UNIVERSIDADE ESTADUAL DE MARINGÁ
CENTRO DE CIÊNCIAS EXATAS
DEPARTAMENTO DE MATEMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA
(Doutorado)

LUCIANO PANEK

CODIFICAÇÃO NA PRESENÇA DO VALOR SEMÂNTICO DA INFORMAÇÃO

Maringá - PR
2012

LUCIANO PANEK

CODIFICAÇÃO NA PRESENÇA DO VALOR SEMÂNTICO DA INFORMAÇÃO

Tese apresentada ao Programa de Pós-Graduação em Matemática do Departamento de Matemática, Centro de Ciências Exatas da Universidade Estadual de Maringá, como requisito parcial para a obtenção do título de Doutor em Matemática.

Orientador: Prof. Dr. Eduardo Brandani da Silva
Co-Orientador: Prof. Dr. Marcelo Firer

Maringá - PR
2012

Codificação na Presença do Valor Semântico da Informação

Luciano Panek

Tese apresentada ao Programa de Pós-Graduação em Matemática do Departamento de Matemática, Centro de Ciências Exatas da Universidade Estadual de Maringá, como requisito parcial para a obtenção do título de Doutor em Matemática pela *Comissão Julgadora* composta pelos membros:

Prof. Dr. Marcelo Firer
Universidade Estadual de Campinas (Presidente)

Prof. Dr. Anderson Clayton Alves Nascimento
Universidade de Brasília

Prof. Dr. Arnaldo Mandel
Universidade de São Paulo

Prof. Dr. Marcelo Muniz Silva Alves
Universidade Federal do Paraná

Prof. Dr. Reginaldo Palazzo Júnior
Universidade Estadual de Campinas

Aprovada em: 03 de fevereiro de 2012.
Local de defesa: Auditório do Departamento de Matemática - DMA, Bloco F-67, campus da Universidade Estadual de Maringá.

Agradecimentos

Agradeço inicialmente o Professor Doutor Marcelo Firer pela orientação.

Sou grato também a minha instituição de origem, a Universidade Estadual do Oeste do Paraná, Campus de Foz do Iguaçu, por permitir minha capacitação.

Agradeço o Programa de Pós-Graduação em Matemática da Universidade Estadual de Maringá pela oportunidade de desenvolver o meu doutorado ali. Em especial, agradeço o Professor Doutor Eduardo Brandani da Silva pelo apoio e pelo suporte no programa.

Agradeço também à Fundação Araucária pelo apoio financeiro.

Agradeço à Professora Doutora Laura Rifo pelo apoio fundamental no trabalho.

Agradeço ao Professor Doutor Cristiano Torezzan pelas frutíferas discussões na UNICAMP sobre proteção desigual de erros.

Agradeço ao meu irmão, Vanderson Martins do Rosario, pelo apoio incondicional criando o software que possibilitou as simulações ilustradas no decorrer desse trabalho. Foi emocionante dividir com você as primeiras imagens geradas pelo programa.

Agradeço à minha esposa Josiane pelo carinho e pela paciência.

Agradeço à minha família pelo apoio nos momentos difíceis.

Agradeço o meu pai, Carlos Alberto Panek (*in memorian*), que a tanto tempo já se foi, que nunca tive a oportunidade de conhecer, e que hoje estaria feliz junto comigo.

Resumo

No contexto clássico da Teoria da Informação todas as informações são consideradas igualmente importantes. Com esta hipótese de uniformidade, a confiabilidade do esquema de codificação é avaliada pela sua probabilidade de erro de decodificação.

Neste trabalho exploramos a possibilidade da codificação de informações com diferentes valores semânticos incorporando esses valores na análise de desempenho do codificador de canal e decodificador. O problema da busca por códigos que minimizam a quantidade de erros é trocado pelo problema da busca por triplas, de código, codificador de canal e decodificador, que minimizam a perda esperada relativa aos erros de decodificação com valor. Neste ambiente exploramos os decodificadores posets por desempenharem um tipo de proteção desigual de informação compatível com a idéia de informação com valor: as informações com maior valor são cercadas por informações de valores aproximados. Definições e resultados similares são também estabelecidos para constelações de sinais em espaços Euclidianos.

Abstract

Classical theorical framework for communication assumes that all information is equally important. With such uniformity assumption, reliability of a communication scheme is measured by the average probability of error over all possible messages to be transmitted.

In this work we explore possibilities for coding when information has different semantic values. We introduce a loss function that expresses the overall performance of a coding scheme for discrete channels and exchange the usual goal of minimizing the probability of error to that of minimizing the expected loss. In this environment we explore the possibilities of using poset decoders to make a message-wise unequal error protection, where the most valuable information is protected by placing in its proximity information words that differ by small valued information. Similar definitions and results are shortly presented also for signal constellations in Euclidean space.

Conteúdo

Lista de Figuras

Lista de Tabelas

Introdução

Os fundamentos da Teoria da Informação foram estabelecidos em 1948 por Claude Shannon no artigo "*A mathematical theory of communication*" ([39]). Shannon mostrou que para determinados canais é sempre possível estabelecer uma comunicação confiável usando esquemas de codificação, desde que a taxa de informação seja inferior a capacidade do canal. Como neste cenário todas as informações são consideradas igualmente importantes, a confiabilidade do esquema de codificação é avaliada pela sua probabilidade de erro de decodificação. Este fato é compatível com a seguinte observação de Shannon:

> "*The semantic aspects of communication are irrelevant to the engeneering problem. The significant aspect is that the actual message is one selected from a set of possible messages.*" - Claude Shannon, 1948 ([39]).

Outra característica da probabilidade de erro de decodificação, herdada da hipótese de que todas as informações são igualmente importantes, é que esta não depende do codificador de canal. Em outras palavras, a probabilidade de erro de decodificação é invariante pelo codificador de canal. Na prática a escolha do codificador de canal pode sim influenciar na percepção da mensagem: na Figura 2 temos a imagem "Hello World" (Figura 1) decodificada depois de ser transmitida usando dois codificadores de canal (para maiores detalhes, ver o Exemplo 2.2 do Capítulo 2).

Figura 1: Imagem original “Hello World”.

A diferença entre as imagens é nítida, embora ambas tenham aproximadamente a mesma quantidade de erros de decodificação. Shannon contorna esse problema demonstrando que assintóticamente existem códigos com taxas de informações não nulas tal que a probabilidade de erro tende a zero. Como a solução teórica de Shannon é inviável para as aplicações práticas, a estratégia adotada na literatura para minimizar os erros consiste em projetar em conjunto o codificador de fonte e o codificador de canal.

