$$\begin{aligned} T(e_j) &= \alpha \widetilde{w} + \sum_{i=1}^{l} \gamma_i v^i + \sum_{i=1}^{n-l-1} \theta_i e_{j_i} \\ &= \alpha \left(w - \sum_{i=1}^{n-l-1} \alpha_i e_{j_i} \right) + \sum_{i=1}^{l} \gamma_i v^i + \sum_{i=1}^{n-l-1} \theta_i e_{j_i} = e_j - \sum_{i=1}^{n-l-1} \alpha \alpha_i e_{j_i}. \end{aligned}$$

Due to the fact that P is hierarchical and $\alpha_i = 0$ for every i such that $j_i \in \Gamma_P^r$ with $r \geq i_0$, the characterization given in Lemma 2.3.6 ensures that T is a linear isometry. □

Observation 3.1.2. The linear isometry T constructed in the previous lemma coincides with the identity map when restricted to elements whose support does not intercept the level i_0, i.e., $T(x) = x$ if $supp(x) \subset [n] \setminus \Gamma_P^{i_0}$.

Theorem 3.1.3. Let P be an $(n : n_1, \dots, n_s)$ hierarchical poset. If $\mathcal{C}$ is a linear P-code, there exists $\widetilde{\mathcal{C}}$ such that $\widetilde{\mathcal{C}} \sim_P \mathcal{C}$,

$$\widetilde{\mathcal{C}} = \mathcal{C}_1 \oplus \mathcal{C}_2 \oplus \cdots \oplus \mathcal{C}_s$$

and $supp(\mathcal{C}_i) \subset \Gamma_P^i$ for every $i \in [s]$.

Proof. If $\mathcal{C}$ is an one-dimensional code, then $\mathcal{C} = \langle v \rangle$ for a suitable $v \in \mathbb{F}_q^n$. Considering $\mathcal{C}' = \{\mathbf{0}\}$ to be the null vector space, since $v \notin \mathcal{C}'$, Lemma 3.1.1 ensures that $\mathcal{C}$ is equivalent to the code generated by the P-clean of v. Suppose $\mathcal{C}$ is a k-dimensional code and $k > 1$. Then $\mathcal{C} = \langle w \rangle \oplus \mathcal{C}'$ for some $w \in \mathcal{C}$ and a $(k-1)$-dimensional code $\mathcal{C}' \subset \mathcal{C}$. By induction, there is a linear isometry T' such that $T'(\mathcal{C}') = \oplus_{i=1}^{s} B_i$ and $supp(B_i) \subset \Gamma_P^i$. Since T' is injective, we get that $T'(w) \notin T'(\mathcal{C}')$. Consider the code $T'(\mathcal{C}) = \langle T'(w) \rangle \oplus T'(\mathcal{C}')$.

Therefore, there exists a level i of P such that the projection of $T'(w)$ in this level does not belong to B_i, i.e, $T'(w)^{(j)} \notin B_i$. Let i_0 be the maximal level with this property. Define the vector $v \in \mathbb{F}_q^n$ as follows: $v^{(i)} = T'(w)^{(i)}$ if $i \leq i_0$ and $v^{(i)} = \mathbf{0}$ if $i > i_0$. Since $v = T'(w) - c$ for some $c \in T'(\mathcal{C}')$ we get that $T'(\mathcal{C}) = \langle v \rangle \oplus T'(\mathcal{C}')$. By Lemma 3.1.1, there is a linear isometry T such that $T(B_i) = B_i$ if $i \neq i_0$ (ensured by the Observation 3.1.2) and $supp(T(B_{i_0} \oplus \langle v \rangle)) = supp(T(B_{i_0}) \oplus \langle \tilde{v} \rangle) \subset \Gamma_P^{i_0}$ where $T(v) = \tilde{v}$. Therefore, denoting $\mathcal{C}_i = B_i$ for every $i \neq i_0$ and $\mathcal{C}_{i_0} = T(B_{i_0}) \oplus \langle \tilde{v} \rangle$, we have that $\widetilde{\mathcal{C}} = TT'(\mathcal{C}) = \oplus_{i=1}^s \mathcal{C}_i$ and $supp(\mathcal{C}_i) \subset \Gamma_P^i$ for every $i \in \{1, \ldots, s\}$. □

Given a linear code $\mathcal{C}$, Theorem 3.1.3 states that there exists a code $\widetilde{\mathcal{C}}$, equivalent to $\mathcal{C}$, such that $\widetilde{\mathcal{C}} = \mathcal{C}_1 \oplus \ldots \oplus \mathcal{C}_s$ and $supp(\mathcal{C}_i) \subset \Gamma_P^i$ for every $i \in \{1, \ldots, s\}$. Assuming P to be naturally labeled, the direct sum can be seen as a product of codes, namely,

$$\widetilde{\mathcal{C}} = \mathcal{C}_1 \oplus \ldots \oplus \mathcal{C}_s = (\widehat{\mathcal{C}_1}, \ldots, \widehat{\mathcal{C}_s}),$$

where the codes $\widehat{\mathcal{C}_i}$ were obtained from $\mathcal{C}_i$ just by deleting the coordinates that do not belong to the level i of the poset (this is known as the punctured code construction, see [27]). Therefore, while $\mathcal{C}_i$ is a sub-code of $\mathbb{F}_q^n$, $\widehat{\mathcal{C}_i}$ is a sub-code of $\mathbb{F}_q^{n_i}$ where $n_i = |\Gamma_P^i|$.

Corollary 3.1.4. If $\mathcal{C}$ is a P-code where P is a naturally labeled $(n : n_1, \ldots, n_s)$ hierarchical poset, there is a code $\mathcal{C}'$, equivalent to $\mathcal{C}$, with generator matrix G in the form

$$G = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & G_s \\ \mathbf{0} & \mathbf{0} & & \iddots & \mathbf{0} \\ \vdots & \vdots & & & \vdots \\ \mathbf{0} & G_2 & \mathbf{0} & \cdots & \mathbf{0} \\ G_1 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix},$$

where G_i is the generator matrix of $\widehat{\mathcal{C}_i}$.

The representation given by Corollary 3.1.4 can be standardized even more. Indeed, by operating in the rows of G we can assume that G_i is in reduced row echelon form, note that this operations only perform changing of basis in the code. If $T \in \mathcal{A}ut(P)$, according to Example 2.3.7, T is a composition of permutations according to each level of the poset. In terms of matrices, assuming that the generator matrix G of a code $\mathcal{C}$ is in the form given by Corollary 3.1.4 and that each G_i is in reduced row echelon form, the description of $\mathcal{A}ut(P)$ ensures that any two rows of G can be permuted if, and only if, the coordinates corresponding to these rows belong to the same level of the poset. Therefore, if P_i are $n_i \times n_i$ permutation matrices, the code generated by

$$G' = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & G_s P_s \\ \mathbf{0} & \mathbf{0} & & \iddots & \mathbf{0} \\ \vdots & \vdots & & & \vdots \\ \mathbf{0} & G_2 P_2 & \mathbf{0} & \cdots & \mathbf{0} \\ G_1 P_1 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}$$

is equivalent to $\mathcal{C}$. With the previous arguments, the following representation holds.

Corollary 3.1.5. If $\mathcal{C}$ is a P-code where P is a naturally labeled $(n : n_1, \ldots, n_s)$ hierarchical poset, there is a code $\mathcal{C}'$, equivalent to $\mathcal{C}$ with generator matrix G given by

$$G = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & [I_{k_s}|A_s] \\ \mathbf{0} & \mathbf{0} & & \iddots & \mathbf{0} \\ \vdots & \vdots & & & \vdots \\ \mathbf{0} & [I_{k_2}|A_2] & \mathbf{0} & \cdots & \mathbf{0} \\ [I_{k_1}|A_1] & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}$$

where k_i is the dimension of $\widehat{\mathcal{C}_i}$ and A_i are $k_i \times (n_i - k_i)$ matrices. This representation is called the *canonical-systematic form.*

As may be seen in [15], the constants $k_1, \ldots, k_s$ are determined by the weight enumerator of the code and, in this sense, the decomposition is canonical.

3.2 Partitions and Decompositions

A *partition* of a subset $J \subseteq [n]$ is a family of subsets $\{J_1, \ldots, J_r\}$ such that

$$J = \bigcup_{i=1}^{r} J_i,$$

where $J_i \cap J_j = \emptyset$ if $i \neq j$ and each $J_i \neq \emptyset$. We will denote such a partition by $\mathcal{J} = (J_i)_{i=1}^{r}$. If we write $J_0 = [n] \setminus J = \{i \in [n] \ : \ i \notin J\}$, the triple $\mathcal{J}^* = (J; J_0; J_i)_{i=1}^{r}$ is called *pointed* *partition*, since J is the union of the subsets J_i, it may be omitted in the notation, therefore the pointed partition $\mathcal{J}^*$ will also be denoted by $(J_0; J_i)_{i=1}^r$. Note that $J_0 = \emptyset$ if, and only if, $J = [n]$. We stress that J_0 has a special role, since it is the only part we allow to be empty. From now on, we consider only pointed partitions, so we will omit the symbol * and the adjective "pointed". A partition $\mathcal{J}$ can be refined in two ways, either by increasing the number of parts or by enlarging the distinguished part J_0. Except for the pointer J_0, the order of the other parts is irrelevant, for example,

$$(J_0; \{1,2\}, \{3,4,5\}) = (J_0; \{5,4,3\}, \{1,2\}) .$$

Definition 3.2.1. An *l-split of a partition* $\mathcal{J} = (J_0; J_i)_{i=1}^r$ is a partition $\mathcal{J}' = (J_0; J_i')_{i=1}^{r+1}$ where $J_i = J_i'$ for each $i \neq l$ and $J_l = J_l' \cup J_{r+1}'$, with both J_l' and J_{r+1}' non-empty. This means that J_l is split into two components and the others are unchanged. An *l-aggregate* of a partition $\mathcal{J} = (J_0; J_i)_{i=1}^r$ is a partition $\mathcal{J}' = (J_0'; J_i')_{i=1}^r$ where $J_i' = J_i$ if $i \notin \{l, 0\}$, $J_l = J_l' \cup J_l^*$ and $J_0' = J_0 \cup J_l^*$ for some $\emptyset \neq J_l^* \subsetneq J_l$, i.e., some elements of J_l were aggregated into the distinguished part J_0.

Definition 3.2.2. We say that a partition $\mathcal{J}'$ is a *1-step refinement of* $\mathcal{J}$ if $\mathcal{J}'$ is obtained from $\mathcal{J}$ by performing a single l-split or a single l-aggregate operation, for some $l < |\mathcal{J}|$. The partition $\mathcal{J}'$ is a *refinement of* $\mathcal{J}$ if $\mathcal{J}'$ can be obtained from $\mathcal{J}$ by a successive number of 1-step refinements. We use the notation $\mathcal{J} \geq \mathcal{J}'$ to denote a refinement and $\mathcal{J} \geq_l \mathcal{J}'$ to denote that $\mathcal{J}'$ is a 1-step refinement of $\mathcal{J}$ performed by an l-split or l-aggregate. When the kind of operation (splitting or aggregation) is relevant, we will use the notation $\mathcal{J} \geq_l^s \mathcal{J}'$ and $\mathcal{J} \geq_l^a \mathcal{J}'$ for an l-split or an l-aggregate, respectively.

Example 3.2.3. The partition $([4]; \emptyset; \{1,2,3,4\})$ can be refined in order to get the partition $([4]; \{1,3\}; \{2\}, \{4\})$ by using the following 1-step refinements:

$$\begin{aligned}(\emptyset; \{1,2,3,4\}) &\geq_1^a (\{3\}; \{1,2,4\}) \\ &\geq_1^a (\{1,3\}; \{2,4\}) \geq_1^s (\{1,3\}; \{2\}, \{4\}) .\end{aligned}$$

By using the set partition previously defined, decompositions of linear codes can be constructed. Each set partition over $[n]$ induces a decomposition of linear odes, as explained below. Such decompositions are the algebraic equivalent of the set partitions over $[n]$.

Definition 3.2.4. We say that $\mathscr{C} = (\mathcal{C}; \mathcal{C}_0; \mathcal{C}_i)_{i=1}^r$ is a *decomposition of* an $[n, k, \delta]_q$ code $\mathcal{C}$ if each $\mathcal{C}_i$ is a subspace of $\mathbb{F}_q^n$ such that:

(a) $\mathcal{C} = \oplus_{i=1}^r \mathcal{C}_i$ with $dim(\mathcal{C}_i) > 0$ for every $i \in \{1, \ldots, r\}$;

(b) $\mathcal{C}_0 = \{(x_1, x_2, \ldots, x_n) \; : \; x_i = 0 \text{ if } i \in supp\,(\mathcal{C})\}$;

(c) $(supp\,(\mathcal{C}_0)\,; supp\,(\mathcal{C}_i))_{i=1}^r$ is a pointed partition over $[n]$.

Definition 3.2.5. An l-split, l-aggregate, 1-step refinement and a refinement $\mathscr{C}' = (\mathcal{C}; \mathcal{C}_0'; \mathcal{C}_i')_{i=1}^{r\prime}$ of a decomposition $\mathscr{C} = (\mathcal{C}; \mathcal{C}_0; \mathcal{C}_i)_{i=1}^r$ are defined according to

$$(supp\,(\mathcal{C}_0')\,; supp\,(\mathcal{C}_i'))_{i=1}^{r\prime}$$

being an l-split, l-aggregate, 1-step refinement or a refinement of $(supp\,(\mathcal{C}_0)\,; supp\,(\mathcal{C}_i))_{i=1}^r$.

Example 3.2.6. Let P be a poset over $[4]$. Consider the $[4, 2, \delta_P]_2$ code $\mathcal{C}$ given by

$$\mathcal{C} = \{0000, 1100, 0010, 1110\}.$$

Then,

$$(\mathcal{C}; \langle e_4 \rangle; \mathcal{C}) \geq_1^s (\mathcal{C}; \langle e_4 \rangle; \mathcal{C}_i)_{i=1}^2$$

where

$$\mathcal{C}_1 = \{0000, 1100\} \text{ and } \mathcal{C}_2 = \{0000, 0010\}.$$

Definition 3.2.7. A decomposition $\mathscr{C} = (\mathcal{C}; \mathcal{C}_0; \mathcal{C}_i)_{i=1}^r$ is said to be *maximal* if it does not admit a refinement.

Let $\beta = \{e_1, \ldots, e_n\}$ be the canonical basis of $\mathbb{F}_q^n$. Given $I \subset [n]$, the I-*coordinate subspace* V_I is defined by

$$V_I = \langle \{e_i \; : \; i \in I\} \rangle = \left\{ \sum_{i \in I} x_i e_i \; : \; x_i \in \mathbb{F}_q \right\}.$$

Given a decomposition $\mathscr{C} = (\mathcal{C}; \mathcal{C}_0; \mathcal{C}_i)_{i=1}^r$ of a linear code $\mathcal{C}$, we say that

$$V_i := V_{supp(\mathcal{C}_i)} = \langle \{e_i \; : \; i \in supp\,(\mathcal{C}_i)\} \rangle$$

is the *support-space of (the component)* $\mathcal{C}_i$. We consider $[n]_{\mathcal{C}} = supp\,(\mathcal{C})$ and $[n]^{\mathcal{C}} = [n]\setminus[n]_{\mathcal{C}}$. In the case where $[n]^{\mathcal{C}} \neq \emptyset$, we write $V_0 = V_{[n]^{\mathcal{C}}}$ and denote $\mathcal{C}_0 = V_0$. We say that the decomposition $(\mathcal{C};\mathcal{C}_0;\mathcal{C}_i)_{i=1}^r$ is supported by the *environment decomposition* $(V_0;V_i)_{i=1}^r$. In the case where $[n]^{\mathcal{C}} = \emptyset$, we have $V_0 = \mathcal{C}_0 = \{\mathbf{0}\}$.

Until now, the metric d_P has not played role in the decomposition of a code. We introduce now a decomposition that depends of the metric and consequently of the poset P.

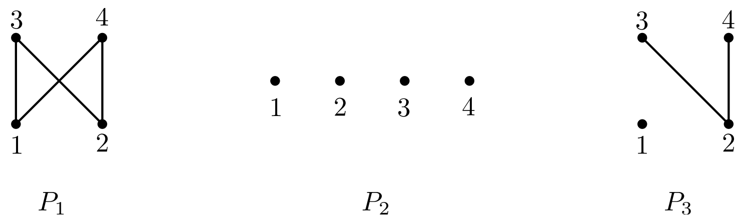
Definition 3.2.8. A *P-decomposition of* $\mathcal{C}$ is a decomposition $\mathscr{C} = (\mathcal{C}';\mathcal{C}_0';\mathcal{C}_i')_{i=1}^r$ of $\mathcal{C}'$ (as in Definition 3.2.4) where $\mathcal{C}' \sim_P \mathcal{C}$. Each $\mathcal{C}_i'$ is called a *component* of the decomposition. A *trivial P-decomposition* of $\mathcal{C}$ is either the decomposition $(\mathcal{C};\mathcal{C}_0;\mathcal{C})$ or any P-decomposition with a unique factor $(\mathcal{C}';\mathcal{C}_0';\mathcal{C}')$ where $|supp\,(\mathcal{C}')| = |supp\,(\mathcal{C})|$ and $|supp\,(\mathcal{C}_0')| = |supp\,(\mathcal{C}_0)|$.

Definition 3.2.9. A code $\mathcal{C}$ is said to be *P-irreducible* if it does not admit a non-trivial P-decomposition.

Example 3.2.10. Consider the $[4,2,\delta_P]_q$ code $\mathcal{C}$ given by

$$\mathcal{C} = \{0000, 1110, 0111, 1001\}.$$

Let P_1 be the $(4:2,2)$ hierarchical poset, P_2 be the anti-chain poset and P_3 be the poset determined by $2 \leq_{P_3} 3$ and $2 \leq_{P_3} 4$. Their Hasse diagrams are as follows:



It is straightforward that $\mathcal{C}$ does not admit a non-trivial P_2-decomposition, then $\mathcal{C}$ is P_2-irreducible. If we consider the d_{P_3}-isometry given by

$$(x_1, x_2, x_3, x_4) \mapsto (x_1, x_2 + x_3, x_3, x_4),$$

then

$$\mathscr{C}' = (\{0000, 1010, 0011, 1001\}; \langle e_2\rangle; \{0000, 1010, 0011, 1001\}).$$

is a P_3-decomposition of $\mathcal{C}$. Note that the previous isometry is also a d_{P_1}-isometry, hence $\mathscr{C}'$ is also a P_1-decomposition of $\mathcal{C}$. Considering the d_{P_1}-isometry given by

$$(x_1, x_2, x_3, x_4) \mapsto (x_1 + x_3 + x_4, x_2, x_3, x_4),$$

we get that

$$\mathscr{C}'' = (\{0000, 0010, 0011, 0001\}; \langle e_1, e_2\rangle; \{0000, 0010\} \oplus \{0000, 0001\})$$

is also a P_1-decomposition for $\mathcal{C}$.