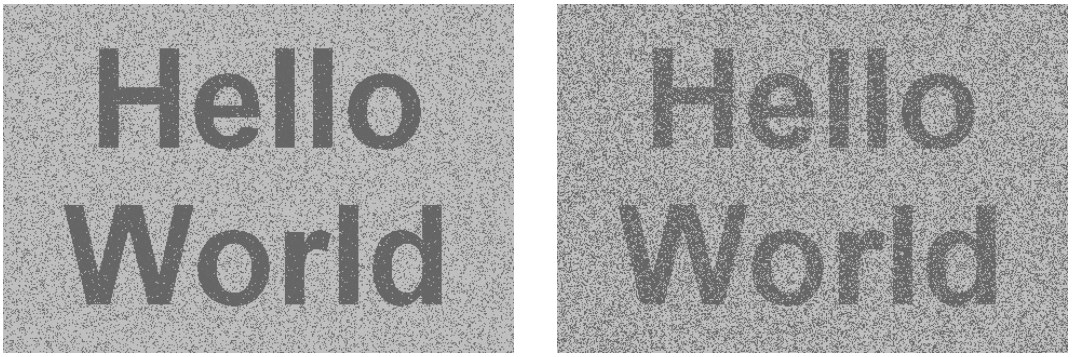

Figura 2: Simulação de uma transmissão com dois codificadores de canal.

Existem várias técnicas de codificação de fonte e canal conjunta (*Joint Source-Channel Coding* - JSCC) conhecidas na literatura especializada (ver [2]). Uma delas é a chamada *proteção desigual de erros* (*Unequal Error Protection* - UEP). Em proteção desigual de erros as informações são separadas em grupos de acordo com a importância de cada informação: por exemplo, em comunicações de redes sem fio (*network wireless*) os sinais de controle do estado do canal são mais relevantes do que os demais sinais (ver [5]).

Neste trabalho incorporamos o valor semântico da informação na análise de desempenho do decodificador e trocamos o problema da busca por pares de código e codificador de canal que maximizam o desempenho do sistema em relação a quantidade de erros pelo problema da busca por triplas, de código, codificador de canal e decodificador, que minimizam a perda relativa aos erros de decodificação com valor. Neste ambiente os decodificadores posets desempenham um papel fundamental protegendo as informações de acordo com o seu valor: as informações com maior valor são cercadas por informações de valores aproximados. Não é supreendente que neste contexto os decodificadores ML não são necessariamente os melhores decodificadores.

Vejamos um exemplo. Considere a imagem "Hello World" dada na Figura 1 acima. Vamos transmitir esta imagem usando uma fonte composta por quatro bits de informação, ou seja, estamos assumindo 16 possíveis tons de cinza na codificação. Vamos codificar as informações com o código binário de Hamming $[7; 4; 3]_2$. Para um determinado codificador de canal simulamos a transmissão da imagem "Hello World" pelo canal binário simétrico com probabilidade de erro $p = 0.4$. A imagem recebida foi decodificada de duas formas: uma usando um decodificador ML (à esquerda na Figura 3); a outra usando um decodificador determinado por uma ordem parcial P (à direita na Figura 3), que chamamos neste momento de P-decodifcador.

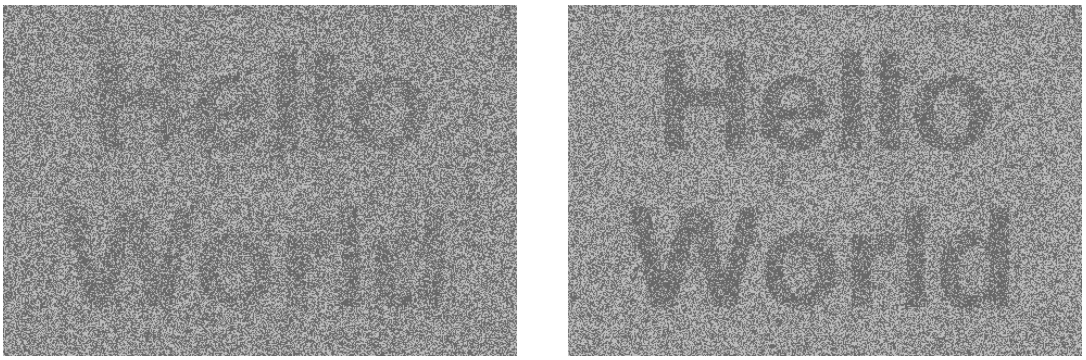

Figura 3: Simulação de uma transmissão com probabilidade de erro $p = 0.4$: à esquerda usando o decodificador ML; à direita usando o P-decodificador

Como podemos ver, a imagem dada pelo P-decodificador fornece uma melhor percepção da imagem original "Hello World" do que a imagem dada pelo decodificador ML, embora esta tenha menos erros de decodificação. A percepção em relação a qualidade das imagens decodificadas é um exemplo da forma como os valores devem ser atribuídos às informações. Este simples exemplo ilustra a nossa proposta de que o valor semântico da informação deve ser incorporado na análise de desempenho do decodificador para se obter melhores resultados do que os produzidos pelos decodificadores ML.

O trabalho está organizado em quatro capítulos. No Capítulo 1 apresentamos os conceitos e resultados básicos da Teoria da Informação. No Capítulo 2 apresentamos os principais conceitos e definições usados neste trabalho: função de valor (Definição 2.6), perda esperada total (Definição 2.7), codificador de Bayes (Definição 2.9) e decodificador de Bayes (Definição 2.9). Na Seção 2.5 estabelecemos a noção de taxa de informação confiável com valor (Definição 2.10) e apresentamos um limitante inferior para a capacidade do canal de transmitir informações com valor (Teorema 2.6). Na sequência apresentamos alguns resultados de existência, que asseguram que nem todo decodificador ML é candidato a decodificador de Bayes (Teorema 2.7 e Teorema 2.8). Posteriormente consideramos as diferenças das perdas esperadas para diferen-

tes decodificadores e caracterizamos estas diferenças como sendo funcionais lineares. A partir dessa caracterização estabelecemos uma condição para a existência de dois subconjuntos não vazios de funções de valor, um contendo funções de valor para as quais a diferença é negativa e outro contendo funções de valor para as quais a diferença é positiva (Teorema 2.9). Em seguida restringimos a definição de diferença entre perdas esperadas para canais q-ários simétricos e estabelecemos mais um resultado de existência (Teorema 2.10). Na Seção 2.9 adaptamos as definições de função de valor e perda esperada total para constelações de sinais sobre canais contínuos (Definições 2.13 e 2.14) e demonstramos que nem sempre os decodificadores determinados pelas regiões de Voronoi são os melhores decodificadores sobre um canal Gaussiano (Teorem 2.16). No Capítulo 3 apresentamos uma família de métricas que são compatíveis com a noção de informação com valor. Começamos apresentando as definições básicas de métrica poset (Definição 3.1) e decodificador P-NN (Definição 3.2). Na Seção 3.3 demonstramos que existe uma família de métricas poset para as quais sempre existe um par codificador-decodificador, em relação a esta família de métricas, que satisfazem as condições do Teorema 2.9 (Teorema 3.2). Na Seção 3.4 calculamos o raio de empacotamento dos códigos dado pelo Teorema 3.2 (Teorema 3.5). No último capítulo descrevemos os codificadores e os decodificadores usados para gerar as imagens ilustradas na Figuras 3 e mostramos que os codificadores de canais usados são de fato codificadores de Bayes para cada um dos decodificadores considerados. Encerramos o trabalho simulando mais algumas transmissões da imagem "Hello World" sobre canais binários simétricos, ilustrando as respectivas imagens dos erros de decodificação.

Todas as imagens ilustradas neste trabalho foram produzidas com o software desenvolvido por *Vanderson Martins do Rosario*, atualmente aluno do primeiro ano do curso de Ciência da Computação da Universidade Estadual de Maringá. O software, bem como sua descrição, encontram-se a disposição no endereço eletrônico

`http://code.google.com/p/error2image/`

Capítulo 1

Fundamentos da Teoria da Informação

A *Teoria da Informação* está fundamentada em problemas práticos de engenharia: *compactação*, *compressão* e *transmissão de dados*. Como disciplina, a Teoria da Informação é normalmente lecionada nos cursos de Engenharia Elétrica, sendo raramente difundida entre os alunos do curso de Matemática. Não é isto que acontece no campo da pesquisa: matemáticos e engenheiros trabalham arduamente no desenvolvimento da Teoria da Informação, sendo os matemáticos responsáveis por boa parte das contribuições em *Teoria dos Códigos*, componente essencial dentro de um modelo de sistema de comunicação.

Neste capítulo faremos uma introdução modesta e resumida dos principais resultados estabelecidos por Shannon em 1948. Estes resultados formam a base da Teoria da Informação e são essenciais para o bom entendimento das idéias que aparecerão ao longo do texto. As referências clássicas para este capítulo são os livros de Robert Gallager [14], Csiszár e Körner [11] e Thomas Cover [9]. Para uma leitura mais descontraída, indicamos o texto de Richard Hamming [16].

1.1 Fonte e Código de Fonte

Começamos enunciando o problema central da Teoria da Informação:

A que taxas podemos transmitir informações de forma confiável?

Classicamente este problema é abordado sem levar em conta o significado da informação transmitida[1]. O parâmetro adotado neste caso é a frequência relativa de ocorrência da informação. Esta estratégia permite uma formulação mais ampla da teoria. Podemos dividir o problema da transmissão de informação em dois subproblemas: o problema da representação mínima da informação (*compactação de dados*), em função da sua frequência, e o problema da transmissão confiável (*transmissão de dados*). Esta foi a estratégia adotada por Claude Shannon em 1948 em seu trabalho pioneiro *A Mathematical Theory of Communication* ([39]). Vamos agora formalizar os comentários deste parágrafo.

Definição 1.1 *Uma* ***fonte de informações discreta*** *é definida como sendo um par* $(\mathcal{S}, \mathcal{P})$ *onde* $\mathcal{S}$ *é um conjunto finito, o chamado* ***alfabeto de fonte****, e* $\mathcal{P} = \{P(s) : s \in \mathcal{S}\}$ *é uma distribuição de probabilidades* $P(s) := \Pr(s)$ *para* $\mathcal{S}$*.*

Note que a definição para fonte de informação não leva em consideração o significado dos *símbolos* $s \in \mathcal{S}$. Ela apenas requer o conhecimento da frequência $P(s)$ relativa a cada símbolo s.

Dada uma fonte de informações $(\mathcal{S}, \mathcal{P})$, definimos a *medida de incerteza* $I(s)$ (ou *medida de surpresa*, ou *medida de informação*) de $s \in \mathcal{S}$ como sendo o valor

$$I(s) := \log_2 \frac{1}{P(s)}.$$

A *entropia* é definida como sendo o valor esperado

$$H(\mathcal{S}) := \sum_{s \in \mathcal{S}} P(s) I(s),$$

[1] "*... semantic aspects of communication are irrelevant to the enginnering problem*" - Claude Shannon, 1948 ([39]).

a quantidade média de informação gerada pela fonte $(\mathcal{S}, \mathcal{P})$. A unidade adotada para $H(\mathcal{S})$ é o *bit*.

Existe uma estreita relação entre a entropia de uma fonte de informação e o problema da representação mínima (compactação de dados) dos elementos desta fonte, ou seja, como representar de forma econômica cada símbolo $s \in \mathcal{S}$ como uma sequência de elementos de um novo alfabeto. O alfabeto de representação considerado será o alfabeto binário, composto pelos elementos 0 e 1, já que estamos adotando o logaritmo na base 2 para a medida de incerteza $I(s)$. O alfabeto de representação, que no nosso caso é o alfabeto binário, também é conhecido como *alfabeto do código*. Diremos que C é um *código de fonte* se C é um subconjunto de $\{0,1\}^*$, o conjunto de todas as sequências finitas com entradas em $\{0,1\}$. Os elementos de C são chamados de *palavras-código*.

Dado um código de fonte C, se para cada $c = c_1 c_2 \dots c_n$ em C tivermos que $c_1 c_2 \dots c_k$ não pertence a C para todo $1 \leq k < n$, então diremos que C é um *código instantâneo*.

Um *esquema de codificação* para uma fonte de informação $(\mathcal{S}, \mathcal{P})$ é um par ordenado (C, f) onde C é um código de fonte e $f : \mathcal{S} \to C$ é uma aplicação bijetora de $\mathcal{S}$ em C.

Estamos prontos para enunciar o *Teorema da Codificação sem Ruído*. Este teorema assegura que toda representação mínima da fonte (no sentido que será estabelecido no enunciado do teorema) está limitada inferiormente pela sua entropia. Este resultado foi estabelecido por Shannon em 1948 ([39]).

Teorema 1.1 *Seja $l(c_s)$ o comprimento da palavra-código $f(s)$ e*

$$m(\mathcal{S}) := \min_{C} \sum_{s \in \mathcal{S}} P(s)\, l(c_s),$$

onde o mínimo é tomado sobre todos os esquemas de codificação (C, f) de $(\mathcal{S}, \mathcal{P})$ tal que C é instantâneo. Então

$$H(\mathcal{S}) \leq m(\mathcal{S}) < H(\mathcal{S}) + 1.$$

Vale a pena mencionar que existem métodos de construção para códigos instantâneos que satisfazem a condição de minimalidade estabelecida no Teorema 1.1.