Given a P-decomposition $\mathscr{C} = (\mathcal{C}'; \mathcal{C}'_0; \mathcal{C}'_i)_{i=1}^{r}$ of $\mathcal{C}$, we have that $[n] = \cup_{i=0}^{r} supp\,(V_i)$. Also, if $\mathcal{C}'_i = \{\mathbf{0}\}$, then $i = 0$. If each $\mathcal{C}'_i$ is P-irreducible, denoting $n_i = dim(V_i)$ and $k_i = dim(\mathcal{C}'_i)$, then

$$\sum_{i=0}^{r} n_i = n = dim(\mathbb{F}_q^n) \text{ and } \sum_{i=1}^{r} k_i = k = dim(\mathcal{C}).$$

Consider two P-decompositions $\mathscr{C}' = (\mathcal{C}'; \mathcal{C}'_0; \mathcal{C}'_i)_{i=1}^{r'}$ and $\mathscr{C}'' = (\mathcal{C}''; \mathcal{C}''_0; \mathcal{C}''_i)_{i=1}^{r''}$ of a code $\mathcal{C}$. Associated to those P-decompositions there are two partitions of $[n]$, namely: $(supp\,(\mathcal{C}'_0)\,; supp\,(\mathcal{C}'_i))_{i=1}^{r'}$ and $(supp\,(\mathcal{C}''_0)\,; supp\,(\mathcal{C}''_i))_{i=1}^{r''}$. By the definition of a P-decomposition, there are isometries $T', T'' \in GL_P(\mathbb{F}_q^n)$ such that $T'(\mathcal{C}) = \mathcal{C}'$ and $T''(\mathcal{C}) = \mathcal{C}''$. Denote $T = T'' \circ (T')^{-1}$. Then T is a linear isometry and $T(\mathcal{C}') = \mathcal{C}''$. By Lemma 2.3.5, T induces an automorphism of order $\phi_T : [n] \to [n]$. The automorphism ϕ_T induces a map on the partition of $[n]$ determined by the P-decomposition $\mathscr{C}'$, namely,

$$\phi_T[(supp\,(\mathcal{C}'_0)\,; supp\ (\mathcal{C}'_i))_{i=1}^{r'}] = (\phi_T\,(supp\,(\mathcal{C}'_0))\,; \phi_T\,(supp\,(\mathcal{C}'_i)))_{i=1}^{r'}\,.$$

Using the previous notations, we can define the analogous of the operations over decompositions to P-decompositions.

Definition 3.2.11. Let $\mathcal{C} \subset \mathbb{F}_q^n$ be a linear code and P be a poset over $[n]$. Let $\mathscr{C}' = (\mathcal{C}'; \mathcal{C}'_0; \mathcal{C}'_i)_{i=1}^{r'}$ and $\mathscr{C}'' = (\mathcal{C}''; \mathcal{C}''_0; \mathcal{C}''_i)_{i=1}^{r''}$ be two P-decompositions of $\mathcal{C}$. We say that $\mathscr{C}'$ is *a P-refinement* (1*-step P-refinement*) of $\mathscr{C}''$ if $\phi_T[(supp\,(\mathcal{C}'_0)\,; supp\,(\mathcal{C}'_i))_{i=1}^{r'}]$ is a refinement (1-step refinement) of the partition $(supp\,(\mathcal{C}''_0)\,; supp\,(\mathcal{C}''_i))_{i=1}^{r''}$.

Similar to what was done with set partitions in Definition 3.2.2, the symbols

$\overset{P}{\geq}{}_l^s$ and $\overset{P}{\geq}{}_l^a$ specify when the refinement was obtained by an l-split or an l-aggregation, respectively.

Example 3.2.12. Let $\mathcal{C} = \{0000, 1100, 0011, 1111\}$ be a 2-dimensional code over $\mathbb{F}_2^4$ and P be the chain poset determined by $1 \leqslant 2 \leqslant 3 \leqslant 4$. Then, starting with the trivial P-decomposition, we have the following refinements:

$$\begin{aligned}
&(\mathcal{C}; \emptyset; \mathcal{C}) \overset{P}{\geq}{}_1^a (\{0000, 1100, 0001, 1101\}; \{0000, 0010\}; \{0000, 1100, 0001, 1101\}) \\
&\overset{P}{\geq}{}_1^a (\{0000, 0100, 0001, 0101\}; \{0000, 0010, 1000, 1010\}; \{0000, 0100, 0001, 0101\}) \\
&\overset{P}{\geq}{}_1^s (\{0000, 0100, 0001, 0101\}; \{0000, 0010, 1000, 1010\}; \{0000, 0100\}, \{0000, 0001\})
\end{aligned}$$

where the first two refinements were obtained by considering the linear P-isometries

$$(x_1, x_2, x_3, x_4) \mapsto (x_1, x_2, x_3 - x_4, x_4)$$

and

$$(x_1, x_2, x_3, x_4) \mapsto (x_1 - x_2, x_2, x_3, x_4),$$

respectively. The third refinement is just a splitting

$$\{0000, 0100, 0001, 0101\} = \{0000, 0100\} \oplus \{0000, 0001\}.$$

Definition 3.2.13. A P-decomposition $\mathscr{C} = (\mathcal{C}; \mathcal{C}_0; \mathcal{C}_i)_{i=1}^r$ is said to be *maximal* if each $\mathcal{C}_i$ is P-irreducible for every $i \in \{1, \ldots, r\}$.

Let us consider the Hamming metric d_H over $\mathbb{F}_2^n$. It is well-known that the group of linear isometries of this metric space is isomorphic to the permutation group $\mathcal{S}_n$. Given a code $\mathcal{C}$, let $\mathscr{C} = (\mathcal{C}; \mathcal{C}_0; \mathcal{C}_i)_{i=1}^r$ be a maximal decomposition of $\mathcal{C}$. If a code $\mathcal{C}'$ is d_H-equivalent to $\mathcal{C}$, then there is $T \in GL_{d_H}(\mathbb{F}_q^n) \sim \mathcal{S}_n$ such that $T(\mathcal{C}) = \mathcal{C}'$. Note that the decomposition $\mathscr{C}' = (\mathcal{C}'; T(\mathcal{C}_0); T(\mathcal{C}_i))_{i=1}^r$ is a maximal decomposition of $\mathcal{C}'$, otherwise, $\mathscr{C}$ would not be a maximal decomposition of $\mathcal{C}$. Therefore, when considering the Hamming metric, a maximal decomposition is also a maximal H-decomposition (H is the anti-chain poset over $[n]$). This is not the general case, as we can see in the following example.

Example 3.2.14. Consider $\mathcal{C}$ to be the 1-dimensional binary code of length n generated by $(1, 1, \ldots, 1) \in \mathbb{F}_2^n$. Let P be a poset defined by the chain order: $1 \leqslant 2 \leqslant \cdots \leqslant n$; and

H be the anti-chain order poset. It follows that $\mathcal{C}$ is H-irreducible but not P-irreducible. Indeed, by Lemma 2.3.6, the map

$$T\left(x_1,\ldots,x_{n-1},x_n\right)=\left(x_1+x_n,\ldots,x_{n-1}+x_n,x_n\right)$$

is a P-isometry because $T(e_i)=e_i$ for every $i\neq n$ and $T(e_n)=\sum_{i=1}^{n}e_i$. Also, $T\left(\mathcal{C}\right)=\langle e_n\rangle$ is the code generated by the vector e_n, hence

$$\mathscr{C}'=\left(\langle e_n\rangle;\langle\{e_1,\ldots,e_{n-1}\}\rangle;\langle e_n\rangle\right)$$

is a maximal P-decomposition of $\mathcal{C}$. Therefore, $\mathcal{C}$ is not P-irreducible. Since $\mathcal{C}$ is a one-dimensional code, we only can perform aggregations, but since $supp(11\ldots1)=[n]$, aggregations change the Hamming weight, so that $\mathcal{C}$ is H-irreducible.

Definition 3.2.15. Let $\mathscr{C}=\left(\mathcal{C}';\mathcal{C}_0;\mathcal{C}_i\right)_{i=1}^{r}$ be a P-decomposition of an $\left[n,k,\delta\right]_q$ code $\mathcal{C}$. The *profile* of $\mathscr{C}$ is the array

$$\text{profile}\left(\mathscr{C}\right):=\left[\left(n_0,k_0\right),\left(n_1,k_1\right),\ldots,\left(n_r,k_r\right)\right],$$

where

$$n_i=\left|supp\left(\mathcal{C}_i\right)\right|\ \text{ and }\ k_i=dim\left(\mathcal{C}_i\right).$$

Observe that $|supp(\mathcal{C}_i)|=dim(V_i)$, then $n=n_0+n_1+\cdots+n_r$ and $k=k_1+k_2+\cdots+k_r$. The following theorem states that the profile of a maximal P-decomposition $\mathscr{C}$ of a code $\mathcal{C}$ depends (essentially) exclusively on $\mathcal{C}$, not on $\mathscr{C}$.

Theorem 3.2.16. Let $\mathcal{C}$ be an $\left[n,k,\delta\right]_q$ code and let P be a poset over $[n]$. Let $\mathscr{C}'$ and $\mathscr{C}''$ be two maximal P-decompositions of $\mathcal{C}$ with

$$\text{profile}\left(\mathscr{C}'\right)=\left[\left(n_0',{k'}_0\right),\left(n_1',k_1'\right),\ldots,\left(n_r',k_r'\right)\right]$$

and

$$\text{profile}\left(\mathscr{C}''\right)=\left[\left(n_0'',k_0''\right),\left(n_1'',k_1''\right),\ldots,\left(n_s'',k_s''\right)\right].$$

Then, $r=s$ and, up to a permutation, profile$\left(\mathscr{C}'\right)=$ profile$\left(\mathscr{C}''\right)$, i.e., there is $\sigma\in\mathcal{S}_r$ such that $(n_i',k_i')=(n_{\sigma(i)}'',k_{\sigma(i)}'')$ and $(n_0',{k'}_0)=(n_0'',k_0'')$.

Proof. Let $\mathscr{C}' = (\mathcal{C}'; \mathcal{C}'_0; \mathcal{C}'_i)_{i=1}^r$ and $\mathscr{C}'' = (\mathcal{C}''; \mathcal{C}''_0; \mathcal{C}''_i)_{i=1}^s$ be two maximal P-decompositions of $\mathcal{C}$. Suppose, without loss of generality, $r < s$. If $T \in GL_P(\mathbb{F}_q^n)$ is an isometry satisfying $T(\mathcal{C}') = \mathcal{C}''$, there is a component $\mathcal{C}'_i$ of $\mathscr{C}'$ such that $T(\mathcal{C}'_i)$ is not contained in any component $\mathcal{C}''_j$ of $\mathscr{C}''$, otherwise $r \geq s$. Hence, there are components $\mathcal{C}'_{i_0}$ of $\mathscr{C}'$ and $\mathcal{C}''_{j_0}, \mathcal{C}''_{j_1}, \ldots, \mathcal{C}''_{j_t}$ of $\mathscr{C}''$ such that

$$T\left(\mathcal{C}'_{i_0}\right) \subset \mathcal{C}''_{j_0} \oplus \mathcal{C}''_{j_1} \oplus \cdots \oplus \mathcal{C}''_{j_t}$$

and $T\left(\mathcal{C}'_{i_0}\right) \cap \mathcal{C}''_{j_l} \neq \emptyset$ for any $l \in \{1, \ldots, t\}$. Therefore,

$$T\left(\mathcal{C}'_{i_0}\right) = \bigoplus_{m=0}^{t} T\left(\mathcal{C}'_{i_0}\right) \cap \mathcal{C}''_{j_m}$$

is a non-trivial P-decomposition for $\mathcal{C}'_{i_0}$, contradicting the fact that each component of a maximal P-decomposition is P-irreducible. It follows that $r = s$. Moreover, for every $i \in \{1, \ldots, r\}$ there is j_i such that $T(\mathcal{C}'_i) \subseteq \mathcal{C}''_{j_i}$. Hence, $n'_i \leq n''_{j_i}$ and $k'_i \leq k''_{j_i}$. Applying the same reasoning to $T^{-1} \in GL_P(\mathbb{F}_q^n)$, we get that $n''_i \leq n'_{j_i}$ and $k''_i \leq k'_{j_i}$, hence $n'_i = n''_{j_i}$ and $k'_i = k''_{j_i}$, so that, up to a permutation, $\text{profile}(\mathscr{C}') = \text{profile}(\mathscr{C}'')$. □

The next Corollary follows straight from Theorem 3.2.16.

Corollary 3.2.17. Let $\mathscr{C}' = (\mathcal{C}'; \mathcal{C}'_0; \mathcal{C}'_i)_{i=1}^r$ and $\mathscr{C}'' = (\mathcal{C}''; \mathcal{C}''_0; \mathcal{C}''_i)_{i=1}^r$ be two maximal P-decompositions of $\mathcal{C}$ and let $T \in GL_P(\mathbb{F}_q^n)$ be a linear isometry such that $T(\mathcal{C}') = \mathcal{C}''$. Then, there is a permutation $\sigma \in \mathcal{S}_r$ such that $T(\mathcal{C}'_i) = \mathcal{C}''_{\sigma(i)}$.

To express the amount of operations (splitting and aggregations) performed in a decomposition, we will define the complexity of a decomposition (do not confuse with the computational complexity of problems).

Definition 3.2.18. Given a P-decomposition $\mathscr{C} = (\mathcal{C}'; \mathcal{C}_0; \mathcal{C}_i)_{i=1}^r$ of $\mathcal{C}$, its *complexity* is defined by

$$\mathcal{O}_P(\mathscr{C}) = \frac{1}{r} + \sum_{i=1}^{r} n_i - k_i,$$

where $n_i = dim(V_i)$ (the dimension of the support-space of $\mathcal{C}_i$) and $k_i = dim(\mathcal{C}_i)$ for every $1 \leq i \leq r$. A P-decomposition with minimum complexity is called *primary P-decomposition.*

Note that the complexity of a P-decomposition is completely determined by its profile. Each aggregation decreases the complexity since some of the parcels $n_i - k_i$ in the summation decrease. A splitting also decreases the complexity since it increases the number of components (the number r). By Theorem 3.2.16, maximal P-decompositions have the same complexity. Thus, the minimum complexity of a decomposition of a code $\mathcal{C}$ will be denoted by $\mathcal{O}(\mathcal{C})$ instead of $\mathcal{O}(\mathscr{C})$. It is straightforward that aggregations and splittings decrease the complexity of a P-decomposition; actually, if $\mathscr{C}'$ is a refinement of $\mathscr{C}$, then $\mathcal{O}_P(\mathscr{C}') < \mathcal{O}_P(\mathscr{C})$. Therefore, we have the following proposition.

Proposition 3.2.19. A P-decomposition of a code $\mathcal{C}$ is maximal if, and only if, it is a primary P-decomposition.

The permutation part $\mathcal{A}ut(P)$ of $GL_P(\mathbb{F}_q^n)$ does not alter the complexity of a decomposition, hence it is irrelevant regarding maximality of P-decompositions.

Lemma 3.2.20. Let $\mathscr{C} = (\mathcal{C}'; \mathcal{C}_0; \mathcal{C}_i)_{i=1}^r$ be a maximal P-decomposition of $\mathcal{C}$. Let $\phi \in Aut\,(P)$ and $T_\phi \in \mathcal{A}ut(P)$ be the isometry induced by ϕ. Then,

$$\mathscr{C}' = \left(T_\phi\left(\mathcal{C}'\right); T_\phi\left(\mathcal{C}_0\right); T_\phi\left(\mathcal{C}_i\right)\right)_{i=1}^r$$

is also a maximal P-decomposition of $\mathcal{C}$.

Proof. Because ϕ is a permutation, for every $i \in \{0, 1, \ldots, r\}$,

$$j \in supp(\mathcal{C}_i) \iff \phi(j) \in supp(T_\phi(\mathcal{C}_i)).$$

Hence, $\mathscr{C}'$ is a P-decomposition of $\mathcal{C}$. Since its profile coincides with the profile of $\mathscr{C}$, $\mathscr{C}'$ is also a maximal P-decomposition of $\mathcal{C}$. □

We recall that the set $\mathcal{P}_n$ of all posets over $[n]$ is itself a partially ordered set. Maximal P-decompositions "behaves well" according to this order in the following way:

Theorem 3.2.21. Let $P, Q \in \mathcal{P}_n$ with $P \leq Q$. Given a code $\mathcal{C}$, there is a maximal P-decomposition of $\mathcal{C}$ which is also a Q-decomposition of $\mathcal{C}$.

Proof. Assume $P, Q \in \mathcal{P}_n$ and $P < Q$. Let $\mathscr{C}' = (\mathcal{C}'; \mathcal{C}_0; \mathcal{C}_i)_{i=1}^r$ be a maximal P-decomposition of $\mathcal{C}$ and $T \in GL_P(\mathbb{F}_q^n)$ such that $T(\mathcal{C}) = \mathcal{C}'$. By the characterization of $GL_P(\mathbb{F}_q^n)$, $T = A \circ T_\phi$ where $A \in \mathcal{G}_P$ and $T_\phi \in \mathcal{A}ut(P)$. From Lemma 3.2.20, we have that

$$\mathscr{C}'' = (T_{\phi^{-1}}(\mathcal{C}'); T_{\phi^{-1}}(\mathcal{C}_0); T_{\phi^{-1}}(\mathcal{C}_i))_{i=1}^r$$

is also a maximal P-decomposition of $\mathcal{C}$. However $T_{\phi^{-1}}(\mathcal{C}') = T_{\phi^{-1}} \circ T(\mathcal{C})$, hence

$$T_{\phi^{-1}}(\mathcal{C}') = T_{\phi^{-1}} \circ A \circ T_\phi(\mathcal{C}).$$

We stress that $\mathcal{G}_P \subset \mathcal{G}_Q$ always that $P \leq Q$. Because $\mathcal{G}_P$ is a normal subgroup of $GL_P(\mathbb{F}_q^n)$, it follows that $T_{\phi^{-1}} \circ A \circ T_\phi \in \mathcal{G}_P$, hence $T_{\phi^{-1}} \circ A \circ T_\phi \in \mathcal{G}_Q$. Therefore, $\mathscr{C}''$ is a Q-decomposition of $\mathcal{C}$. □

As a direct consequence of the previous theorem, we obtain a relation among primary decompositions of posets and the natural order over $\mathcal{P}_n$.

Corollary 3.2.22. Let $P, Q \in \mathcal{P}_n$ with $P \leq Q$. Then, $\mathcal{O}_Q(\mathcal{C}) \leq \mathcal{O}_P(\mathcal{C})$ for every linear code $\mathcal{C}$.