Um dos métodos mais conhecidos na literatura é o *método de codificação de Huffman* (para maiores detalhes, ver [9] ou [16]).

Exemplo 1.1 *Seja* $\mathcal{S} = \{e, h, l, o\}$ *com* $P(e) = \frac{1}{8}$, $P(h) = \frac{1}{8}$, $P(l) = \frac{1}{2}$, $P(o) = \frac{1}{4}$. *Assuma que o esquema de codificação é dado por* $c_e = 111$, $c_h = 110$, $c_l = 0$, $c_o = 10$. *Os símbolos menos prováveis são representados por palavras de maior comprimento*: $I(e) = I(h) = 3$, $I(o) = 2$ *e* $I(l) = 1$. *Temos que* $C = \{c_e, c_h, c_l, c_o\}$ *é um código instantâneo. Mais ainda,*

$$H(\mathcal{S}) = \sum_{s \in \mathcal{S}} P(s)\, l(c_s) = 1.75 \text{ bits},$$

ou seja, C *é uma representação ótima para* $\mathcal{S}$ *como código instantâneo. Note agora que qualquer sequência de bits pode ser unicamente decodificada como sendo uma sequência de símbolos. Por exemplo, a mensagem*

$$1101110010$$

é decodificada como sendo

$$110, 111, 0, 0, 10 = hello.$$

Se tivéssemos considerado o código $\widetilde{C} = \{c_e = 111, c_h = 011, c_l = 0, c_o = 01\}$ *não poderíamos decodificar cada símbolo da mensagem "hello" instantaneamente*: *não podemos interpretar o primeiro dígito* 0 *da mensagem* 0111110001 *instantaneamente como sendo o símbolo* l, *já que o dígito* 0 *é prefixo das palavras-código* 0, 01 *e* 011.

É comum encontrarmos na literatura especializada a hipótese de que a distribuição de probabilidade $P(s)$ de uma fonte de informação $\mathcal{S}$ é *uniforme*, ou seja, $P(s) = \frac{1}{|\mathcal{S}|}$ para todo $s \in \mathcal{S}$. Neste trabalho também vamos assumir esta condição simplificadora.

Exemplo 1.2 *Seja* $(\mathcal{S}, \mathcal{P})$ *uma fonte de informação com distribuição de probabili-*

dade $\mathcal{P}$ uniforme. Assuma que $|\mathcal{S}| = 2^k$. Nestas condições

$$\begin{aligned} H(\mathcal{S}) &= \sum_{s\in\mathcal{S}} P(s)\, I(s) \\ &= \frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}} \log_2 |\mathcal{S}| \\ &= \log_2 2^k \\ &= k. \end{aligned}$$

Note agora que $\{0,1\}^k$ é um código de fonte instantâneo para $\mathcal{S}$ e que

$$m(\mathcal{S}) = \sum_{s\in\mathcal{S}} P(s)\, l(c_s) = \frac{1}{|\mathcal{S}|}\cdot|\mathcal{S}|\cdot k = k.$$

Como $m(\mathcal{S}) = H(\mathcal{S})$, concluímos que $\{0,1\}^k$ é uma representação minimal para a fonte $(\mathcal{S},\mathcal{P})$.

1.2 Canal Discreto e Informação Mútua

Na primeira seção consideramos o problema da compressão de dados para fontes de informações discretas. Agora trataremos do problema da transmissão de dados para canais discretos. A questão central aqui é: *qual a taxa máxima de bits por símbolo de entrada que podemos transmitir pelo canal de tal forma que a informação possa ser recuperada com baixa probabilidade de erro?*

O *canal de informação* é o meio (geralmente físico) pelo qual a informação é transmitida. Numa transmissão sem fio o canal pode ser a atmosfera (no caso do celular) ou o espaço sideral (no caso da transmissão de imagens via satélite). No caso do telefone, o canal usado são os cabos de fibra óptica. A informação transmitida pelo canal pode ser corrompida pelos variados tipos de ruídos relativos ao meio físico. Por exemplo: ruídos gerados pelos dispositivos eletrônicos, ruídos gerados pelas ignições de partida dos automóveis, ruídos gerados pelas descargas elétricas durante uma tempestade. Devido a natureza aleatória dos processos interferentes, os modelos teóricos adotados para os canais de informação são os modelos probabilísticos.

Vamos começar com o modelo de canal para fontes discretas.

Definição 1.2 *Um **canal discreto** (DC) é uma tripla* $(\mathcal{X}, P(\mathcal{Y}|\mathcal{X}), \mathcal{Y})$ *onde* $\mathcal{X} = \{x_1, \ldots, x_r\}$ *é o alfabeto de entrada,* $\mathcal{Y} = \{y_1, \ldots, y_s\}$ *é o alfabeto de saída e* $P(\mathcal{Y}|\mathcal{X})$ *é uma distribuição de probabilidades condicionais* $P(y_j|x_i)$*. Aqui* $P(y_j|x_i)$ *representa a probabilidade de recebermos* y_j *dado que* x_i *foi transmitido.*

Diremos que um canal discreto é *sem memória* (DMC - Discrete Memoryless Channel) se

$$P(y_1, \ldots, y_n|x_1, \ldots, x_n) = P(y_1|x_1) \cdot \ldots \cdot P(y_n|x_n)$$

para todo $n \in \mathbb{N}$, onde $x_1, \ldots, x_n$ e $y_1, \ldots, y_n$ representam n símbolos de entrada e saída respectivamente. *Neste trabalho assumiremos que todos os canais discretos são sem memória.* Assim, para nós, DC e DMC têm o mesmo significado.

Vamos listar agora alguns exemplos de canais discretos. Descreveremos estes exemplos considerando a *matriz do canal*

$$\mathbf{P} := \begin{pmatrix} P(y_1|x_1) & \cdots & P(y_s|x_1) \\ \vdots & \ddots & \vdots \\ P(y_1|x_r) & \cdots & P(y_s|x_r) \end{pmatrix}.$$

Começamos com os exemplos extremos: o *canal sem utilidade* e o *canal sem ruído.*

Exemplo 1.3 *Diremos que o canal é **sem utilidade** se as linhas de* $\mathbf{P}$ *são idênticas. Se para cada* $1 \leq j \leq s$ *tivermos* $P(y_j|x_l) = P(y_j|x_m)$ *para todo* $1 \leq l, m \leq r$*, então*

$$\begin{aligned} P(y_j) &= \sum_i P(y_j|x_i) P(x_i) \\ &= P(y_j|x_i) \sum_i P(x_i) \\ &= P(y_j|x_i). \end{aligned}$$

Em outras palavras, o canal é sem utilidade se os símbolos de saída não dependem dos símbolos de entrada, ou seja, não temos informação sobre os símbolos transmitidos.