Concerning primary P-decompositions, there is always a code that it is differently decomposed depending on the poset, indeed:

Proposition 3.2.23. Let $P, Q \in \mathcal{P}_n$ with $P < Q$. Then, there is a code $\mathcal{C}$ such that

$$\mathcal{O}_Q(\mathcal{C}) < \mathcal{O}_P(\mathcal{C}).$$

Proof. Let us suppose $P < P_1 \leq Q$ and that P_1 covers P. Then, there are $i_0, j_0 \in [n]$ such that

$$i_0 \not\leqslant_P j_0 \text{ and } i_0 \leqslant_{P_1} j_0.$$

Let $\mathcal{C}$ be the one-dimensional code generated by $e_{i_0} + e_{j_0}$ and let $\mathcal{C}'$ be the one-dimensional code generated by e_{j_0}. The linear map T defined by $T(e_j) = e_j$ for every $j \in [n] \setminus \{i_0, j_0\}$, $T(e_{i_0}) = e_{i_0}$ and $T(e_{j_0}) = e_{j_0} - e_{i_0}$, is a P_1-linear isometry by the characterization of $GL_P(\mathbb{F}_q^n)$ given in Lemma 2.3.6. Therefore, the P_1-decomposition $\mathscr{C} = (\mathcal{C}'; \mathcal{C}_0'; \langle e_{j_0} \rangle)$ of $\mathcal{C}$ is a primary P_1-decomposition. We note that $\mathcal{C}_0' = \mathbb{F}_q^n \setminus \langle e_{j_0} \rangle$ and therefore, $\mathcal{O}_{P_1}(\mathcal{C}) = 1$.

On the other hand, we claim that $\mathscr{C}' = (\mathcal{C}; \mathcal{C}_0; \mathcal{C})$, where $\mathcal{C}_0 = \oplus_{i \neq i_0, j_0} V_i$, is a P-primary decomposition of $\mathcal{C}$. Indeed, suppose, for a contradiction that $\mathscr{C}'$ is not a P-primary decomposition, the only way to refine it would be by constructing a 1-dimensional code $\mathcal{C}''$ with $|supp\,(\mathcal{C}'')| = 1$ and $\mathcal{C}'' \sim_P \mathcal{C}$. Hence, $\mathcal{C}''$ would be generated by a canonical vector e_d for some $d \in [n]$. However, each $T \in GL_P(\mathbb{F}_q^n)$ determines an order automorphism $\phi_T \in Aut\,(P)$ as in Lemma 2.3.5. Suppose $T \in GL_P(\mathbb{F}_q^n)$ and $T(\mathcal{C}) = \mathcal{C}''$, therefore $T(e_{i_0} + e_{j_0}) = e_d$ for some $d \in [n]$. Considering $S = T^{-1}$, it follows that $Max(\langle supp(S(e_d))\rangle_P) = \{i_0, j_0\}$, which is a contradiction since S does not determine an order automorphism as in Lemma 2.3.5. Thus, such isometry does not exist. Therefore, $\mathscr{C}'$ is a primary P-decomposition for $\mathcal{C}$ and

$$1 = \mathcal{O}_{P_1}(\mathcal{C}) < \mathcal{O}_P(\mathcal{C}) = 2. \tag{3.1}$$

Since $P_1 \leq Q$, Corollary 3.2.22 ensures that

$$\mathcal{O}_Q(\mathcal{C}) \leq \mathcal{O}_{P_1}(\mathcal{C}). \tag{3.2}$$

Inequalities 3.1 and 3.2 imply $\mathcal{O}_Q(\mathcal{C}) < \mathcal{O}_P(\mathcal{C})$. □

We remark that Corollary 3.2.22 together with Proposition 3.2.23 implies that primary decomposition is a characterization of posets, in the sense that a given poset P may be reconstructed from the profile of codes according to P. Moreover, looking at the proof of Proposition 3.2.23, one may notice that the reconstruction can be done by considering only the $n(n-1)/2$ pairs of vectors (e_i, e_j).

3.2.1 Hierarchical Bounds

Hierarchical poset metrics are well understood. In particular, if P is a hierarchical poset, the profile of a primary P-decomposition of a code $\mathcal{C}$ is uniquely (and easily) determined by the weight hierarchy of the code. Moreover, as we can see in [30], this property is exclusive of hierarchical posets. For this reason, when considering a general poset P, we aim to establish bounds for $\mathcal{O}_P(\mathcal{C})$ considering the easy-to-compute primary P-decompositions relatively to hierarchical posets. In the next section, we will characterize the generator matrices that determine maximal P-decompositions providing an algorithm to obtain these matrices. But the complexity to determine these matrices will not be discussed.

We say that a poset P is *hierarchical at level* i if levels Γ_P^i and Γ_P^j relate hierarchically for every $j \in \{1, \ldots, i-1\}$, i.e., if $a \in \Gamma_P^j$ for some $j \in \{1, \ldots, i-1\}$ and $b \in \Gamma_P^i$ then $a \leqslant b$. Let

$$\mathcal{H}(P) = \{i \in [r] \ : \ P \text{ is hierarchical at } i\}.$$

It is clear that P is hierarchical if, and only if, $\mathcal{H}(P) = [r]$.

We stress that for every integers a and b with $a < b$, the notation $[a, b]$ denotes the set $\{a, a+1, \ldots, b\}$. Given a poset P with height r, then:

(1) Consider

$$s_1 = \min\{i \in [r] \ : \ P \text{ is not hierarchical at level } i+1\}$$

and

$$s_2 = \min\{i \in [r] \ : \ i > s_1 \text{ and } P \text{ is hierarchical at level } i+1\}.$$

Denote

$$J_i = [n_1 + \cdots + n_{i-1} + 1, n_1 + \cdots + n_i]$$

for every $1 \leq i < s_1$ and

$$J_{s_1} = [n_1 + \cdots + n_{s_1-1} + 1, n_1 + \cdots + n_{s_2}].$$

(2) Consider

$$s_3 = \min\{i \in [r] \ : \ i > s_2 \text{ and } P \text{ is not hierarchical at level } i+1\}$$

and

$$s_4 = \min\{i \in [r] \ : \ i > s_3 \text{ and } P \text{ is hierarchical at level } i+1\}.$$

Thus, denote

$$J_{s_1+i} = [n_1 + \cdots + n_{s_2+i-1} + 1, n_1 + \cdots + n_{s_2+i}]$$

for every $i \geq 1$ satisfying $s_2 + i < s_3$ and

$$J_{s_1+(s_3-s_2)} = [n_1 + \cdots + n_{s_3-1} + 1, n_1 + \cdots + n_{s_4}].$$

(3) Consider

$$s_5 = \min\{i \in [r] \ : \ i > s_4 \text{ and } P \text{ is not hierarchical at level } i+1\}$$

and

$$s_6 = \min\{i \in [r] \ : \ i > s_5 \text{ and } P \text{ is hierarchical at level } i+1\}.$$

Therefore,

$$J_{s_1+(s_3-s_2)+i} = [n_1 + \cdots + n_{s_4+i-1} + 1, n_1 + \cdots + n_{s_4+i}]$$

for every $i \geq 1$ satisfying $s_4 + i < s_5$ and

$$J_{s_1+(s_3-s_2)+(s_5-s_4)} = [n_1 + \cdots + n_{s_5-1} + 1, n_1 + \cdots + n_{s_6}].$$

Proceeding in this way, we construct a partition of $[n]$ given by $J_1 \cup \ldots \cup J_h$ for some h. An arbitrary iteration of the process can be described as follows. **(General Case)** Consider t to be an odd integer, then

$$s_t = \min\{i \in [r] \ : \ i > s_{t-1} \text{ and } P \text{ is not hierarchical at level } i+1\}$$

and,

$$s_{t+1} = \min\{i \in [r] \ : \ i > s_t \text{ and } P \text{ is hierarchical at level } i+1\}.$$

Furthermore,

$$J_{s_1+(s_3-s_2)+\cdots+(s_{2t-1}-s_{2t-2})} = [n_1 + \cdots + n_{s_{2t-1}-1} + 1, n_1 + \cdots + n_{2t}]$$

and

$$J_{s_1+(s_3-s_2)+\cdots+(s_{2t-1}-s_{2t-2})+i} = [n_1 + \cdots + n_{s_{2t}+i-1} + 1, n_1 + \cdots + n_{2t+i}].$$

Therefore, out of P we can define two hierarchical posets:

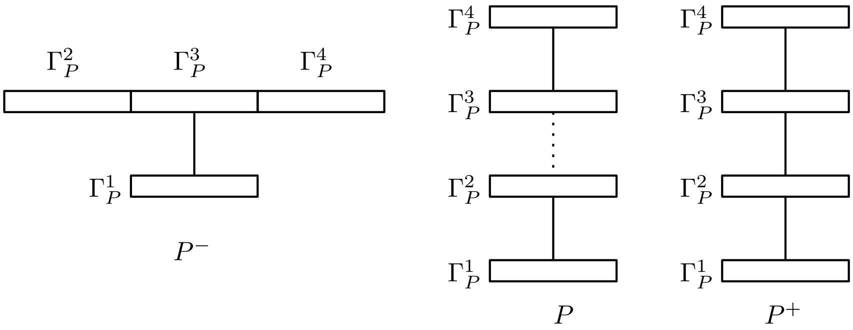
Upper neighbor: Let P^+ be the poset over $[n]$ with the same level decomposition of P and for every $a \in \Gamma_P^i$ and $b \in \Gamma_P^j$ with $a \neq b$, define $a \leqslant_{P^+} b$ if, and only if, $i < j$.

Lower neighbor: Let P^- be the poset over $[n]$ with level hierarchy $[n] = J_1 \cup J_2 \cup \ldots \cup J_h$ and for $a \in \Gamma_P^i$ and $b \in \Gamma_P^j$ with $a \neq b$, we have $a \leqslant_{P^-} b$ if, and only if,

$i < j$.

In order to clarify the definitions of the upper and lower neighbors, we shall present an illustrative example.

Example 3.2.24. Let P be a poset with 4 levels. Suppose $\mathcal{H}(P) = \{1, 2, 4\}$. Consider the simplified Hasse diagram representing whether the poset is hierarchical at a particular level or not: if two consecutive levels are joined by a dotted line, then the poset is not hierarchical at the level above the other; if two consecutive levels are joined by a line, then the poset is hierarchical at the level above. Therefore, the diagrams of P^-, P and P^+ are as follows:



We stress that $\Gamma_{P^-}^2 = \Gamma_P^2 \cup \Gamma_P^3 \cup \Gamma_P^4$ since the smallest level where P is hierarchical at this level is the third one. Hence, in order to construct P^-, all levels of P above the level 2 are gathered in the second level of P^-.

It is easy to see that:

(a) P^+ and P^- are hierarchical posets and P is hierarchical if, and only if, $P = P^+ = P^-$;

(b) Considering the natural order $\leq$ on $\mathcal{P}_n$, we have that $P^- \leq P \leq P^+$. Moreover,

$$P^- = \max \{Q \in \mathcal{P}_n \ : \ Q \leq P \text{ and } Q \text{ is hierarchical}\}$$

and

$$P^+ = \min \{Q \in \mathcal{P}_n \ : \ P \leq Q \text{ and } Q \text{ is hierarchical}\}.$$

The next proposition follows directly from Corollary 3.2.22.

Proposition 3.2.25. For any linear code $\mathcal{C}$,

$$\mathcal{O}_{P^+}(\mathcal{C}) \leq \mathcal{O}_P(\mathcal{C}) \leq \mathcal{O}_{P^-}(\mathcal{C}).$$

Example 3.2.26. Consider the posets of Example 3.2.10 and denote $P = P_3$, then $P^+ = P_1$ and $P^- = P_2$. Furthermore, using the decompositions in that example (which are maximal), we conclude that

$$\mathcal{O}_{P^+}(\mathcal{C}) = 1/2,\ \mathcal{O}_P(\mathcal{C}) = 2 \text{ and } \mathcal{O}_{P^-}(\mathcal{C}) = 3.$$

If P is not hierarchical, both inequalities are strict for some code $\mathcal{C}$. Moreover, the bounds are tight, in the sense that, given a poset P, there are codes $\mathcal{C}_1$ and $\mathcal{C}_2$ such that $\mathcal{O}_{P^+}(\mathcal{C}_1) = \mathcal{O}_P(\mathcal{C}_1)$ and $\mathcal{O}_{P^-}(\mathcal{C}_2) = \mathcal{O}_P(\mathcal{C}_2)$ (just consider any code $\mathcal{C}$ with $supp(\mathcal{C}) \subset \Gamma_P^1$).

3.2.2 Packing Radius Bounds

Maximal P-decompositions may be useful to determine bounds for the packing radius of a code. We remark that the packing radius of codes according to hierarchical metrics was completely characterized in Proposition 2.3.12 and it depends essentially on the minimum distance of the code, see [30]. We also remark that for non-hierarchical metrics, the complexity to determine the packing radius of a single vector is NP-hard (see [12]) and the packing vector is not necessarily a vector with minimum distance.

Proposition 3.2.27. Let $\mathscr{C} = (\mathcal{C}'; \mathcal{C}_0; \mathcal{C}_i)_{i=1}^r$ be a P-decomposition for $\mathcal{C}$. Then,

$$\mathcal{R}_P(\mathcal{C}) \leq \min_{i \in \{1,\ldots,r\}} \mathcal{R}_P(\mathcal{C}_i).$$

Proof. Note that $\mathcal{R}_P(\mathcal{C}) = \mathcal{R}_P(\mathcal{C}')$. Furthermore, since each $\mathcal{C}_i$ is a subcode of $\mathcal{C}'$, it follows that $\mathcal{R}_P(\mathcal{C}) \leq \mathcal{R}_P(\mathcal{C}_i)$ for every $i \in \{1, \ldots, r\}$. □

Proposition 3.2.28. If $P \leq Q$, then $\mathcal{R}_P(\mathcal{C}) \leq \mathcal{R}_Q(\mathcal{C})$ for every linear code $\mathcal{C}$.

Proof. If follows directly from the fact that

$$B_Q(r, c) \cap B_Q(r, \mathbf{0}) \subset B_P(r, c) \cap B_P(r, \mathbf{0})$$

for every $c \in \mathcal{C}$ and any integer $r \geq 0$. □

Consider the projections $\pi_i : \mathbb{F}_q^n \longrightarrow \mathbb{F}_q^{n_i}$ (note that $n_i = |supp(\mathcal{C}_i)|$) defined by

$$\pi_i(x_1, \dots, x_n) = (x_{i_1}, \dots, x_{i_s})$$

where $i_1 < \cdots < i_s$ and $\{i_1, \dots, i_s\} = supp(\mathcal{C}_i)$. Consider on $\mathbb{F}_q^{n_i}$ the metric d_{π_i} induced by π_i in the sense that $\pi_i : V_i \to \mathbb{F}_q^{n_i}$ is an isometry. Then, by the definition of d_{π_i},

$$\mathcal{R}_P(\mathcal{C}_i) = \mathcal{R}_{d_{\pi_i}}(\pi_i(\mathcal{C}_i)). \tag{3.3}$$

Using Proposition 3.2.28 and the upper and lower neighbors P^+ and P^- defined in the previous section, bounds for the packing radius of codes according to hierarchical posets may be obtained.

Proposition 3.2.29. Given a maximal P-decomposition $\mathscr{C} = (\mathcal{C}'; \mathcal{C}_0; \mathcal{C}_i)_{i=1}^r$ for $\mathcal{C}$,

$$\mathcal{R}_{P^-}(\mathcal{C}_i) \leq \mathcal{R}_P(\mathcal{C}_i) \leq \mathcal{R}_{P^+}(\mathcal{C}_i)$$

for every $i \in \{1, \dots, r\}$.

Propositions 3.2.27 and 3.2.29 yield the following upper and lower bounds for the packing radius:

$$\mathcal{R}_P(\mathcal{C}) \leq \min_{i \in \{1,\dots,r\}} \mathcal{R}_{P^+}(\mathcal{C}_i) \tag{3.4}$$

for every $i \in \{1, \dots, r\}$.

Since P^+ and P^- are hierarchical, these bounds are obtained just by finding the minimum distance of each code $\mathcal{C}_i$. If P is hierarchical, the bounds obtained in Proposition 3.2.29 and in Inequality (3.4) are tight.

3.2.3 Construction of Maximal P-Decompositions

In order to find maximal P-decompositions of a code $\mathcal{C}$, we need to find codes that are equivalent to $\mathcal{C}$ and verify if one of theirs maximal decompositions is a refinement of the given maximal decomposition of $\mathcal{C}$. To do so, we will provide a construction that outputs a generator matrix of a code $\mathcal{C}' \sim_P \mathcal{C}$, which determines a maximal P-decomposition of $\mathcal{C}$.

Given a poset P over $[n]$ and an $[n,k,\delta]_q$ code $\mathcal{C}$, let $G=(g_{ij})$ be a $k\times n$ generating matrix of $\mathcal{C}$. As we have already noted, we lose no generality by assuming that G is in a reduced row echelon form, obtained by elementary operations on rows. In order to obtain a maximal P-decomposition, we need to change the classical definition of the reduced row echelon form. For each i, let

$$j\,(i)=\max\{j\ :\ g_{ij}\neq 0\}$$

be the right-most non-zero column of the i-th row of G. Performing elementary row operations on G, we may assume that

$$j(1)>j(2)>\cdots>j(k)\ \text{ and }\ g_{ij(l)}=0\ \text{ if }\ i\neq l. \tag{3.5}$$

We say that G is in *reduced row echelon form* if the entries of G satisfy 3.5. From now on, we assumed that generator matrices have this form. In order to clarify the difference of the proposed reduced row echelon form and the classical one, we present the next example.