Exemplo 1.4 *Em um* ***canal sem ruído*** *temos que* $\mathcal{X} = \mathcal{Y}$ *e* $\mathbf{P}$ *é a matriz identidade. Em outras palavras, estamos assumindo que o símbolo recebido é exatamente o símbolo transmitido*: $P(y_j|\, x_i) = 1$ *se, e somente se,* $x_i = y_j$.

O terceiro exemplo de canal discreto é um dos mais explorados na literatura especializada.

Exemplo 1.5 *O canal é dito* ***simétrico*** *se cada linha de* $\mathbf{P}$ *contém os mesmos elementos e se cada coluna de* $\mathbf{P}$ *também contém os mesmos elementos. Um canal é dito* q***-ário simétrico,*** *com* ***probabilidade de erro*** $0 < p < \frac{1}{2}$, *se este é simétrico,* $\mathcal{X} = \mathcal{Y} = \{0, 1, \ldots, q-1\}$ *e*

$$\mathbf{P} = \begin{pmatrix} 1-p & \cdots & \frac{p}{q-1} & \frac{p}{q-1} \\ \vdots & \ddots & & \frac{p}{q-1} \\ \frac{p}{q-1} & & \ddots & \vdots \\ \frac{p}{q-1} & \frac{p}{q-1} & \cdots & 1-p \end{pmatrix}.$$

Em um canal q*-ário simétrico a probabilidade de recebermos um símbolo incorretamente é igual a* p *e nesse caso cada um dos* $q-1$ *símbolos incorretos têm a mesma probabilidade* $\frac{1}{q-1}$ *de serem recebidos.*

Para um dado canal discreto $(\mathcal{X}, P(\mathcal{Y}|\, \mathcal{X}), \mathcal{Y})$ sabemos que $H(\mathcal{X})$ e $H(\mathcal{Y})$ medem a quantidade média de incerteza (ou informação) contidas nos alfabetos de entrada e saída $\mathcal{X}$ e $\mathcal{Y}$ respectivamente: se $P(x)$ é uma distribuição de probabilidades para $\mathcal{X}$, então

$$H(\mathcal{X}) = \sum_{x \in \mathcal{X}} P(x) \log_2 \frac{1}{P(x)}$$

e

$$H(\mathcal{Y}) = \sum_{y \in \mathcal{Y}} P(y) \log_2 \frac{1}{P(y)}$$

com

$$P(y) = \sum_{x \in \mathcal{X}} P(y|\, x)\, P(x).$$

A quantidade média de incerteza que resta sobre $\mathcal{X}$ depois que observamos $\mathcal{Y}$ é dada pela *entropia condicional*

$$H(\mathcal{X}|\mathcal{Y}) := \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P(x,y)\log_2\left(\frac{1}{P(x|y)}\right).$$

Aqui $P(x,y)$ denota a probabilidade conjunta entre x e y: $P(x,y) = P(x|y)\,P(y)$. A entropia condicional $H(\mathcal{X}|\mathcal{Y})$ também é conhecida na literatura como *equívoco*: para o canal sem ruído temos que $H(\mathcal{X}|\mathcal{Y}) = 0$, ou seja, não há equívoco durante a transmissão; para o canal sem utilidade temos que $H(\mathcal{X}|\mathcal{Y})$ é máximo, ou seja, $H(\mathcal{X}|\mathcal{Y}) = H(\mathcal{X})$, e neste caso o equivoco é total.

A quantidade média de informação de $\mathcal{X}$ que fica pelo canal durante uma transmissão é dada pela diferença entre a quandidade média de informação $H(\mathcal{X})$ de $\mathcal{X}$ e a quantidade média de incerteza $H(\mathcal{X}|\mathcal{Y})$ que resta sobre $\mathcal{X}$ depois de observado $\mathcal{Y}$. Esta diferença,

$$I(\mathcal{X};\mathcal{Y}) := H(\mathcal{X}) - H(\mathcal{X}|\mathcal{Y}),$$

é chamada *informação mútua média* entre $\mathcal{X}$ e $\mathcal{Y}$. A informação mútua média desempenha um papel crucial na resposta para a questão estabelecida no início desta seção: *qual a taxa máxima de bits por símbolo de entrada que podemos transmitir pelo canal de tal forma que a informação possa ser recuperada com baixa probabilidade de erro?* Vejamos os detalhes.

1.3 Capacidade do Canal

Seja $\mathcal{I}$ um conjunto de informações contendo M elementos. Para transmitirmos as informações de $\mathcal{I}$ pelo canal $(\mathcal{X}, P(\mathcal{Y}|\mathcal{X}), \mathcal{Y})$ associamos cada informação $i \in \mathcal{I}$ a uma única *palavra-código*

$$\mathbf{c}(i) = (c_1(i), \ldots, c_N(i))$$

de $\mathcal{X}^N$, o espaço de todas as N-uplas sobre $\mathcal{X}$, assumindo que $\mathbf{c}(i) \neq \mathbf{c}(j)$ se $i \neq j$. Diremos que $C = \{\mathbf{c}(i) : i \in \mathcal{I}\}$, o conjunto de todas as palavras-código, é um (M,N) *código de bloco* (ou simplesmente, um (M,N) *código*). A aplicação

injetora $f : \mathcal{I} \to C$ dada por $f(i) = \mathbf{c}(i)$ é chamada *codificador de canal* de C, ou simplesmente, o *codificador*. A *taxa de informação* para C é definida como sendo

$$R := \frac{\log_2 M}{N}.$$

Um *decodificador* para C é uma aplicação sobrejetora

$$a : \mathcal{Y}^N \to C$$

que gera uma estimativa $a(\mathbf{y})$ em C para cada $\mathbf{y} = (y_1, \ldots, y_N)$ em $\mathcal{Y}^N$. A *probabilidade de erro de decodificação* é dada pela média

$$P_e(C) = \sum_{\mathbf{c} \in C} P_e(\mathbf{c}) P(\mathbf{c})$$

das probabilidades de erros de decodificação

$$P_e(\mathbf{c}) = \sum_{\mathbf{y} \notin a^{-1}(\mathbf{c})} P(\mathbf{y}|\,\mathbf{c}) = 1 - \sum_{\mathbf{y} \in a^{-1}(\mathbf{c})} P(\mathbf{y}|\,\mathbf{c})$$

de cada palavra-código $\mathbf{c}$ de C. Aqui $P(\mathbf{c})$ denota a probabilidade de $\mathbf{c}$ ser enviada.

Definição 1.3 *Um número real $R > 0$ é uma* ***taxa de informação confiável***[2] *se para todo $\delta > 0$ existe um código C com taxa de informação R tal que $P_e(C) < \delta$.*

O problema "*Qual a taxa máxima de bits por símbolo de entrada que podemos transmitir pelo canal de tal forma que a informação possa ser recuperada com baixa probabilidade de erro?*" pode ser agora reformulado como: *qual é a máxima taxa de informação confiável possível?*

Definição 1.4 *A máxima taxa de informação confiável para um canal discreto é definida como sendo a* ***capacidade do canal****.*

A resposta do problema acima para canais discretos foi estabelecida por Shannon em seu trabalho pioneiro ([39]). Este é o conhecido *Teorema Fundamental da Teoria da Informação.*

[2]Esta definição é relativa aos decodificadores MAP, que serão definidos na Seção 2.1 do Capítulo 2.

Teorema 1.2 (Shannon - 1948) *Seja* $(\mathcal{X}, P(\mathcal{Y}|\mathcal{X}), \mathcal{Y})$ *um canal discreto e* $\mathcal{C}$ *sua capacidade. Então*

$$\mathcal{C} = \max_{P(x)} I(\mathcal{X};\mathcal{Y})$$

onde o máximo é tomado sobre todas as possíveis distribuições de probabilidade $P(x)$ *de* $\mathcal{X}$*.*

Os argumentos originais de Shannon para a prova do Teorema Fundamental da Teoria da Informação são pouco formais, porém muito intuitivos. Uma demonstração completa deste resultado pode ser encontrada em [9].

Encerramos esta seção calculando a capacidade do canal q-ário simétrico. Para tanto usaremos o fato de que

$$I(\mathcal{X};\mathcal{Y}) = H(\mathcal{Y}) - H(\mathcal{Y}|\mathcal{X}).$$

Exemplo 1.6 *Seja* $(\mathcal{X}, P(\mathcal{Y}|\mathcal{X}), \mathcal{Y})$ *um canal* q*-ário simétrico com probabilidade de erro* p*. Nestas condições temos que*

$$P(x_i, y_j) = \begin{cases} (1-p)\cdot P(x_i) & se \quad i = j \\ \dfrac{p}{q-1}\cdot P(x_i) & se \quad i \neq j \end{cases}$$

e daí que

$$\sum_i P(x_i, y_i)\log_2\left(\frac{1}{P(y_i|x_i)}\right) = (1-p)\log_2\left(\frac{1}{1-p}\right)$$

e

$$\sum_{i\neq j} P(x_i, y_j)\log_2\left(\frac{1}{P(y_j|x_i)}\right) = p\log_2\left(\frac{q-1}{p}\right).$$

Juntando estas informações obtemos que

$$H(\mathcal{Y}|\mathcal{X}) = (1-p)\log_2\left(\frac{1}{1-p}\right) + p\log_2\left(\frac{q-1}{p}\right).$$

Note agora que $H(\mathcal{Y}|\mathcal{X})$ *não depende da distribuição de probabilidades* $P(x)$ *de* $\mathcal{X}$*. Como* $H(\mathcal{Y})$ *depende da distribuição* $P(x)$ *de* $\mathcal{X}$*,* $I(\mathcal{X};\mathcal{Y})$ *é máximo quando* $H(\mathcal{Y})$

é máximo. Agora $H(\mathcal{Y})$ *é máximo* $(= \log_2 q)$ *quando* $P(y) = \frac{1}{q}$ *para todo* $y \in \mathcal{Y}$*, ou equivalentemente, quando* $P(x) = \frac{1}{q}$ *para todo* $x \in \mathcal{X}$*. Disto tudo concluímos que*

$$\mathcal{C} = \log_2 q + (1-p)\log_2(1-p) + p\log_2\left(\frac{p}{q-1}\right).$$

Em particular, para um canal binário simétrico

$$\mathcal{C} = 1 + (1-p)\log_2(1-p) + p\log_2 p.$$

1.4 Canais Contínuos

Quando assumimos que o canal é discreto estamos agrupando algumas etapas do sistema de comunicação, analisando apenas as probabilidades dos símbolos recebidos em função dos símbolos transmitidos. No entanto cada símbolo de entrada é associado a um sinal para na sequência ser transmitido pelo canal. Este sinal, corrompido pelo ruído, chega ao destino distorcido. Neste momento o decodificador de sinal toma uma decisão, entregando ao demodulador uma estimativa do sinal transmitido. O canal discreto é o agrupamento das etapas modulador-canal-demodulador. A análise separada destas três etapas dá origem ao chamado *canal contínuo.*

Definição 1.5 *Um* ***canal contínuo*** *é uma tripla* $(\mathcal{X}, P(\mathcal{Y}|\mathcal{X}), \mathcal{Y})$ *onde* $\mathcal{X}$ *é um subconjunto de* $\mathbb{R}$*,* $\mathcal{Y} = \mathbb{R}$ *e* $P(\mathcal{Y}|\mathcal{X})$ *é uma família de funções de densidade de probabilidade* $p(y|x)$*. Vamos assumir que todos os canais contínuos são* ***sem memória****, ou seja,*

$$p(y_1 \ldots y_N | x_1 \ldots x_N) = p(y_1|x_1) \cdot \ldots \cdot p(y_n|x_n),$$

com $x_i \in \mathcal{X}$ *e* $y_i \in \mathcal{Y}$*, para todo* n*.*

Exemplo 1.7 *Um* ***canal Gaussiano*** *é um canal contínuo com funções de densidade de probabilidade*

$$p(y|x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y-x)^2}{2\sigma^2}\right).$$

Consequentemente, como estamos assumindo que todo canal é sem memória,

$$p\left(y_1 \ldots y_N \middle| x_1 \ldots x_N\right) = \frac{1}{\left(2\pi\sigma^2\right)^{\frac{N}{2}}} \exp\left(-\sum_{i=1}^{N} \frac{\left(y_i - x_i\right)^2}{2\sigma^2}\right).$$