Example 3.2.30. Let

$$G=\begin{bmatrix}1&0&1&1&0\\1&1&0&1&1\\0&1&0&1&1\end{bmatrix}$$

be a generator matrix for a $[5,3,1]_2$ d_H-code $\mathcal{C}$. Then,

$$G_1=\begin{bmatrix}1&0&0&0&0\\0&1&0&1&1\\0&0&1&1&0\end{bmatrix}\text{ and } G_2=\begin{bmatrix}0&1&1&0&1\\0&0&1&1&0\\1&0&0&0&0\end{bmatrix}$$

where G_1 was obtained by the classical construction of reduced row echelon form and G_2 was obtained by the proposed form. Note that

(a) if $j'(i)=min\{j\ :\ g_{ij}\neq 0\}$. Considering G_1 we get that

$$j'(1)=1,\ j'(2)=2 \text{ and } j'(3)=3, \text{ hence } j'(1)<j'(2)<j'(3);$$

(b) on the other hand, considering G_2,

$$j(1) = 5,\ j(2) = 4 \text{ and } j(3) = 1, \text{ hence } j(1) > j(2) > j(3).$$

A generator matrix G determines a unique decomposition of $\mathcal{C}$ in the following sense:

Construction of $\mathscr{C}$:

Suppose $\beta_1 = \{v_1, \ldots, v_k\}$ is the set of all rows of G and $I \subset [n]$ is the index set of the null columns of G. Then, define

$$\mathcal{C}_0 = \{v \in \mathbb{F}_q^n \ : \ v = \sum_{i \in I} x_i e_i\}.$$

Take $w_1 \in \beta_1$ and let $\gamma_1 = \{v_{i_1}, \ldots, v_{i_r}\} \subset \beta_1$ be the set of all rows of G such that, if $v \in \gamma_1$, then $supp(w_1) \cap supp(v) \neq \emptyset$. Denote

$$\mathcal{C}_1 = \{c \in \mathcal{C} \ : \ c = \sum_{j=1}^{r} x_i v_{i_j}\}.$$

Take $\beta_2 = \beta_1 \backslash \gamma_1$, if it is empty, the matrix G has determined the decomposition $(\mathcal{C}; \mathcal{C}_0; \mathcal{C}_1)$. If $\beta_2 \neq \emptyset$, take $w_2 \in \beta_2$ and $\gamma_2 = \{v_{i_1}, \ldots, v_{i_s}\} \subset \beta_2$ defined by the elements of β_2 whose support intercepts the support of w_2 and define

$$\mathcal{C}_2 = \left\{ c \in \mathcal{C} \ : \ c = \sum_{j=1}^{s} x_i v_{i_j} \right\}.$$

Proceeding this way, at some point, $\beta_{r+1} = \emptyset$ and $\beta_r \neq \emptyset$ for some r. Then, $\mathscr{C} = (\mathcal{C}; \mathcal{C}_0; \mathcal{C}_i)_{i=1}^{r}$ determines a decomposition of $\mathcal{C}$. It is clear that the choice of w_i does not interfere in the construction of the decomposition, therefore, the decomposition constructed is unique, up to a permutation of its components.

Example 3.2.31. Considering the generator matrix G given in Example 3.2.30, it follows that the decomposition determined by G is the trivial one $(\mathcal{C}; \emptyset; \mathcal{C})$ since every two rows of G have intersection in their supports. On the other hand, the decomposition obtained by the matrix G_2 is not trivial, indeed, the third row has disjoint support from the first and second rows, therefore, if

$$\mathcal{C}_1 = \{00000, 01101, 00110, 01011\}$$

and

$$\mathcal{C}_2 = \{00000, 10000\}.$$

The decomposition $(\mathcal{C}; \emptyset; \mathcal{C}_i)_{i=1}^2$ is the one determined by G_2.

A generator matrix G is said to be in the *generalized reduced row echelon form* if there is a permutation in the rows of G such that the resultant matrix is in a reduced row echelon form.

Proposition 3.2.32. Let $\mathcal{C}$ be an $[n, k, \delta]_q$ code. If G is a generator matrix for $\mathcal{C}$ in a generalized reduced row echelon form, then G determines a maximal decomposition for $\mathcal{C}$.

Proof. Let G be a generator matrix for $\mathcal{C}$ in a generalized reduced row echelon form. Without loss of generality, we will assume that the decomposition determined by G (as in the construction previously presented) has only one component, i.e., $\mathscr{C} = (\mathcal{C}; \mathcal{C}_0; \mathcal{C})$. Suppose, for contradiction, there is a refinement for $\mathscr{C}$.

Claim 1: Aggregations are not allowed: if $i_0 \in supp(\mathcal{C})$, every basis of $\mathcal{C}$ would have an element v such that $i_0 \in supp(v)$ but $supp(\mathcal{C}_0)$ is the set of null columns of G.

Claim 2: Splittings are not allowed: Suppose there is a splitting for $\mathscr{C}$, in other words, suppose there is a decomposition $\mathscr{C}' = (\mathcal{C}; \mathcal{C}_0; \mathcal{C}_i)_{i=1}^2$. Thus, $\mathcal{C} = \mathcal{C}_1 \oplus \mathcal{C}_2$ with $supp(\mathcal{C}_1) \cap supp(\mathcal{C}_2) = \emptyset$. Let $w_1, \ldots, w_{k_1}$ be a basis for $\mathcal{C}_1$ and $w_{k_1+1}, \ldots, w_k$ be a basis for $\mathcal{C}_2$. The $k \times n$ matrix G_1 having $w_1, \ldots, w_k$ as its rows, is a generator matrix for $\mathcal{C}$ and its decomposition coincides with $\mathscr{C}'$. Then, $AG = G_1$ for some $k \times k$ matrix A, i.e., each row of G_1 is a linear combination of rows of G. Let β_1 be the minimum set of rows of G generating $w_1, \ldots, w_{k_1}$, i.e., for every $1 \leq i \leq k_1$,

$$w_i = \sum_{v \in \beta_1} x_v^i v$$

where $x_v^i \in \mathbb{F}_q$ and there is no i such that $x_v^i = 0$ for every $v \in \beta_1$. Let β_2 be the minimum set of rows of G generating $w_{k_1+1}, \ldots, w_k$, i.e.,

$$w_i = \sum_{v \in \beta_2} x_v^i v$$

for every $k_1 + 1 \leq i \leq k$ and there is no i such that $x_v^i = 0$ for every $v \in \beta_2$. Suppose $v_0 \in \beta_1 \cap \beta_2$. Since G is in a reduced row form, there is a coordinate i_0 of v_0 such that it is the only non-null entry of the i_0-th column of G. Then, there exist $1 \leq j_1 \leq k_1$ and $k_1 + 1 \leq j_2 \leq k$ such that the coordinate i_0 of both w_{j_1} and w_{j_2} are non-null, which is a contradiction, therefore $\beta_1 \cap \beta_2 = \emptyset$. By construction, $supp(\beta_1) \cap supp(\beta_2) \neq \emptyset$, otherwise the decomposition defined by G would be not trivial. Let $i_0 \in supp(\beta_1) \cap supp(\beta_2)$, then $i_0 \notin supp(\mathcal{C}_0)$ and, supposing $i_0 \in supp(\mathcal{C}_2)$, it follows that $i_0 \notin supp(\mathcal{C}_1)$. Since $i_0 \in supp(\beta_1)$, there exists a row w of G_1 such that its coordinate i_0 is non-null. Therefore, $i_0 \in supp(\mathcal{C}_1)$, which is a contradiction. Hence, the decomposition determined by G does not admits refinements, therefore, it is maximal. □

The previous proposition ensures that each maximal decomposition is obtained by taking a generator matrix in a generalized reduced row echelon form. In the following, we will consider the triangular subgroup of isometries $\mathcal{G}_P \subseteq GL_P(\mathbb{F}_q^n)$. By Lemma 3.2.20, these are the only isometries that matter when we are looking for maximal P-decompositions. By definition of $\mathcal{G}_P$, the following two operations over a generator matrix G will provide matrices generating equivalent codes to the one generated by G:

(OP 1) If $g_{i_0 j_0} \neq 0$, $g_{i j_0} = 0$ for every $i \neq i_0$ ($g_{i_0 j_0}$ is the only non-zero entry on the j_0-th column) and $r \leqslant j_0$ ($r \neq j_0$), we may assume $g_{i_0 r} = 0$;

This is equivalent to choosing the isometry $T \in \mathcal{G}_P$ such that $T(e_j) = e_j$ for every $j \neq j_0$ and

$$T(e_{j_0}) = e_{j_0} - g_{i_0 r} g_{i_0 j_0}^{-1} e_r.$$

(OP 2) More generally, if $r \leqslant j_1, j_2, \ldots, j_s$ ($r \neq j_i$ for every i) and there two rows i_1 and i_2 of G such that $g_{i_1 r} = \sum_{l=1}^{s} x_l g_{i_1 j_l}$ and $g_{i_2 r} = \sum_{l=1}^{s} x_l g_{i_2 j_l}$ for some choice of $x_1, \ldots, x_l$, and $g_{i j_l} = 0$ for every $i \neq i_1, i_2$ ($g_{i_1 j_k}$ and $g_{i_2 j_k}$ are the only non-zero entry on the j_k-th column for every $k \in \{1, \ldots, s\}$), we may assume $g_{i_1 r} = g_{i_2 r} = 0$. Even more generally, the procedure can be performed simultaneously to many lines, all those entries may be considered to be 0. If the column r is a linear combination of columns

$j_1, \ldots, j_s$, then one may exchange the column r by the null column. Let $\{g_1, \ldots, g_n\}$ be the set of columns of G and suppose

$$g_r = \sum_{i=1}^{s} x_i g_{j_i}.$$

Then, in order to perform the exchange of the r-th column of G by a null column we have to consider the isometry $T \in \mathcal{G}_P$ defined by $T(e_i) = e_i$ for every $i \notin \{j_1, \ldots, j_s\}$ and

$$T(e_{j_i}) = e_{j_i} - x_i e_r$$

for every $i \in \{1, \ldots, s\}$.

Definition 3.2.33. Let G be a generator matrix for an $[n, k, \delta]_q$ code $\mathcal{C}$. If G is in a generalized reduced row echelon form and no operations as defined in (1) and (2) can be performed over G, we say that G is in a *P-canonical form.*

Example 3.2.34. Let $\mathcal{C}$ be the $[6, 3]_2$ code with generator matrix given by

$$G = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Consider the poset P_1 with order relations $1 \leqslant 2$ and $3 \leqslant 4$. By using operations (OP 1) and (OP 2) we get the following matrix in a P_1-canonical form:

$$G' = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Furthermore, if we consider P_2 a poset such that $P_1 \subset P_2$ and $4 \leqslant 5$, then using operation (OP 1) we get

$$G'' = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

It is clear that G'' is in a P_2-canonical form. It is also clear that it determines a maximum P_2-decomposition for $\mathcal{C}$.

In the following, given $T \in \mathcal{G}_P$, denote by $A_T = (a_{ij}) \in G_P$ its matrix according to the canonical basis.

Theorem 3.2.35. Let G be a generator matrix of a code $\mathcal{C}$ and let $\mathscr{C} = (\mathcal{C}; \mathcal{C}_0; \mathcal{C}_i)_{i=1}^{r}$ be the decomposition determined by G. Then, if G is in a P-canonical form, the decomposition $\mathscr{C}$ is a maximum P-decomposition of $\mathcal{C}$.

Proof. We will first show that $\mathscr{C}$ does not admit aggregations. Suppose there is a P-decomposition $\mathscr{C}'$ determined by the linear isometry $T \in \mathcal{G}_P$ such that this decomposition is obtained by an aggregation from $\mathscr{C}$, i.e., $\mathscr{C}' = (T(\mathcal{C}); \mathcal{C}_0'; \mathcal{C}_i')_{i=1}^{r}$ and $\phi_T(supp(\mathcal{C}_0)) \subset supp(\mathcal{C}_0')$. Since $T \in \mathcal{G}_P$, the map ϕ_T coincides with the identity map, therefore $supp(\mathcal{C}_0) \subset supp(\mathcal{C}_0')$. Consider the matrix $G_1 = (g_{ij}^1)$ which i-th row is obtained by the action of T in the i-th row of G. Hence, G_1 generates $T(\mathcal{C})$. We remark that $A_T = (a_{ij}) \in G_P$ is the matrix obtained by T. If $i_0 \in supp(\mathcal{C}_0') \setminus supp(\mathcal{C}_0)$, then i_0 is a null column of G_1 obtained by the aggregation. The characterization of G_P ensures that

$$0 = g_{li_0}^1 = \sum_{\substack{j \\ i_0 \leqslant j}} a_{i_0 j} g_{lj}$$

for every $l \in \{1, \ldots, k\}$. Since $a_{i_0 i_0} \neq 0$, it follows that

$$g_{li_0} = \sum_{\substack{j \\ i_0 \leqslant j, i_0 \neq j}} \left(-\frac{a_{i_0 j}}{a_{i_0 i_0}} \right) g_{lj} \tag{3.6}$$

for every $l \in \{1, \ldots, k\}$. Since G is in a P-canonical form, i_0 can not be a column $j(i)$ for every $i \in \{1, \ldots, k\}$. Equation 3.6 together with Operation (2) ensures that the i_0-th column of G is already null, so $i_0 \in supp(\mathcal{C}_0)$. Therefore, $supp(\mathcal{C}_0') \subset supp(\mathcal{C}_0)$, so $\mathscr{C}'$ can not be obtained by aggregations from $\mathscr{C}$.

We now prove that $\mathscr{C}$ does not admit a splitting. To do so, we will assume, without loss of generality, $\mathscr{C} = (\mathcal{C}; \mathcal{C}_0; \mathcal{C})$ is the decomposition determined by G. Let $T \in \mathcal{G}_P$ be the isometry determining the splitting for $\mathscr{C}$ and consider G_1 as in the first part of the proof. Hence, the rows of G_1 can be split into two sets with disjoint support, i.e., there exist two disjoint sets I_1 and I_2 formed by rows of G_1 such that $|I_1|= k_1$, $|I_2|= k_2$ and $k_1 + k_2 = k$. Therefore, I_1 and I_2 generate $\mathcal{C}_1$ and $\mathcal{C}_2$ respectively, and $\mathscr{C}' = (T(\mathcal{C}); \mathcal{C}_0; \mathcal{C}_i)_{i=1}^{2}$ is the maximum decomposition determined by G_1. Since the rows of G cannot be split in this form, let i_0 be the right-most column of G such that i_0 belongs to the support of the first k_1 and also of the last k_2 rows of G. Since the support of the subspaces generated by the first k_1 and the last k_2 rows of G_1 respectively are disjoint and $i_0 \notin supp(\mathcal{C}_0')$ since $i_0 \notin supp(\mathcal{C}_0)$, then i_0 or belongs to the support of the first k_1 or of the last k_2 rows of G_1. Without loss of generality, suppose i_0 belongs to the support of the last k_2 rows of G_1. Therefore, if l is an integer such that $0 \leq l \leq k_1$, then

$$0 = g^1_{li_0} = \sum_{\substack{j \\ i_0 \leqslant j}} a_{i_0 j} g_{lj}.$$

Thus,

$$g_{li_0} = \sum_{\substack{j \\ i_0 \leqslant j, i_0 \neq j}} \left(-\frac{a_{i_0 j}}{a_{i_0 i_0}} \right) g_{lj} \tag{3.7}$$

for every $l \in \{1, \dots, k_1\}$.

Therefore, i_0 is not the last column of G such that $g_{li_0} \neq 0$ for every $l \in \{1, \dots, k_1\}$. Each column contributing to the summation 3.7 has support either in the first k_1 or in the last k_2 rows of G, therefore, Operation (OP 2) ensures that every vector in the first k_1 rows of G has zero in the coordinate i_0. Hence, i_0 is not in the support of the first k_1 rows of G. Therefore, $\mathcal{C} = \mathcal{C}_1 \oplus \mathcal{C}_2$ where $\mathcal{C}_1$ and $\mathcal{C}_2$ are generated by the first k_1 and the last k_2 rows of G respectively. We conclude that $\mathscr{C}$ and $\mathscr{C}'$ have the same profile, since G_1 is also in a reduced row form and by Proposition 3.2.32, $\mathscr{C}'$ is a maximal decomposition of $\mathcal{C}'$, therefore, $\mathscr{C}$ is a maximal P-decomposition. □

Each time we can exchange a column by a null column, we exchange a code $\mathcal{C}$ with P-decomposition $\mathscr{C} = (\mathcal{C}, \mathcal{C}_0, \mathcal{C}_i)_{i=1}^r$ by a P-equivalent code with P-decomposition $\mathscr{C}' = (\mathcal{C}', \mathcal{C}_0', \mathcal{C}_i')_{i=1}^r$ where $dim(\mathcal{C}_0') = dim(\mathcal{C}_0) + 1$. If the Operation (OP 2) is performed in a proper subset of the k-lines of G, we do not increase $dim(\mathcal{C}_0)$ but, we may split some of the $\mathcal{C}_i$ into $\mathcal{C}_i = \mathcal{C}_{i_1}' \oplus \mathcal{C}_{i_2}'$ with $supp(\mathcal{C}_{i_1}') \cap supp(\mathcal{C}_{i_2}') = \emptyset$.

Note that the role of P in such operations rests solely on the condition $j \leqslant j_1, j_2, \dots, j_{n_1}$. For the two extremal posets, namely, the anti-chain and the chain poset, the picture is absolutely clear: If P is an anti-chain (hence we are considering the Hamming metric), no such operation may be performed (since $i \leqslant j \iff i = j$). Hence, maximal P-decompositions coincide with maximal decompositions (see the paragraph after Definition 3.2.13) and the P-canonical form presented in Proposition 2.1.14 provides it; if P is a chain with $1 \leqslant 2 \leqslant \cdots \leqslant n$, then the first operation may be performed to every $j\,(i)$ in the reduced row echelon matrix G, i.e., we have that $\mathcal{C}$ is equivalent to a code that has a generating matrix $G = (g_{ij})$ such that $g_{ij(i)} = 1$ and $g_{ij} = 0$ if $j \neq j\,(i)$ (already known from [38]).

The other case that can be easily described is the case of hierarchical posets. The algorithm to find P-decompositions according to the levels of the poset was first proposed in [15]. In that work, the basis of the code providing such decomposition was also constructed. We explicit the matrix of this basis in Corollary 3.1.5. Its form is called canonical-systematic and provides a maximal P-decomposition for $\mathcal{C}$. Furthermore, the canonical-systematic form is a standard representation in the sense that every code has such decomposition. In [30], it was proved that for every non-hierarchical poset, it is not possible to give a standard representation like the one obtained in Corollary 3.1.5. Actually, this was the main motivation to define a canonical form as we did in Definition 3.2.33.

3.2.4 Complexity of Syndrome Decoding Algorithm

As seen in section 1.2.1, considering a metric determined by a weight, hence invariant by translations, a syndrome algorithm, which depends on the choice of the coset leaders, is an implementation of an MD-decoder. This is the case for poset metrics. Given an $[n, k, \delta]_q$ code $\mathcal{C}$, a look-up table for syndrome decoding of $\mathcal{C}$ (the table composed by the coset leaders) has q^{n-k} elements. This is an algebraic invariant and does not depend on the metric. Nevertheless, we can use refinements of decompositions in order to minimize the search space (number of cosets) by isometrically immersing the code into a smaller dimension space or by performing syndrome in each component of the decomposition. In this section we describe the cases for which such improvements are possible.

Let $\mathscr{C} = (\mathcal{C}; \mathcal{C}_0; \mathcal{C}_i)_{i=1}^{r}$ be a maximal P-decomposition of an $[n, k, \delta]_q$ code $\mathcal{C}$. Initially, note that in order to perform decoding, we can ignore the component V_0 (recall that V_i is the support-space of $\mathcal{C}_i$) since we know that any codeword $c = (c_1, \ldots, c_n) \in \mathcal{C}$ should have $c_j = 0$ for every $j \in [n]^{\mathcal{C}}$. Consider the projection $\pi : \mathbb{F}_q^n \longrightarrow \mathbb{F}_q^{n_1+\cdots+n_r}$ (recall that $n_i = |supp(\mathcal{C}_i)|$) defined by

$$\pi(x_1, \ldots, x_n) = (x_{i_1}, \ldots, x_{i_s})$$

where $i_1 < \cdots < i_s$ and $\{i_1, \ldots, i_s\} = supp(\mathcal{C})$. The map $\pi_{1,\ldots,r} = \pi|_{\oplus_{i=1}^r V_i}$, the restriction of π to the space $\oplus_{i=1}^r V_i$, is a bijection. Therefore, by pushing forward the metric in the restriction, we obtain the metric d_P^π in $\mathbb{F}_q^{n_1+\cdots+n_r}$ defined by

$$d_P^\pi(x,y) := d_P(\pi_{i,\ldots,r}^{-1}(x), \pi_{i,\ldots,r}^{-1}(y))$$

for every $x, y \in \mathbb{F}_q^{n_1+\cdots+n_r}$. The metric d_P^π turns the restriction $\pi_{1\ldots r}$ into a linear isometry. Because $\mathcal{C}$ is a subspace of $V_1 \oplus \ldots \oplus V_r$, the proof of the proposition below follows straight from these observations.

Proposition 3.2.36. The metric-decoding criteria of $\mathbb{F}_q^n$ for $\mathcal{C}$ is equivalent to the metric-decoding criteria of $\mathbb{F}_q^{n_1+\cdots+n_r}$ for $\pi(\mathcal{C})$, i.e.,

$$d_P(c', y) = \min_{c\in\mathcal{C}} d_P(c,y) \iff d_P^\pi(\pi(c'), \pi(y)) = \min_{c\in\pi(\mathcal{C})} d_P^\pi(c, \pi(y)).$$

By Proposition 3.2.36, to perform syndrome decoding, instead of using a look-up table with $|\mathbb{F}_q^n/\mathcal{C}|= q^{n-k}$ elements, we can reduce the number of cosets to $|\mathbb{F}_q^{n_1+\cdots+n_r}/\pi(\mathcal{C})|= \prod_{i=1}^r q^{n_i-k_i}$ elements. Note that

$$q^{n_0} \times \prod_{i=1}^r q^{n_i-k_i} = q^{n-k}.$$

Therefore, P-decompositions having the maximum possible number of elements in the support of $\mathcal{C}_0$ are the best ones in order to perform syndrome decoding.

Besides this possible (and a-posteriori irrelevant) gain in the cardinality of the syndrome look-up table obtained considering aggregations, there is a more significant gain that can be obtained through the splitting operation.

Proposition 3.2.37. If

$$\langle supp(\mathcal{C}_i)\rangle_P \cap \langle supp(\mathcal{C}_j)\rangle_P = \emptyset$$

for all $i \neq j$ and $i, j \neq 0$, then given $y \in \mathbb{F}_q^n$,

$$\min_{c\in\mathcal{C}} d_P(c,y) = \sum_{i=1}^r \min_{c\in\mathcal{C}_i} d_P^{\pi_i}(\pi_i(c), \pi_i(y)).$$

Proof. Note that if $c \in \mathcal{C}$, then $c = c_1 + \cdots + c_r$ with $c_i \in \mathcal{C}_i$ and for every $y_i \in \mathbb{F}_q^{n_i}$,

$$supp(y_i - c_i) \subset \langle supp(\mathcal{C}_i)\rangle = supp(V_i).$$

Then,

$$d(y, c) = d(y_1, c_1) + \cdots + d(y_r, c_r),$$

where $y = y_1 + \cdots + y_r$ and $y_i \in \mathbb{F}_q^{n_i}$ for all $i \in [r]$. □

Due to Proposition 3.2.37, if the ideals generated by each component are disjoint, syndrome decoding can be done independently in each component $\mathcal{C}_i$. Therefore, the number of cosets can be reduced from q^{n-k} elements to $\sum_{i=1}^{r} q^{n_i - k_i}$ elements, where the last one is the sum of the cosets in each quotient $\mathbb{F}_q^{n_i}/\pi_i(\mathcal{C}_i')$. More generally, if r is a disjoint union of subsets I_i, i.e., $[r] = I_1 \sqcup \ldots \sqcup I_s$, and

$$\langle supp(\oplus_{i\in I_j}\mathcal{C}_i)\rangle_P \cap \langle supp(\oplus_{i\in I_l}\mathcal{C}_i)\rangle_P = \emptyset,$$

for every $j \neq l$, then decoding can be separately done in each projection of $\oplus_{i\in I_j}\mathcal{C}_i$ into $\mathbb{F}_q^N$ where $N = \sum_{i\in I_j} n_i$.

Until now, in order to obtain the reductions, it was not necessary to change the syndrome decoding algorithm; we just performed it in a different (smaller) space. The hierarchical relation among elements of the poset will provide us a "quasi-independent" syndrome decoding algorithm that is performed by choosing first coordinates that hierarchically dominate others, therefore we will call this algorithm a *Leveled Syndrome Decoding.*

If $I, J \subset [n]$, we say that I and J are hierarchically related if every element in I is smaller than every element in J. Suppose $[r]$ is an ordered disjoint union of sets, $[r] = I_1 \sqcup \ldots \sqcup I_s$, such that $supp(\oplus_{i\in I_j}\mathcal{C}_i)$ is hierarchically related with $supp(\oplus_{i\in I_{j+1}}\mathcal{C}_i)$ for every $j \in \{1, \ldots, s-1\}$, so if $i_0 \in supp(\oplus_{i\in I_{j_0}}\mathcal{C}_i)$, then $i_0 \leqslant i$ for every $i \in supp(\oplus_{i\in I_l}\mathcal{C}_i)$ with $l > j_0$. By using the syndrome decoding algorithm described in Section 1.2.1, the leveled syndrome decoding algorithm is as follows:

Leveled Syndrome Decoding Algorithm 1.

Input: $y = y_1 + \cdots + y_s \in \mathbb{F}_q^n$ where $y_i \in \oplus_{j\in I_i} V_j$

For each $i \in \{1, \ldots, s\}$ do

```
        Decode $\pi_{j\in I_i}(y_i) \in \mathbb{F}_q^{\sum_{j\in I_i} n_j}$ using syndrome $(d_P^{\pi_{j\in I_i}})$ outputting $c_i \in \pi_{j\in I_i}(\oplus_{j\in I_i}\mathcal{C}_j)$
```

Output: $c = \pi_{j\in I_1}^{-1}(c_1) + \cdots + \pi_{j\in I_s}^{-1}(c_s)$.

As a particular case of a syndrome algorithm, we can order the levels to be decoded and do the following:

Leveled Syndrome Decoding Algorithm 2.

Input: $y = y_1 + \cdots + y_s \in \mathbb{F}_q^n$ where $y_i \in \oplus_{j\in I_i} V_j$

```
For $i = s$ to 1, do
  If $\pi_{j\in I_i}(y_i) \in \oplus_{j\in I_i}\mathcal{C}_j$ do
    $c_i = \pi_{j\in I_i}(y_i)$;
  else do
    Decode $\pi_{j\in I_i}(y_i) \in \mathbb{F}_q^{\sum_{j\in I_i} n_j}$ using syndrome $(d_P^{\pi_{j\in I_i}})$ outputting $c_i \in \pi_{j\in I_i}(\oplus_{j\in I_i}\mathcal{C}_j)$;
    Go to Output;
  end if;
end For;
```

Output: $c = \pi_{j\in I_i}^{-1}(c_i) + \pi_{j\in I_{i+1}}^{-1}(c_{i+1}) + \cdots + \pi_{j\in I_s}^{-1}(c_s)$.

The first algorithm was described in [15] for the hierarchical poset case. The second one is also a minimum distance algorithm according to the poset metric d_P (see the next proposition), but if there is an error in a particular level, the decoder outputs the null vector in the levels covered by the one where the error happened.

Proposition 3.2.38. Algorithm 2 determines a minimum distance decoder according to the metric d_P.

Proof. Given $y \in \mathbb{F}_q^n$ and $c \in \mathcal{C}$, then

$$d_P(y, c) = d_P^{\pi_{j\in I_i}}(\pi_{j\in I_i}(y), \pi_{j\in I_i}(c))$$

where i is the largest integer such that $\pi_{j\in I_i}(y) \neq \pi_{j\in I_i}(c)$. Therefore, if $c' \in \mathcal{C}$ satisfies

$$d_P(y, c') = \min_{c\in\mathcal{C}} d_P(y, c),$$

then $c'' = c'_s + c'_{s-1} + \cdots + c'_i$ also attains the minimum and this is the codeword returned by algorithm 2. □

Both the algorithms are needed to store the look-up tables of each quotient $\mathbb{F}_q^N/\pi_{j\in I_i}(\oplus_{j\in I_i}\mathcal{C}_j)$ where $\sum_{j\in I_i} n_j$. The total number of elements we need to store is

$$\sum_{i=1}^{s}\prod_{j\in I_i} q^{n_j-k_j}. \tag{3.8}$$

The difference between Algorithm 1 and 2 is that while the search in the algorithm 1 is performed over all elements of the look-up table, the search in Algorithm 2 is restricted to the first level where an error has occurred, i.e., if this happens in the level i_0, then the search is performed only over the look-up table of $\pi_{j\in I_{i_0}}(\oplus_{j\in I_{i_0}}\mathcal{C}_j)$ which has

$$\prod_{j\in I_{i_0}} q^{n_j-k_j} \tag{3.9}$$

elements. Note that if the poset has only one level, the values of Expressions (3.8) and (3.9) coincide.

Chapter 4

Expected Loss

In this chapter we explore the framework of expected loss introduced in [18] and [16]. This concept allows us to give different treatments for distinct kind of errors. As showed in [18], that seems important when dealing with image transmission. In this context, we express the mean of the expected losses according to a given decoder and, in a particular case, we show that the lexicographic encoder is better than the mean. Due to experimental results we conjcture that the lexicographic encoder is a Bayes encoder in the analized case. In the last section, we relate the theory of expected loss with unequal error protection presenting some conjectures and coding constructions to achieve unequal error protection.

Given an $[n, k, \delta]_q$ linear code $\mathcal{C} \subset \mathbb{F}_q^n$, the error probability of $\mathcal{C}$ (Equation 1.1) may be rewritten in the following way

$$P_e^D(\mathcal{C}) = \frac{1}{q^k} \sum_{c \in \mathcal{C}} \sum_{y \in \mathbb{F}_q^n} W(y|c) \sum_{\substack{c' \in \mathcal{C} \\ c' \neq c}} D(c'|y),$$

where D is a decoder (stochastic map) and W is a channel conditional probability. Let $\mathbb{R}_+$ be the set of non-negative real numbers. Consider the well-known *indicator function*

$$\mu_{\text{0-1}} : \mathcal{C} \times \mathcal{C} \to \mathbb{R}_+$$

given by

$$\mu_{\text{0-1}}(c, c') = \begin{cases} 0 & \text{if} \quad c = c' \\ 1 & \text{if} \quad c \neq c' \end{cases} .$$

Therefore, the error probability may be expressed as

$$P_e^D(\mathcal{C}) = \frac{1}{q^k} \sum_{c \in \mathcal{C}} \sum_{y \in \mathbb{F}_q^n} W(y|c) \sum_{c' \in \mathcal{C}} \mu_{0\text{-}1}(c, c') D(c'|y). \tag{4.1}$$

The function $\mu_{0\text{-}1}$ detects decoding errors but does not distinguish such errors. In some cases, depending on the nature of the information, we may assign for each pair $(c, c') \in \mathcal{C} \times \mathcal{C}$, a real value representing the amount of loss obtained provided that c is transmitted but is decoded as c' by the receiver. Such loss measures will be used to define the expected loss of an encoding-decoding scheme.

4.1 Expected Loss

The extension of the error probability definition to expected loss was explained in [18]. The motivation of this extension arises from the fact that in many real-world situations, such as the transmission of digital images, it is reasonable to attribute different values to different errors, contrasting with the indicator function that appears in the error probability definition. To attribute different values for each exchange of information, we shall replace the indicator function $\mu_{0\text{-}1}$ by a value function that may assume any (non-negative) real value.

Until now, the information set was identified with $\mathbb{F}_q^k$. From now on, we will assume that the information set is any set $\mathcal{I}$ with q^k elements. This is important since the identification with $\mathbb{F}_q^k$ may raise some confusion, because the measures that will be defined in this chapter are determined according to the nature of the information.

An *error value function* for the information set $\mathcal{I}$ is a map μ that associates to each pair of information (element of $\mathcal{I}$) a non-negative real number

$$\mu : \mathcal{I} \times \mathcal{I} \to \mathbb{R}_+$$

where $\mu(\iota_1, \iota_2)$ is the cost of exchanging ι_2 by ι_1. If $\mathcal{C}$ is a code and $f : \mathcal{I} \to \mathcal{C}$ is an encoder, we denote by

$$\mu_f : \mathcal{C} \times \mathcal{C} \to \mathbb{R}_+$$

the *error value function induced by the encoder* f, i.e., given $\iota_1, \iota_2 \in \mathcal{I}$,

$$\mu_f\left(f\left(\iota_1\right), f\left(\iota_2\right)\right) := \mu\left(\iota_1, \iota_2\right).$$

We shall refer to μ and μ_f as just an *error value function.* By considering such value function, we are interested in evaluating the errors that may occur during the process consisting of coding, transmitting and decoding. Therefore, it is reasonable to assume that μ has the following properties:

(a) μ is symmetric, i.e., $\mu(\iota_1, \iota_2) = \mu(\iota_2, \iota_1)$;

(b) $\mu(\iota, \iota) = 0$ for all $\iota \in \mathcal{I}$.

Given an encoding-decoding scheme $(\mathcal{C}, f, D)$ and an error value function μ, let us denote by $\mathbb{E}(\mathcal{C}, \mu_f, D)$ the *expected loss* of $(\mathcal{C}, f, D)$ with respect to μ, i.e.,

$$\mathbb{E}(\mathcal{C}, \mu_f, D) = \frac{1}{q^k} \sum_{c \in \mathcal{C}} \sum_{y \in \mathbb{F}_q^n} W(y|c) \sum_{c' \in \mathcal{C}} \mu_f(c, c') D(c'|y). \tag{4.2}$$

The only difference from this equation to Equation 4.1 is the exchanging of $\mu_{0\text{-}1}$ by μ_f. As noticed in the first chapter, the error probability of $\mathcal{C}$ does not depend on the choice of the encoder f. In other words, given a decoder and a channel, the error probability of $\mathcal{C}$ is invariant under permutations of the encoder f (which is a bijection between $\mathcal{I}$ and $\mathcal{C}$). Since μ is defined using the nature of the information set and the decoder uses only the characteristics of the code, for a given code $\mathcal{C}$, the expected loss is not invariant according to the choice of an encoder for $\mathcal{C}$. Therefore, a new variable (the encoder) is introduced when dealing with expected loss. In this case, the problem of minimizing both the error and refusal probabilities can be generalized to the problem of minimizing the expected loss and the refusal probability.

Definition 4.1.1. Given an information set $\mathcal{I}$ and a error value function μ, an encoding-decoding scheme $(\mathcal{C}^*, f^*, D^*)$ is an $(n,k)_q$*-Bayes scheme according to* μ if it satisfies

$$\mathbb{E}(\mathcal{C}^*, \mu_{f^*}, D^*) + P_{ref}^{D^*}(\mathcal{C}^*) = \min_{(\mathcal{C}, f, D)} \left(\mathbb{E}(\mathcal{C}, \mu_f, D) + P_{ref}^{D}(\mathcal{C})\right).$$

Note that we are not chosing previously the code $\mathcal{C}$ as in Definition 1.1.9. Unequal error protection uses basically two techniques in order to unequally protect errors:

adding more redundancy in some coordinates than others, or; increasing the decision region for some codewords; such strategies are respectively called bit-wise and message-wise unequal error protection; see [4] for more details. In [13], for a given code, it was noticed that protection against different kinds of errors depends not only on the decoder, but also upon the encoder. In that work, optimal decoders were characterized (optimal according to a concept called *separation*). Our concept of optimal decoders and encoders are going to be defined by decoders and encoders minimizing the expected loss according to a given error value function.

Similarly to what was done in the first chapter for error probability, considering a normalized error value function ($\mu(\iota_1, \iota_2) \leq 1$ for all $\iota_1, \iota_2 \in \mathcal{I}$), we can assume that the refusal probability is zero and that decoders are deterministic maps. Therefore, from this point forward, if μ is a loss function, then $\mu : \mathcal{I} \times \mathcal{I} \to [0,1]$ where $[0,1] \subset \mathbb{R}$.

Proposition 4.1.2. Given a code $\mathcal{C}$, for every decoder $D \in \mathcal{D}_{\mathbb{F}_q}(\mathcal{C})$ and encoder f, there exists a decoder $\widetilde{D} \in \mathcal{D}_{\mathbb{F}_q}(\mathcal{C})$ such that

$$\mathbb{E}(\mathcal{C}, \mu_f, \widetilde{D}) \leq P_{ref}^{D}(\mathcal{C}) + \mathbb{E}(\mathcal{C}, \mu_f, D)$$

and $P_{ref}^{\widetilde{D}}(\mathcal{C}) = 0$.