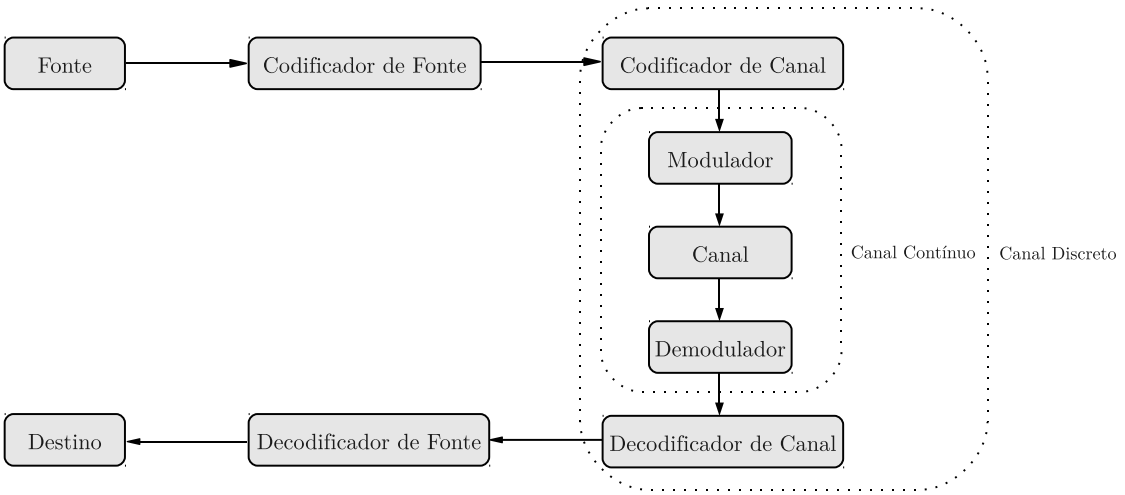

Figura 1.1: Elementos básicos de um sistema de comunicação.

Uma (M, N) *constelação de sinais* para um canal contínuo $\left(\mathcal{X}, P\left(\mathcal{Y}\middle|\mathcal{X}\right), \mathcal{Y}\right)$ é um subconjunto finito S de $\mathcal{X}^N$ contendo M elementos juntamente com uma aplicação sobrejetora $f : \mathcal{Y}^N \to S$, o *decodificador de canal*. Os elementos de uma constelação S serão chamados de *sinais*.

O desempenho de uma constelação S é medido pela sua *probabilidade* $P_e(S)$ *de erro de decodificação*:

$$P_e(S) = \sum_{\mathbf{s} \in S} P_e(\mathbf{s}) P(\mathbf{s})$$

onde $P(\mathbf{s})$ é a *probabilidade de transmitirmos* $\mathbf{s}$ e

$$P_e(\mathbf{s}) = 1 - \int_{\mathbf{y}: f(\mathbf{y}) = \mathbf{s}} p\left(\mathbf{y}\middle|\mathbf{s}\right) dy$$

é a probabilidade do decodificador entregar para o destino um sinal distinto do sinal transmitido $\mathbf{s}$.

Considerando as definições apropriadas de entropia e informação mútua para canais contínuos (para maiores detalhes ver [14]), é possível mostrar que

$$C = \frac{1}{2} \ln \left(1 + \frac{\sigma_x^2}{\sigma_\eta^2} \right)$$

é a máxima taxa de informação confiável para um canal Gaussiano, onde σ_η^2 é a variância do ruído Gaussiano e σ_x^2 é a variância do sinal transmitido x dado pela distribuição Gaussiana de média zero.

1.5 Proteção Desigual de Erros

Atualmente a Teoria da Informação divide os sistemas de comunicação em duas classes: a classe dos *sistemas ponto-a-ponto* e a classe dos *sistemas multiusuários*. O Teorema Fundamental da Teoria da Informação de Shannon diz respeito somente aos sistemas de comunicação ponto-a-ponto. Para estes sistemas o codificador de fonte e o codificador de canal são projetados separadamente sem comprometer a confiabilidade da transmissão. Por este motivo o Teorema de Shannon também é conhecido como *Teorema da Separação*.

Para os sistemas multiusuários a situação é bem mais delicada. Vamos considerar o exemplo dos *sistemas de transmissão de imagens de alta resolução* (HDTV - *High Definition Television*), que pertencem a classe dos *canais de radiodifusão*. Num canal de radiodifusão um único transmissor difunde a informação para vários receptores. No nosso exemplo, do sistema HDTV, o transmissor é a emissora de televisão e os receptores são os televisores em nossas residências. Se não levarmos em conta as diferentes distâncias e os diferentes relevos entre as casas e a emissora, é provável que sob condições severas de mau tempo as residências mais afastadas não recuperem o sinal (Figura 1.2). A solução ótima neste caso, descrita por Cover em [8], é obtida quando separamos os bits de informações em classes com níveis diferenciados de proteção: a imagem de baixa resolução deve ser mais protegida do que a imagem de alta resolução, assegurando aos usuários mais distantes no mínimo a recepção da imagem de baixa resolução (veja também [35]).

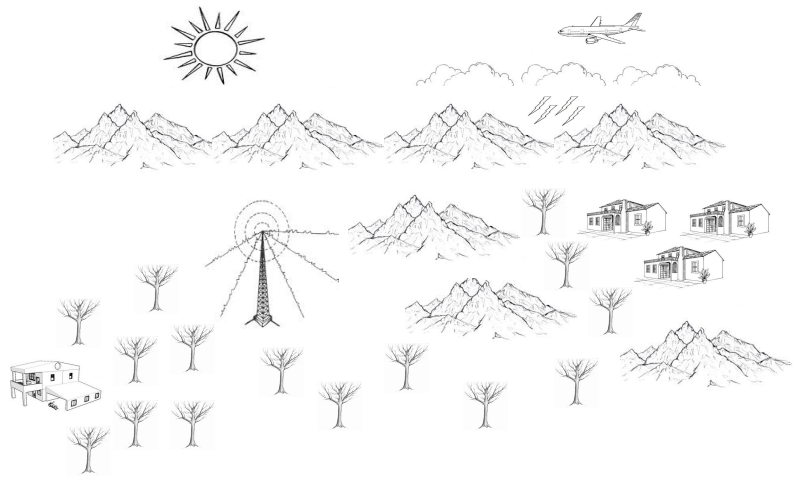

Figura 1.2: Transmissão do sinal de HDTV para muitas residências.

Os comentários acima mostram que a abordagem heterogênea é mais compatível com a natureza dos sistemas multiusuários do que a abordagem clássica onde as informações são consideradas homogêneas e o canal é estacionário. A estratégia adotada na literatura para otimizar o desempenho de um sistema multiusuário consiste em projetar em conjunto o codificador de fonte e o codificador de canal.

Existem várias técnicas para *codificação de fonte e canal conjunta* conhecidas na literatura (JSCC - *Joint Source-Channel Coding*) (ver [2]). Destacamos duas destas técnicas: *proteção desigual de bit* (bit-wise Unequal Error Protection - BUEP) e *proteção desigual de informação* (message-wise Unequal Error Protection - MUEP). A grosso modo, em BUEP os bits mais protegidos são aqueles com menores probabilidades de erro de decodificação de bit (BER - *Bit Error Rate*) e em MUEP as informações mais protegidas são aquelas com menores probabilidades de erro de decodificação de mensagem (WEP - *Word Error Probability*).