Proof. Given a decoder $D \in \mathcal{D}_{\mathbb{F}_q}(\mathcal{C})$, suppose there exists $y_0 \in \mathbb{F}_q^n$ such that $D(\infty|y_0) = 1$. Define a new decoder $D^* \in \mathcal{D}_{\mathbb{F}_q}(\mathcal{C})$ satisfying $D^*(y) = D(y)$ for all $y \neq y_0$ but $D^*(y_0)$ is a new distribution which does not refuse y_0, so $D^*(\infty|y_0) = 0$. Using the same reasoning of the Proposition 1.1.10, we get that

$$P_{ref}^{D}(\mathcal{C}) = P_{ref}^{D^*}(\mathcal{C}) + \frac{1}{q^k}\sum_{c\in\mathcal{C}} W(y_0|c)$$

and

$$\mathbb{E}(\mathcal{C}, \mu_f, D) = \mathbb{E}(\mathcal{C}, \mu_f, D^*) - \frac{1}{q^k}\sum_{c\in\mathcal{C}} W(y_0|c) \sum_{c'\in\mathcal{C}} \mu_f(c, c')D^*(c'|y_0).$$

Therefore,

$$\mathbb{E}(\mathcal{C}, \mu_f, D^*) + P_{ref}^{D^*}(\mathcal{C}) = P_{ref}^{D}(\mathcal{C}) + \mathbb{E}(\mathcal{C}, \mu_f, D) + \frac{1}{q^k}\sum_{c\in\mathcal{C}} W(y_0|c)\left(\sum_{c'\in\mathcal{C}} \mu_f(c, c')D^*(c'|y_0) - 1\right).$$

Since $\mu_f(c,c') \leq 1$ for every $c, c' \in \mathcal{C}$, then for every $c \in \mathcal{C}$,

$$\sum_{c' \in \mathcal{C}} \mu_f(c,c')D^*(c'|y_0) - 1 \leq 0.$$

Thus,

$$\mathbb{E}(\mathcal{C}, \mu_f, D^*) + P^{D^*}_{ref}(\mathcal{C}) \leq P^{D}_{ref}(\mathcal{C}) + \mathbb{E}(\mathcal{C}, \mu_f, D).$$

Therefore, if $P^{D^*}_{ref}(\mathcal{C}) = 0$, then $\widetilde{D} = D^*$, otherwise, applying the same arguments for D^* until we run out of refused elements, $\widetilde{D}$ may be constructed. □

Proposition 4.1.3. Let $\mathcal{C}$ be an $[n,k,\delta]_q$ linear code, $D \in \mathcal{D}_{\mathbb{F}_q}(\mathcal{C})$ be a decoder satisfying $P^D_{ref}(\mathcal{C}) = 0$ and f be an encoder for $\mathcal{C}$. Then, there exists a deterministic decoder g such that

$$\mathbb{E}(\mathcal{C}, \mu_f, g) \leq \mathbb{E}(\mathcal{C}, \mu_f, D).$$

Proof. If D is a stochastic map modeling a decoder for $\mathcal{C}$ with no refusals, then by 4.2,

$$\mathbb{E}(\mathcal{C}, \mu_f, D) = \frac{1}{q^k} \sum_{y \in \mathbb{F}_q^n} \left(\sum_{c' \in \mathcal{C}} \left[\sum_{c \in \mathcal{C}} W(y|c)\mu_f(c,c') \right] D(c'|y) \right)$$

where f is an encoder and μ is an error value function. For each $y \in \mathbb{F}_q^n$, choose $c_y \in \mathcal{C}$ such that

$$A_y = \sum_{c \in \mathcal{C}} W(y|c)\mu_f(c,c_y) \leq \sum_{c \in \mathcal{C}} W(y|c)\mu_f(c,c')$$

for every $c' \in \mathcal{C}$. Note that $\sum_{c' \in \mathcal{C}} D(c'|y) = 1$, then

$$\frac{1}{q^k} \sum_{y \in \mathbb{F}_q^n} A_y = \frac{1}{q^k} \sum_{y \in \mathbb{F}_q^n} A_y \sum_{c' \in \mathcal{C}} D(c'|y) = \frac{1}{q^k} \sum_{y \in \mathbb{F}_q^n} \sum_{c' \in \mathcal{C}} A_y D(c'|y) \tag{4.3}$$

$$\leq \frac{1}{q^k} \sum_{y \in \mathbb{F}_q^n} \sum_{c' \in \mathcal{C}} \sum_{c \in \mathcal{C}} W(y|c)\mu_f(c,c')D(c'|y) = \mathbb{E}(\mathcal{C}, \mu_f, D). \tag{4.4}$$

Denote by g the map defined by $g(y) = c_y$, then g is a deterministic decoder for $\mathcal{C}$. Furthermore,

$$\mathbb{E}(\mathcal{C}, \mu_f, g) = \frac{1}{q^k} \sum_{y \in \mathbb{F}_q^n} \sum_{c \in \mathcal{C}} W(y|c)\mu_f(c, g(y)) = \frac{1}{q^k} \sum_{y \in \mathbb{F}_q^n} A_y \leq \mathbb{E}(\mathcal{C}, \mu_f, D).$$

□

In order to minimize the sum of the expected loss and the refusal probability, similarly to what was done in the first chapter, Propositions 4.1.2 and 4.1.3 ensure we can assume that decoders are deterministic maps and their refusal probabilities are zero (note that error value functions are considered to be normalized). Therefore, the expected loss can be rewritten as follows

$$\mathbb{E}(\mathcal{C}, \mu_f, g) = \frac{1}{q^k} \sum_{y \in \mathbb{F}_q^n} \sum_{c \in \mathcal{C}} W(y|c)\mu_f(c, g(y)).$$

In a general setting, we consider the following data to be given:

(1) The error value function μ, determined by the nature of the information;

(2) The size of the code $\mathcal{C}$, determined by the size of the information set: $|\mathcal{I}|= q^k$;

(3) The information rate, determined by cost constraints;

(4) The channel model W, determined by physical conditions.

In such a setting, we say that the triple $(\mathcal{C}^*, f^*, g^*)$ is an $(n, k)_q$-*Bayes coding-decoding scheme* if

$$\mathbb{E}\left(\mathcal{C}^*, \mu_{f^*}, g^*\right) = \min_{(\mathcal{C}, f, g)} \mathbb{E}\left(\mathcal{C}, \mu_f, g\right)$$

where the minimum is taken over all encoding-decoding schemes for $\mathcal{I}$ over W.

We may consider each of the variables $\mathcal{C}$, f and g independently.

Definition 4.1.4. A decoder g^* is a *Bayes decoder* of the pair $(\mathcal{C}, f)$ if

$$\mathbb{E}(\mathcal{C}, \mu_f, g^*) = \min_{g} \mathbb{E}(\mathcal{C}, \mu_f, g).$$

Definition 4.1.5. An encoder f^* is a *Bayes encoder* of the pair $(\mathcal{C}, g)$ if

$$\mathbb{E}(\mathcal{C}, \mu_{f^*}, g) = \min_{f} \mathbb{E}(\mathcal{C}, \mu_f, g).$$

It is clear that a decoder g is a Bayes decoder if, and only if, for any $y \in \mathbb{F}_q^n$,

$$\sum_{c \in \mathcal{C}} W(y|c)\mu_f(c, g(y)) = \min \left\{ \sum_{c \in \mathcal{C}} W(y|c)\mu_f(c, c') \ : \ c' \in \mathcal{C} \right\}.$$

Note that if g is a Bayes decoder and μ is the indicator function $\mu_{0\text{-}1}$, then

$$\sum_{\substack{c\in\mathcal{C}\\ c\neq g(y)}} W(y|c) = \sum_{c\in\mathcal{C}} W(y|c)\mu_{0\text{-}1}(c,g(y)) \leq \sum_{c\in\mathcal{C}} W(y|c)\mu_{0\text{-}1}(c,c') = \sum_{\substack{c\in\mathcal{C}\\ c\neq c'}} W(y|c)$$

for every $c' \in \mathcal{C}$. Therefore,

$$W(y|g(y)) \geq W(y|c')$$

for every $c' \in \mathcal{C}$, i.e.,

$$W(y|g(y)) = \max\{W(y|c) \ : \ c \in \mathcal{C}\}.$$

In other words, g is an ML decoder.

4.2 Bayes Encoders

Even considering the situation when the code and the decoder are previously chosen, finding Bayes encoders may still be a difficult problem that, up to our knowledge, has not been explored in the literature. A characterization of Bayes encoders in a general setting depends on the loss function. We stress that for an $[n,k,\delta]_q$ code $\mathcal{C}$, the "complexity" to find a Bayes encoder is $q^k!$ (the number of encoders for $\mathcal{C}$). To estimate a qualitative measure of a proposed encoder without running over all possible encoders, the average expected loss may be used.

Lemma 4.2.1. Let g be a decoder of an $[n,k,\delta]_q$ linear code $\mathcal{C}$ and μ an error value function. Considering the function μ_{mean} where $\mu_{mean}(c,c) = 0$ for every $c \in \mathcal{C}$ and

$$\mu_{mean}(c_1,c_2) = \frac{\sum_f \mu_f(c_1,c_2)}{q^k!}$$

for every $c_1 \neq c_2$. Then, the average of the expected losses is the expected loss given by μ_{mean}, i.e.,

$$\mathbb{E}(\mathcal{C},\mu_{mean},g) = \frac{\sum_f \mathbb{E}(\mathcal{C},\mu_f,g)}{q^k!}.$$

Proof. It follows straight from the fact that

$$\frac{\sum_f \mathbb{E}(\mathcal{C},\mu_f,g)}{q^k!} = \frac{1}{q^k!}\sum_{c\in\mathcal{C}}\sum_{y\in\mathbb{F}_q^n} W(y|c)P(c)\sum_f \mu_f(g(y),c). \tag{4.5}$$

□

With the same statements of the previous Lemma, we have the following proposition:

Proposition 4.2.2. For every $c_1, c_2 \in \mathcal{C}$ with $c_1 \neq c_2$,

$$\mu_{mean}(c_1, c_2) = \frac{1}{q^k(q^k-1)} \sum_{s=1}^{t} |A_{j_s}| j_s$$

where $\{j_1, \ldots, j_t\}$ is the set of all possible real values assumed by μ and $A_{j_k} = \{(\iota_1, \iota_2) \in \mathcal{I} \times \mathcal{I} \ : \ \mu(\iota_1, \iota_2) = j_k\}$ for every $k \in [t]$.

Proof. For every $c_1, c_2 \in \mathcal{C}$ with $c_1 \neq c_2$,

$$\begin{aligned}\sum_f \mu_f(c_1, c_2) = \sum_{s=1}^{t} \sum_{\substack{f \\ \mu_f(x,y)=j_s}} j_s &= \sum_{s=1}^{t} (q^k-2)!\, |A_{j_s}| j_s \\ &= (q^k-2)! \sum_{s=1}^{t} |A_{j_s}| j_s,\end{aligned}$$

where $A_{j_s} = \{(\iota_1, \iota_2) \in \mathcal{I} \times \mathcal{I} \ : \ \mu(\iota_1, \iota_2) = j_s\}$. Therefore,

$$\frac{\sum_f \mu_f(c_1, c_2)}{q^k!} = \frac{1}{q^k(q^k-1)} \sum_{s=1}^{t} |A_{j_s}| j_s.$$

□

An information set, as any finite set, may be endowed with an additive group structure. Despite that, the group operation does not, in general, translate the significance of the information. When the information set may naturally be endowed with an additive group structure and the error value function is invariant by this operation, i.e., if μ satisfies $\mu(\iota_1 + \iota_3, \iota_2 + \iota_3) = \mu(\iota_1, \iota_2)$ for all $\iota_1, \iota_2, \iota_3 \in \mathcal{I}$, in [17], it was presented a characterization of the *linear-Bayes* encoders (the linear encoders minimizing among all linear encoders, the expected loss). If g is a decoder for $\mathcal{C}$, the decision regions for each $c \in \mathcal{C}$, i.e., $g^{-1}(c)$, determine a partition of $\mathbb{F}_q^n$, i.e.,

$$\mathbb{F}_q^k = \bigsqcup_{c \in \mathcal{C}} g^{-1}(c).$$

Then,

$$\mathbb{E}\left(\mathcal{C}, \mu_f, g\right) = \frac{1}{q^k} \sum_{c \in \mathcal{C}} \sum_{c' \in \mathcal{C}} \sum_{y \in g^{-1}(c')} \mu_f\left(c, c'\right) W\left(y|c\right).$$

Therefore,

$$\mathbb{E}\left(\mathcal{C},\mu_f,g\right)=\frac{1}{q^k}\sum_{(c,c')\in\mathcal{C}\times\mathcal{C}}G_g\left(c,c'\right)\mu_f\left(c,c'\right) \tag{4.6}$$

where

$$G_g\left(c,c'\right)=\sum_{y\in g^{-1}(c')}W\left(y|c\right). \tag{4.7}$$

If μ is invariant by translations and f is a linear encoder, then μ_f is also invariant by translations. Therefore,

$$\mathbb{E}\left(\mathcal{C},\mu_f,g\right)=\frac{1}{q^k}\sum_{(c,c')\in\mathcal{C}\times\mathcal{C}}G_g\left(c,c'\right)\mu_f\left(c-c',\mathbf{0}\right).$$

Thus, writing $u=c-c'$,

$$\mathbb{E}\left(\mathcal{C},\mu_f,g\right)=\frac{1}{q^k}\sum_{c\in\mathcal{C}}\sum_{u\in\mathcal{C}}G_g\left(c,c-u\right)\mu_f\left(u,\mathbf{0}\right)=\frac{1}{q^k}\sum_{u\in\mathcal{C}}\left(\sum_{c\in\mathcal{C}}G_g\left(c,c-u\right)\right)\mu_f\left(u,\mathbf{0}\right).$$

Proposition 4.2.3. [17] Let $\mathcal{C}=\{c_1,\dots,c_{q^k}\}$ be an $[n,k,\delta]_q$ linear code, g a decoder and μ an error value function invariant by translations. Suppose, without loss of generality,

$$\sum_{c\in\mathcal{C}}G_g(c,c-c_1)\geq\cdots\geq\sum_{c\in\mathcal{C}}G_g(c,c-c_{q^k}).$$

Then, f is a linear-Bayes encoder if, and only if,

$$\mu_f(c_1,\mathbf{0})\leq\cdots\leq\mu_f(c_{q^k},\mathbf{0}).$$

Even though the characterization obtained by Proposition 4.2.3 provides a simple way to construct linear-Bayes encoders, invariance by translation is an artificial condition for the majority of the applications. As we shall see, the general case is much harder.

Suppose $\mathcal{C}=\{c_1,\dots,c_{q^k}\}$, then

$$\mathbb{E}(\mathcal{C},\mu_f,g)=\frac{1}{q^k}\sum_{i=1}^{q^k}\sum_{j=1}^{q^k}G_g(c_i,c_j)\mu_f(c_i,c_j) \tag{4.8}$$

where G_g is as in (4.7). Using Equation (4.8), if $A^g=(a_{ij})$ and $B^f=(b_{ij})$ are matrices defined by

$$a_{ij} = G_g(c_i, c_j) \text{ and } b_{ij} = \mu_f(c_i, c_j),$$

then

$$\mathbb{E}(\mathcal{C}, \mu_f, g) = \frac{1}{q^k} Tr(A^g B^f)$$

where $Tr(.)$ is the matrix trace function.

Given an encoder f, if P is a permutation matrix (P is obtained by permuting the rows or columns of the identity matrix), then $H = PB^f P^T$ is a matrix constructed by using an encoder h, i.e., $H = B^h$. In addition, if f and h are two encoders, there is a permutation matrix P such that $B^h = PB^f P^T$. Therefore, we have the following theorem.

Theorem 4.2.4. Given a decoder g', an encoder f' and the corresponding matrices $A^{g'}$ and $B^{f'}$, then

$$\min_f \mathbb{E}(\mathcal{C}, \mu_f, g') = \frac{1}{q^k} \min_P \; Tr(A^{g'} P B^{f'} P^t),$$

where the minimum on the left side is over all permutation matrices.

The previous characterization brings the Bayes encoder problem into a class of well-studied problems, "minimize the trace of matrices". Let f be an encoder. If $\sigma : [q^k] \to [q^k]$ is a permutation (bijection) and $B^{f,\sigma}$ is the matrix with entries $B^{f,\sigma}_{ij} = \mu_f(c_{\sigma(i)}, c_{\sigma(j)})$, then the problem of finding a Bayes encoder is equivalent to the problem of finding a permutation σ of the codewords such that $A^g B^{f,\sigma}$ minimizes the trace (note that $B^{f,\sigma} = PB^f P^t$ for some permutation matrix P). It is clear that the search space has $q^k!$ elements (the number of permutations in a set with q^k elements). In the next section we will work with a particular (the simplest in terms of codes) case.

4.3 A particular Case

From now on, we will assume that the channel is a binary DSMC channel with conditional probabilities given by

$$W(y|c) = (1-p)^n \left(\frac{p}{1-p} \right)^{d_H(y,c)},$$

where y and c are words with length n, $0 \leq p \leq 1/2$ is the error probability of each symbol and d_H is the Hamming distance. Let us consider the unusual case of a code with no redundancy at all, i.e., let $\mathcal{C}$ be an $[n, n]_2$ code. In such a case, to find a Bayes encoder-decoder pair is equivalent to find a Bayes encoder, since there is a unique decision to be made: accept as true whatever the message you receive. Codes like those (without redundancy) are rare but still may be used, as we can see in the recent use, for image transmission, in the satellite CBERS-2 (http://www.cbers.inpe.br/ingles/). However, more than looking for possible applications, we believe that understanding this instance may be a key for the general case (where $\mathcal{C}$ has redundancy).

We stress that a permutation $\sigma \in \mathcal{S}_n$ acts on $\mathbb{F}_q^n$ by permuting the coordinates.

Proposition 4.3.1. If $\mathcal{C} = \mathbb{F}_q^n$ is a code without redundancy over a DSMC channel, then the expected loss is invariant over permutations of the coordinates.

Proof. Let $f : \mathcal{I} \to \mathcal{C}$ be an encoder minimizing the expected loss, without loss of generality, suppose $f(i_j) = c_j$. Let σ be a permutation within the code and denote $c_j^\sigma = (c_j^{\sigma(1)}, \ldots, c_j^{\sigma(n)})$. It is clear that there is a permutation τ over q^n such that $c_{\tau(j)} = c_j^\sigma$. Define $h : \mathcal{C} \to \mathcal{C}$ such that $h(c_j) = c_{\tau(j)}$, as $d_H(c_i, c_j) = d_H(c_{\tau(i)}, c_{\tau(j)})$ then $h \circ f$ is an encoder satisfying

$$\mathbb{E}(\mathcal{C}, \mu_f, g) = \mathbb{E}(\mathcal{C}, \mu_{h\circ f}, g)$$

□

As an immediate consequency,

Corollary 4.3.2. When $\mathcal{C} = \mathbb{F}_q^n$, there are at least $n!$ Bayes encoders.

Suppose $\mathcal{I} = \{i_0, i_1, \ldots, i_{2^n-1}\}$ for some integer n, so that $|\mathcal{I}|= 2^n$. Motivated by applications in image transmission, there is a natural normalized error value function μ in $\mathcal{I}$, namely:

$$\mu(i_s, i_t) = \frac{1}{(2^n-1)^2}|i_s - i_t|^2 = \frac{1}{(2^n-1)^2}|s-t|^2 \quad \text{for all } \ i_s, i_t \in \mathcal{I}. \tag{4.9}$$

Let $\mathcal{C} = \mathbb{F}_2^n$ and consider $\mathcal{C} = \{c^1, \ldots, c^{2^n}\}$ where the codewords are lexicographically ordered, i.e., if $c^i = (c_1^i, \ldots, c_n^i)$ and $c^j = (c_1^j, \ldots, c_n^j)$, then

$$i \leq j \iff \sum_{l=1}^{n} c_l^i 2^{l-1} \leq \sum_{l=1}^{n} c_l^j 2^{l-1}.$$

Setting $j_s = (s/(2^n-1))^2$ for every $s \in \{0,1,\dots,2^n-1\}$, the set $\{j_0, j_1, \dots, j_{2^n-1}\}$ is the image of μ. Thus,

$$|A_{j_s}| = 2(2^n - s)$$

for every $s \in [2^n - 1]$ and $|A_{j_0}| = 2^n$, where A_{j_s} was defined in Proposition 4.2.2. Therefore,

$$\begin{aligned}
\mathbb{E}(\mathcal{C}, \mu_{mean}, g) &= \frac{1}{2^n} \sum_{i=1}^{2^n} \sum_{\substack{j \neq i \\ j=1}}^{2^n} \frac{\sum_{s=1}^{2^n-1}(2^n-s)s^2}{2^{n-1}(2^n-1)^3} W(c^j|c^i) \\
&= \frac{(1-p)^n}{2^n} \sum_{i=1}^{2^n} \sum_{\substack{j \neq i \\ j=1}}^{2^n} \frac{\sum_{s=1}^{2^n-1}(2^n-s)s^2}{2^{n-1}(2^n-1)^3} \left(\frac{p}{1-p}\right)^{d_H(c^i,c^j)} \\
&= \frac{(1-p)^n}{2^n(2^n-1)^2} \sum_{i=1}^{2^n} \sum_{\substack{j \neq i \\ j=1}}^{2^n} \frac{2^{n-1}(2^n+1)}{3} \left(\frac{p}{1-p}\right)^{d_H(c^i,c^j)},
\end{aligned}$$

and the last equality holds because

$$\frac{1}{3} 2^{2n-2}(2^n-1)(2^n+1) = \sum_{s=1}^{2^n-1} (2^n - s)s^2.$$

An encoder f defined by $f(i_j) = c^{j+1}$ is called a *lexicographic encoder*. For such an encoder we have that

$$\mathbb{E}(\mathcal{C}, \mu_f, g) = \frac{(1-p)^n}{2^n(2^n-1)^2} \sum_{i=1}^{2^n} \sum_{\substack{j \neq i \\ j=1}}^{2^n} (i-j)^2 \left(\frac{p}{1-p}\right)^{d_H(c^i,c^j)}. \tag{4.10}$$

It is intuitive and quite obvious that $\mathbb{E}(\mathcal{C}, \mu_f, g)$ is independent on the encoder and decoder. For $p = 0$ and $p = 1/2$,

$$\mathbb{E}(\mathcal{C}, \mu_f, g) = \mathbb{E}(\mathcal{C}, \mu_{mean}, g). \tag{4.11}$$

Based on experimentations and the knowledge that the result is true in small dimensions (the Example 4.3.4 describe the 8-dimensional case), we have the following conjecture:

Conjecture 4.3.3.

$$\sum_{i=1}^{2^n} \sum_{\substack{j \neq i \\ j=1}}^{2^n} (i-j)^2 \left(\frac{p}{1-p}\right)^{d_H(c^i,c^j)} < \sum_{i=1}^{2^n} \sum_{\substack{j \neq i \\ j=1}}^{2^n} \frac{2^{n-1}(2^n+1)}{3} \left(\frac{p}{1-p}\right)^{d_H(c^i,c^j)}.$$

for every positive integer n and every $0 < p < 1/2$.

If the previous conjecture is true, then for every $0 < p < 1/2$ the expected loss according to f is smaller than the mean of the expected losses, i.e.,

$$\mathbb{E}(\mathcal{C}, \mu_f, g) < \mathbb{E}(\mathcal{C}, \mu_{mean}, g).$$

In the following, we shall describe a particular case (with $n = 8$), based on the 256 gray levels of the RGB color model, which fits well with the expected loss μ described in (4.9), in [18] two others error value functions were used in this context. In this case, we manage to conclude that the expected loss obtained by the lexicographic encoder is better than the mean of the expected losses.

Example 4.3.4. Suppose $n = 8$ and that $\mathcal{I} = \{0, 1, \dots, 255\}$ is the set of all gray colors in the RGB color model (black is represented by 0 and white by 255). Then,

$$\mu(i_s, i_t) = \frac{1}{255^2}|i_s - i_t|^2 = \frac{1}{255^2}|s - t|^2 \quad \forall\ i_s, i_t \in \mathcal{I}.$$

Suppose $\mathcal{C} = \{c^1, \dots, c^{256}\}$ and $f(i_j) = c^{j+1}$ a lexicographic encoder, for every $j \in [2^8]$. It follows that

$$\begin{aligned}\mathbb{E}(\mathcal{C}, \mu_{mean}, g) &= \frac{(1-p)^8}{16646400} \sum_{i=1}^{2^8} \sum_{\substack{j \neq i \\ j=1}}^{2^8} \frac{32896}{3} \left(\frac{p}{1-p}\right)^{d_H(c^i, c^j)} \\ &= \frac{(1-p)^8}{16646400} \left(\frac{67371008r}{3} + \frac{235798528r^2}{3} + \frac{471597056r^3}{3} + \frac{589496320r^4}{3} \right. \\ &\quad \left. + \frac{471597056r^5}{3} + \frac{235798528r^6}{3} + \frac{67371008r^7}{3} + \frac{8421376r^8}{3} \right)\end{aligned}$$

where $r = p/(1-p)$. On the other hand,

$$\begin{aligned}\mathbb{E}(\mathcal{C}, \mu_f, g) &= \frac{(1-p)^8}{16646400} \sum_{i=1}^{2^8} \sum_{\substack{j \neq i \\ j=1}}^{2^8} (i-j)^2 \left(\frac{p}{1-p}\right)^{d_H(c^i, c^j)} \\ &= \frac{(1-p)^8}{16646400} \left(6379776r + 40988416r^2 + 119295232r^3 + 195767040r^4 \right. \\ &\quad \left. + 193932032r^5 + 115625216r^6 + 38366976r^7 + 5462272r^8 \right).\end{aligned}$$

Therefore,

$$q(r) = \frac{\mathbb{E}(\mathcal{C}, \mu_f, g)}{\mathbb{E}(\mathcal{C}, \mu_{mean}, g)} =$$

$$\frac{3(24921 + 160111r + 465997r^2 + 764715r^3 + 757547r^4 + 451661r^5 + 149871r^6 + 21337r^7)}{32896(8 + 28r + 56r^2 + 70r^3 + 56r^4 + 28r^5 + 8r^6 + r^7)}$$

for every $0 < r < 1$. Since the derivative

$$q'(r) = \frac{62475(1+r)^6(28+56r+70r^2+56r^3+28r^4+8r^5+r^6)}{32896(8+28r+56r^2+70r^3+56r^4+28r^5+8r^6+r^7)^2},$$

is positive in the interval $(0,1)$, q is strictly increasing in this interval. If $r = 1$, then $p = 1/2$, therefore, by (4.11), $q(1) = 1$. Hence, $0 < q(r) < 1$ for all $r \in (0,1)$, thus for every $p \in (0, 1/2)$,

$$\mathbb{E}(\mathcal{C}, \mu_f, g) < \mathbb{E}(\mathcal{C}, \mu_{mean}, g).$$

As we will see below, we have reasons to believe that the encoder f as constructed is a Bayes encoder for the n-dimensional case. The graphic 4.1 represents, for the 8-dimensional case, the expected losses for p varying from 0 to 1/2. The lexicographic encoder is represented by the red line. The blue dots correspond to encoders randomly sampled from each p in the set $\{0, 0.005, 0.01, \ldots, 0.5\}$. This picture not only suggests that lexicographic encoders are optimal, but they are rare and there is a large concentration close to the mean encoder performance (the green line).

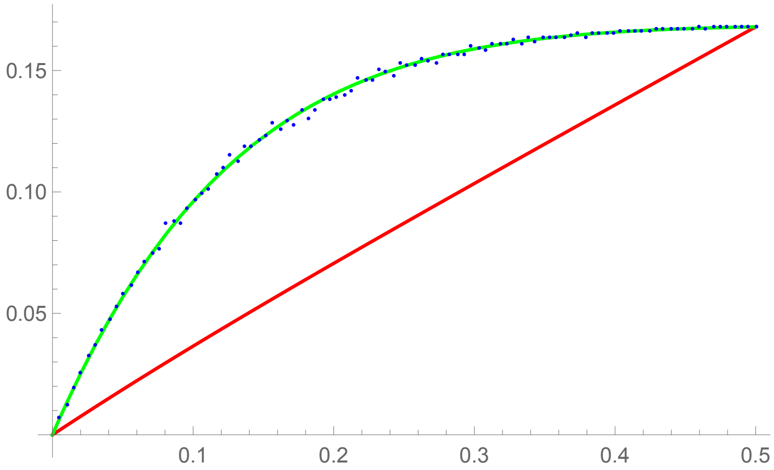


Figure 4.1: p in the interval $[0, 1/2]$.

The construction we did for the 8-dimensional case is similar for other dimensions, but proving that lexicographic encoders are optimal for general n and p is still an open problem.

4.4 Expected Loss and Unequal Error Protection

Given a decoder g, an encoder f and an error value function μ, in this section we propose a construction of codes with unequal error protection. The goal is to improve the expected loss of the code obtained by the direct product of two codes, each one "optimal" for the given parameters. Even though this construction may not be an optimal construction, we will see that in some cases the performance of this code, according to expected loss, is improved when compared to the product code.

Let $\mathcal{I} = \mathbb{F}_2^k$ be the information set. Consider the following data to be given:

- W is a binary DSMC;
- f is a lexicographic encoder;
- g is a minimum distance decoder determined by the Hamming metric d_H;

One of the existing formulations of unequal error protection (UEP) is the *bitwise UEP*, in this formulation, the coordinates (bits) of the information set are partitioned into subsets and the decoding errors in different parts of bits are viewed as different kind of errors. Since $\mathcal{I} = \mathbb{F}_2^k$, we will assume that k is a positive even integer and that the first $k/2$ coordinates are less important than the last $k/2$ coordinates. In other words, the last $k/2$ coordinates need more protection against errors than the first $k/2$ coordinates. One usual approach [4] is to encode each space $\mathbb{F}_2^{k/2}$ separately and take the Cartesian product of the codes. Suppose the low-priority bits (the fist $k/2$ coordinates) are encoded using an $[n_1, k/2, \delta_1]_2$ linear code $\mathcal{C}_1$ and that the high-priority bits are encoded with an $[n_2, k/2, \delta_2]_2$ linear code $\mathcal{C}_2$ with $n_2 > n_1$. Therefore, the code $\mathcal{C} = \mathcal{C}_1 \times \mathcal{C}_2$ has the unequal error protection property with more protection devoted to the more significant information. We will call such a code of a *2-levels product code* (in order to recall that the code was constructed using two levels of UEP).

There is a natural error value function associated with the 2-levels product code, the one expressing the difference between errors that may occur in the two distinct levels. Given $v = (v_1, \dots, v_k) \in \mathbb{F}_2^k$ and $v' = (v'_1, \dots, v'_k) \in \mathbb{F}_2^k$, let $r \in \mathbb{R}_+$ with $r < 1$ and let μ^r be an error value function defined by

$$\mu^r(v,v') = \begin{cases} 0 & \text{if} \quad v = v' \\ r & \text{if} \quad \sum_{i=1}^{k}(v_i - v_i')2^{i-1} < 2^{k/2+1} \\ 1 & \text{otherwise} \end{cases} .$$

This function outputs 1 if an error occurs in the high-priority bits and $r < 1$ for errors occurred in the low-priority bits. We recall that the number r is an invariant depending on the difference among errors occurred in the high-priority and low-priority bits.

In the information theory literature, it is common to reject a message if the errors are above some threshold. Since we are assuming decoders with refusal probability zero, it is more important to concentrate the errors in the less important information coordinates (where the error value is low), and this is achieved by protecting more the high-priority bits (where the error value is high). By using the concept of expected loss in this ambient, we may allow more errors provided that they are concentrated in the low-priority bits, therefore, in general, expected loss does not minimize the number of expected errors.

The Plotkin's construction is a well-known method to construct codes. In particular, the Reed-Muller family of codes are obtained using this construction, see [3]. Basically, given an $[n_1, k_1, \delta_1]_2$ code $\mathcal{C}_1$ and an $[n_2, k_2, \delta_2]_2$ code $\mathcal{C}_2$ and suppose $n_1 = n_2$, the *Plotkin construction* is the code $\mathcal{C}$ given by

$$\mathcal{C} = \mathcal{C}_1 * \mathcal{C}_2 = \{(u + v, v) \ : \ u \in \mathcal{C}_1 \text{ and } v \in \mathcal{C}_2\}.$$

Here we generalize this construction. Suppose $n_1 \leq n_2$, then consider an $n_2 \times n_1$ matrix $A = (a_{ij})$ such that $a_{ij} \in \mathbb{F}_2$. Hence, the *A-generalized Plotkin's construction* is given by

$$\mathcal{C} = \mathcal{C}_1 *_A \mathcal{C}_2 = \{(u + vA, v) \ : \ u \in \mathcal{C}_1 \text{ and } v \in \mathcal{C}_2\}.$$

When $n_1 = n_2$ and A is the identity matrix, the generalized construction coincides with the classical Plotkin's construction. When A is the null matrix, we have the Cartesian product of two codes which is denoted by $\mathcal{C}_1 \times \mathcal{C}_2$.

Suppose $k_1 = k_2$ and $n_1 < n_2$ (the information encoded in $\mathcal{C}_2$ is more protected). Note that the error value function μ^r is defined in such a way that errors occurred simultaneously in the high-priority and low-priority bits have the same error value of the errors occurred only in the high priority bits. In other words, the high-priority bits dominate the low-priority bits in terms of errors. Therefore, a good choice of a matrix A would be the matrix A having a maximal number of $v \in \mathcal{C}_2$ such that $vA \notin \mathcal{C}_1$. To do so, we will describe the construction of A used in our simulations which, we believe, due to the previous observations, it is a reasonable construction. Let $\beta = \{c_1, \ldots, c_{k_1}\}$ be a basis for $\mathcal{C}_1$ and consider

$$\beta' = \{c_1, \ldots, c_{k_1}, v_1, \ldots, v_{n_1-k_1}\}$$

a basis for $\mathbb{F}_2^{n_1}$ extended from β. If $k_1 > n_1 - k_1$, choose A having v_i as its i-th row and for every $j > n_1 - k_1$ the j-th row of A is null. Therefore, the image of $\mathcal{C}_2$ by A is the space generated by $\{v_1, \ldots, v_{n_1-k_1}\}$. If $k_1 \leq n_1 - k_1$ take v_i as the i-th row of A for every $i \leq k_1$ and consider the remaining rows to be null. In this case, the image of $\mathcal{C}_2$ by A is the subspace generated by $\{v_1, \ldots, v_{k_1}\}$. Using the A-generalized Plotkin's construction according to the matrix A previously constructed, we have the following conjecture:

Conjecture 4.4.1. Let $0 < p < 1/2$ be the error probability of the binary DSMC. For every $0 < r < 1$,

$$\mathbb{E}(\mathcal{C}_1 *_A \mathcal{C}_2, \mu^r_{f'}, g') \leq \mathbb{E}(\mathcal{C}_1 \times \mathcal{C}_2, \mu^r_{f''}, g'')$$

where f' and f'' are lexicographic encoders and g' and g'' are syndrome decoders according to the Hamming metric.

The *Best Known* linear code $\mathcal{C}_b$ is the code with length n and dimension k minimizing the error probability, among all known codes with the same parameters.

Conjecture 4.4.2. If $0 < p < 1/2$ is the error probability of the binary DSMC, then there exists $0 < r_0(p) < 1$ such that for every $r < r_0(p)$,

$$\mathbb{E}(\mathcal{C}_1 *_A \mathcal{C}_2, \mu^r_{f'}, g') \leq \mathbb{E}(\mathcal{C}_b, \mu^r_{f''}, g'').$$

Also, if $p \to 0$ then $r_0(p) \to 0$.

Besides the simulations suggesting the truthfulness of the two previous conjectures, the fact that the A-generalized Plotkin's construction uses the coordinates of $\mathcal{C}_1$

as redundancy for the code $\mathcal{C}_2$ suggests that the code $\mathcal{C}_2$ is even more protected. One of the experiments that strengthen the truthfulness of such conjectures will be described in the rest of this section as an example.

Example 4.4.3. Let $\mathcal{C}_1$ be a $[6,3,3]_2$ code and $\mathcal{C}_2$ be a $[10,3,5]_2$ code with generator matrices

$$G_1 = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} \text{ and } G_2 = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \end{bmatrix}$$

respectively. Let $\mathcal{C}_3$ be the best known linear code with parameters $[16,6,6]_2$. By using the Magma Computational Algebra System version 2.21-10, this code has generator matrix

$$G_3 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

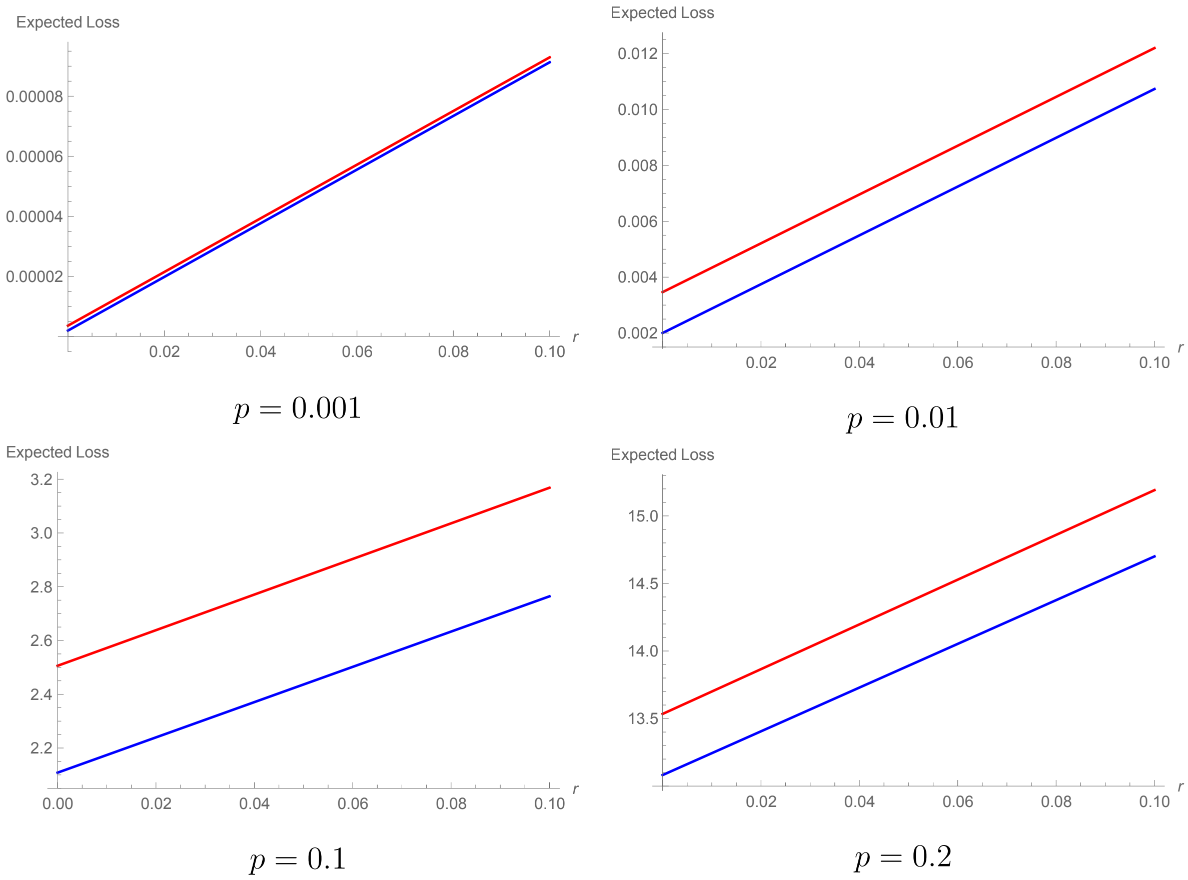
Consider the 10×6 matrix A given by

$$A = \left[\begin{array}{cccccc} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ \hline \multicolumn{6}{c}{\mathbf{0}_{7\times 6}} \end{array}\right]$$

where $\mathbf{0}_{7\times 6}$ is a null submatrix with order 7×6. Then, the code $\mathcal{C}_1 *_A \mathcal{C}_2$ has generator matrix

$$G_4 = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \end{bmatrix}$$

In this scenario, considering the error probability p of the binary DSMC to be 0.001, 0.01, 0.1 and 0.2 and considering r (r is the parameter given by the error value function μ^r) as a variable, we have the following graphics:



$p = 0.001$

$p = 0.01$

$p = 0.1$

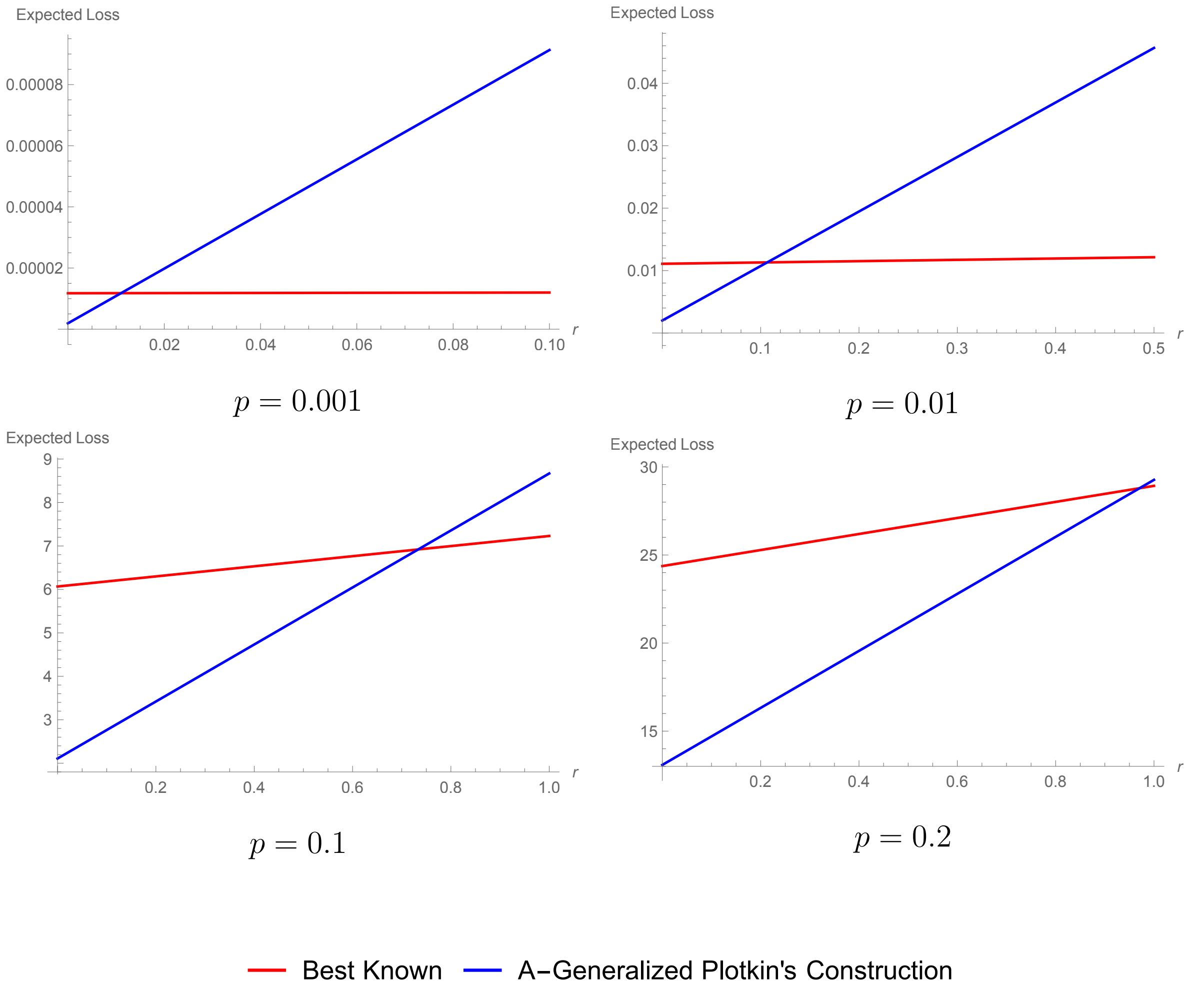
$p = 0.2$

Note that the previous graphs were plotted using $r \in (0, 0.1)$, since the for small error probability p, the lines are so close that we cannot distinguish them if we would plot for every r in the interval $(0, 1)$. Due to the characterization of μ^r, it is possible to prove that $\mathbb{E}(\mathcal{C}, \mu_f^r, g)$ is linear in r. Hence, by analyzing the inclination of these straight lines (what was done using Magma), we get that for every $p \in \{0.001, 0.01, 0.1, 0.2\}$ and $0 \leq r \leq 1$,

$$\mathbb{E}(\mathcal{C}_1 *_A \mathcal{C}_2, \mu_{f'}^r, g') \leq \mathbb{E}(\mathcal{C}_1 \times \mathcal{C}_2, \mu_{f''}^r, g'')$$

as expected. Furthermore, comparing the A-generalized Plotkin's construction with the best known linear code, using the same values of p, we get the following graphics:

🟥 Direct Product 🟦 A−Generalized Plotkin's Construction



$p = 0.001$

$p = 0.01$

$p = 0.1$

$p = 0.2$

In order to make visible the intersection points of lines in the graphs, the interval plotted for both the first and second graphs were reduced. Note that the 4 graphs suggest that the intersection point converges to zero if $p \to 0$. We stress that in this case we can find the exact point of intersection of the two straight lines since the dimension is small and can be computationally calculated.

FUTURE PERSPECTIVES

Extended Poset Metrics

By characterizing a big family of metrics, the description of all possible decoders with some characteristics can be solved, this would be very impressive if this family of metrics respect some particular construction, as in the poset metrics. Motivated by this characterization problem, we will propose an even more general family of metrics than the poset one.

Given a set Y, remark that any subset $X \subset \mathcal{P}(Y)$ of the powerset of Y is a poset with the inclusion relation. In the following, consider $X \subset \mathcal{P}(Y)$ such that $Y \subset \cup_{A \in X} A$. Remark that if $A, B \in X$ and $A \subset B$, B is said a covering of A if there is no $C \in X$ such that $A \subsetneq C \subsetneq B$. Given a poset $P = (X, \leqslant)$ over X, if $B \subset Y$ and I is an ideal in P such that $B \subset \cup_{A \in I} A$, I is a *cover ideal* of B if there is no ideal J in P such that $B \subset \cup_{A \in J} A$ and $|J| < |I|$. Denote by $N_P(B)$ the cardinality of the ideal covering B,

$$N_P(B) = \min\{|I| : \ I \text{ covers } B\}.$$

If I and J are cover ideals of B, then $|I| = |J|$, therefore the function N_P is well defined.

Proposition 4.4.4. $N_P(A \cup B) \leq N_P(A) + N_P(B)$

Proof. If I and J are ideals covering A and B respectively, then $I \cup J$ is an ideal satisfying $A \cup B \subset I \cup J$, then there is an ideal U covering $A \cup B$ such that $A \cup B \subset U \subset I \cup J$, therefore $N_P(A \cup B) \leq N_P(A) + N_P(B)$. □

Definition 4.4.5. With the notations above, supposing $Y = [n]$, if $x \in \mathbb{F}_q^n$, the *extended poset P-weight* is defined by being the smallest cardinality of the ideals covering the support of x,

$$w_P(x) = N_P(supp(x)).$$

Proposition 4.4.6. Extended poset weights preserve support.

Proof. If w_P is an extended poset weight and $x, y \in \mathbb{F}_q^n$ are vectors such that $supp(x) \subset supp(y)$, then every ideal covering $supp(y)$ is an ideal containing $supp(x)$, then $w_P(x) \leq w_P(y)$. □

The weight function w is clearly non-negative, furthermore, $w(x) = 0$ if, and only if, $x = \mathbf{0}$ since $[n] \subset \cup_{A \in X} A$. Given $x, y \in \mathbb{F}_q^n$, because $supp(x + y) \subset supp(x) \cup supp(y)$, by Propositions 4.4.4 and 4.4.6, it follows that $w(x+y) \leq w(x)+w(y)$. Therefore, the extended poset weight w_P is a norm function and metrics can be defined according to these norms.

Definition 4.4.7. Given an extended poset weight w_P, the *extended metric* (or extended P-metric) is the metric induced by w_P, i.e.,

$$d_P(x, y) = w_P(x - y)$$

for every $x, y \in \mathbb{F}_q^n$.

Despite the definition of these metrics is quite broad, these metrics possess a nice structure, indeed, they are invariant by translations since are defined by norms, and these norms preserve support, as stated in Proposition 4.4.6.

In [11], it was showed that, up to a decoding equivalence, any metric space may be embedded into a hypercube with the Hamming metric. Therefore, for decoding purposes, the Hamming metric can be always used, but the decoding complexity may inclease since the embedding immerse the code into a higher dimensional space. A natural question rises when working in this level of generality, among the metrics preserving support, does the family of extended poset metrics characterizes, up to decoding equivalence, any minimum distance decoder? We do not have the answer for this question but seems to be a fruitful line of research.

Better Hierarchical Bounds

The set $\mathcal{P}_n$ of all posets over $[n]$ is, by itself, a poset, as we saw in Chapter 2. If we consider the graph that has $\mathcal{P}_n$ as the set of vertices and an edge connects P to Q if, and only if, P covers Q or Q covers P, then we have a natural distance defined between two elements $P, Q \in \mathcal{P}_n$: $d\,(P, Q)$ is the minimal length of a path in the graph connecting P to Q. We remark that given the upper and lower neighbors P^- and P^+, in general, there may be a hierarchical poset Q with $d\,(P, Q) \leq d\,(P, P^-)$ and $d\,(P, Q) \leq d\,(P, P^+)$. It is possible that this finer notion of proximity of posets allow us to determine better bounds than the ones obtained by the upper and lower neighbors P^+ and P^-.

Bibliography

[1] Marcelo Muniz S Alves. “A standard form for generator matrices with respect to the Niederreiter-Rosenbloom-Tsfasman metric”. In: *Information Theory Workshop (ITW)*. IEEE. 2011, pp. 486–489.

[2] Marcelo Muniz S Alves, Luciano Panek, and Marcelo Firer. “Error-block codes and poset metrics”. In: *Advances in Mathematics of Communications* 2.1 (2008), pp. 95–111.

[3] D J Baylis. *Error Correcting Codes: A Mathematical Introduction*. Vol. 15. CRC Press, 1997.

[4] Shashi Borade, Bariş Nakiboğlu, and Lizhong Zheng. “Unequal error protection: An information-theoretic perspective”. In: *Information Theory, IEEE Transactions on* 55.12 (2009), pp. 5511–5539.

[5] Thomas Britz and Peter Cameron. “Partially ordered sets”. In: *J. of Formalized Mathematics* 1 (2002).

[6] Richard A Brualdi, Janine Smolin Graves, and K Mark Lawrence. “Codes with a poset metric”. In: *Discrete Mathematics* 147.1 (1995), pp. 57–72.

[7] Antonio Campello et al. “Perfect codes in the lp metric”. In: *European Journal of Combinatorics* 53 (2016), pp. 72–85.

[8] Sung Hee Cho and Dae San Kim. “Automorphism group of the crown-weight space”. In: *European Journal of Combinatorics* 27.1 (2006), pp. 90–100.

[9] Serban D Constantin and TRN Rao. “On the theory of binary asymmetric error correcting codes”. In: *Information and Control* 40.1 (1979), pp. 20–36.

[10] Michel Marie Deza and Elena Deza. *Encyclopedia of distances*. Springer, Berlin, 2009.

[11] Rafael GL D'Oliveira and Marcelo Firer. "Channel Metrization". In: *arXiv preprint arXiv:1510.03104* (2015).

[12] Rafael Gregorio Lucas D'Oliveira and Marcelo Firer. "The packing radius of a code and partitioning problems: The case for poset metrics on finite vector spaces". In: *Discrete Mathematics* 338.12 (2015), pp. 2143–2167.

[13] Larry A Dunning and Woodrow E Robbins. "Optimal encodings of linear block codes for unequal error protection". In: *Information and control* 37.2 (1978), pp. 150–177.

[14] Ben Dushnik and Edwin W Miller. "Partially ordered sets". In: *American Journal of Mathematics* (1941), pp. 600–610.

[15] Luciano Viana Felix and Marcelo Firer. "Canonical-systematic form for codes in hierarchical poset metrics". In: *Advances in Mathematics of Communications* 6.3 (2012), pp. 315–328.

[16] Marcelo Firer, Luciano Panek, and Jerry Anderson Pinheiro. "Coding and Decoding Schemes for MSE and Image Transmission". In: *arXiv preprint arXiv:1411.1139* (2014).

[17] Marcelo Firer, Luciano Panek, and Laura Rifo. "Coding in the presence of semantic value of information: Unequal error protection using poset decoders". In: *arXiv preprint arXiv:1108.3832* (2011).

[18] Marcelo Firer, Laura L Ramos Rifo, and Luciano Panek. "Coding and decoding schemes tailor made for image transmission". In: *Information Theory and Applications Workshop (ITA), 2013*. IEEE. 2013, pp. 1–8.

[19] Marcelo Firer and Judy L Walker. "Matched Metrics and Channels". In: *arXiv preprint arXiv:1506.03782* (2015).

[20] Ernst Gabidulin. "A brief survey of metrics in coding theory". In: *Mathematics of Distances and Applications* (2012), pp. 66–84.

[21] Ernst Gabidulin. "Combinatorial metrics in coding theory". In: *2nd International Symposium on Information Theory*. Akadémiai Kiadó. 1973.

[22] S Golomb. "A general formulation of error metrics (Corresp.)" In: *Information Theory, IEEE Transactions on* 15.3 (1969), pp. 425–426.

[23] Marcus Greferath et al. “MacWilliams’ Extension Theorem for Bi-Invariant Weights Over Finite Principal Ideal Rings”. In: *arXiv preprint arXiv:1309.3292* (2013).

[24] Jonathan I Hall. *Notes on coding theory.* FreeTechBooks. com, 2003.

[25] Richar W. Hamming. “Error detecting and error correcting codes”. In: *Bell System technical journal* 29.2 (1950), pp. 147–160.

[26] Abramo Hefez and Maria Lúcia T Villela. *Códigos corretores de erros.* Instituto de Matematica Pura e Aplicada, 2008.

[27] W Cary Huffman and Vera Pless. *Fundamentals of error-correcting codes.* Cambridge university press, 2003.

[28] C. Y. Lee. “Some properties of nonbinary error-correcting codes”. In: *IRE Transactions on Information Theory* 4.2 (1958), pp. 77–82.

[29] Kwankyu Lee. “The automorphism group of a linear space with the Rosenbloom–Tsfasman metric”. In: *European Journal of Combinatorics* 24.6 (2003), pp. 607–612.

[30] Roberto Assis Machado, Jerry Anderson Pineiro, and Marcelo Firer. “Characterization of metrics induced by hierarchical posets”. In: *arXiv preprint arXiv:1508.00914* (2015).

[31] Florence Jessie MacWilliams and Neil James Alexander Sloane. *The theory of error correcting codes.* Elsevier, 1977.

[32] Burt Masnick and Jack Wolf. “On linear unequal error protection codes”. In: *Information Theory, IEEE Transactions on* 13.4 (1967), pp. 600–607.

[33] James L. Massey. *Notes on coding theory, class notes for course 6.575 (spring).* M.I.T., Cambridge, MAA, 1967.

[34] Joseph Neggers and Hee Sik Kim. *Basic posets.* World Scientific, 1998.

[35] Harald Niederreiter. “A combinatorial problem for vector spaces over finite fields”. In: *Discrete Mathematics* 96.3 (1991), pp. 221–228.

[36] Harald Niederreiter. “Orthogonal arrays and other combinatorial aspects in the theory of uniform point distributions in unit cubes”. In: *Discrete mathematics* 106 (1992), pp. 361–367.

[37] Harald Niederreiter. “Point sets and sequences with small discrepancy”. In: *Monatshefte für Mathematik* 104.4 (1987), pp. 273–337.

[38] Luciano Panek, Marcelo Firer, and Marcelo Muniz Silva Alves. “Classification of niederreiter–rosenbloom–tsfasman block codes”. In: *Information Theory, IEEE Transactions on* 56.10 (2010), pp. 5207–5216.

[39] Luciano Panek et al. “Groups of linear isometries on poset structures”. In: *Discrete Mathematics* 308.18 (2008), pp. 4116–4123.

[40] Woomyoung Park and Alexander Barg. “Linear ordered codes, shape enumarators and parallel channels”. In: *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*. IEEE. 2010, pp. 361–367.

[41] Claudio Qureshi and Sueli IR Costa. “On perfect q-ary codes in the maximum metric”. In: *Information Theory and Applications* (2016).

[42] Gérald Séguin. “On metrics matched to the discrete memoryless channel”. In: *Journal of the Franklin Institute* 309.3 (1980), pp. 179–189.

[43] Danilo Silva and Frank R Kschischang. “On metrics for error correction in network coding”. In: *Information Theory, IEEE Transactions on* 55.12 (2009), pp. 5479–5490.

[44] Werner Ulrich. “Non-Binary Error Correction Codes”. In: *Bell System Technical Journal* 46.6 (1957), pp. 1341–1388.

[45] G Viswanath and B Sundar Rajan. “Matrix characterization of linear codes with arbitrary Hamming weight hierarchy”. In: *Linear algebra and its applications* 412.2 (2006), pp. 396–407.

Index