As técnicas de proteção desigual de erros foram inicialmente exploradas por Masnick e Wolf em [26]. Na formulação original de Masnick e Wolf os bits de informação são separados em níveis de proteção, e os erros de decodificação em cada nível são avaliados de forma diferenciada. Também podemos encontrar na literatura trabalhos onde as informações são separadas em grupos de bits, cada grupo com um nível de proteção (ver por exemplo [27]). Mais ainda, o assunto se estende para a

classe dos códigos convolucionais (ver por exemplo [42]). No contexto da Teoria da Informação temos o trabalho de Csiszár [10], sobre proteção desigual de informação, e mais recentemente o trabalho [5] de Borade, Nakiboğlu e Zheng sobre BUEP e MUEP.

Neste trabalho estamos interessados no problema da transmissão ponto-a-ponto considerando agora o valor semântico da informação. No nosso caso a fonte de informação é caracterizada como sendo uma tripla $(\mathcal{S}, \mathcal{P}, \mathcal{V})$ onde $(\mathcal{S}, \mathcal{P})$ é uma fonte de informações no sentido clássico e $\mathcal{V}$ é um conjunto de valores semânticos[3] para $\mathcal{S}$. Mais precisamente, $\mathcal{V}$ é uma aplicação do tipo

$$\mathcal{V} : \mathcal{S} \to \mathbb{R}_+$$

onde $\mathbb{R}_+$ é o conjunto dos números reais não-negativos. Na nossa abordagem a informação será protegida de acordo com o seu valor semântico impondo que o codificador de canal e o decodificador minimizem a perda esperada relativa a transmissão de informação com valor. Como veremos, não é surpreendente que as melhores estratégias para uma transmissão ponto-a-ponto homogênea (onde todas as informações têm o mesmo valor) não são necessariamente as melhores estratégias para uma transmissão ponto-a-ponto heterogênea (onde as informações têm valores distintos).

[3]No trabalho [23] Juba e Sudan consideram a semântica da informação. No nosso caso estamos apenas considerando a possibilidade de atribuir um valor para a semântica, o que deve ser estabelecido pelos especialistas das diferentes fontes de informações.

Capítulo 2

Funções de Valor e Perdas Esperadas

Na formulação original de Shannon todas as informações são igualmente importantes: o codificador de canal não é relevante para a análise de desempenho e não há distinção entre os diferentes tipos de erros de decodificação. Neste trabalho propomos uma mudança nesta formulação:

Considerar os valores semânticos das informações
e incorporar estes à Teoria da Codificação.

Como já podemos imaginar, os esquemas convencionais para transmissões de informações não serão necessariamente os melhores esquemas neste nova formulação, já que agora as informações não são mais igualmente concebidas e igualmente protegidas. O grande desafio consiste em estabelecer estruturas matemáticas que sejam compatíveis com essa nova abordagem e que também sejam computacionalmente viáveis.

Neste capítulo estabeleceremos precisamente a noção de *valor semântico da informação* e introduziremos o conceito de *perda esperada total*, instrumento que medirá o desempenho dos sistemas de comunicação na presença das informações com valores. Como veremos, os conceitos de decodificadores MAP, ML e NN são a base para estas definições. Por este motivo abriremos o capítulo revendo estes conceitos.

2.1 Decodificadores MAP, ML e NN

Começamos este capítulo revendo os conceitos básicos de decodificadores MAP, ML e NN. Como veremos, estes conceitos darão origem às idéias que aparecerão nas seções seguintes. Lembre que neste trabalho todos os canais considerados são sem memória.

Seja C um (M, N) código sobre um canal discreto $(\mathcal{X}, P(\mathcal{Y}|\mathcal{X}), \mathcal{Y})$ e $a : \mathcal{Y}^N \to C$ um decodificador para C. Temos que a probabilidade de erro de decodificação é dada pela média

$$P_e(C) = \sum_{\mathbf{c}\in C} P_e(\mathbf{c})\, P(\mathbf{c})$$

onde

$$P_e(\mathbf{c}) = 1 - \sum_{\mathbf{y}\in a^{-1}(\mathbf{c})} P(\mathbf{y}|\,\mathbf{c})$$

é a probabilidade do decodificador entregar ao destino uma palavra-código distinta da palavra transmitida $\mathbf{c}$. Podemos reescrever $P_e(C)$ considerando a probabilidade condicional $P(a(\mathbf{y})|\,\mathbf{y})$ de termos transmitido exatamente o que será entregue pelo decodificador. De fato,

$$\begin{aligned} P_e(C) &= \sum_{\mathbf{c}\in C} P_e(\mathbf{c})\, P(\mathbf{c}) \\ &= \sum_{\mathbf{c}\in C} \left(1 - \sum_{\mathbf{y}\in a^{-1}(\mathbf{c})} P(\mathbf{y}|\,\mathbf{c})\right) P(\mathbf{c}) \\ &= \sum_{\mathbf{c}\in C} P(\mathbf{c}) - \sum_{\mathbf{c}\in C} \sum_{\mathbf{y}\in a^{-1}(\mathbf{c})} P(\mathbf{y}|\,\mathbf{c})\, P(\mathbf{c}) \\ &= 1 - \sum_{\mathbf{y}\in\mathcal{Y}^N} P(\mathbf{y}|\,a(\mathbf{y}))\, P(a(\mathbf{y}))\,. \end{aligned}$$

Como $P(\mathbf{y}|\,\mathbf{c})\, P(\mathbf{c}) = P(\mathbf{c}|\,\mathbf{y})\, P(\mathbf{y})$, concluímos que

$$P_e(C) = 1 - \sum_{\mathbf{y}\in\mathcal{Y}^N} P(a(\mathbf{y})|\,\mathbf{y})\, P(\mathbf{y})\,.$$

Esta última expressão assegura que os decodificadores que minimizam a probabilidade de erro $P_e(C)$ são aqueles que maximizam as probabilidades condicionais $P(a(\mathbf{y})|\,\mathbf{y})$.

Definição 2.1 *Um decodificador de canal* $a : \mathcal{Y}^N \to C$ *que satisfaz a condição*

$$P\left(a\left(\mathbf{y}\right)|\,\mathbf{y}\right) = \max\left\{P\left(\mathbf{c}|\,\mathbf{y}\right) : \mathbf{c} \in C\right\}$$

é dito um ***decodificador de máxima probabilidade a posteriori*** (*MAP - Maximum a Posteriori Probability*).

Os comentários acima mostram que:

Teorema 2.1 *Os decodificadores MAP são os decodificadores que minimizam a probabilidade de erro* $P_e\left(C\right)$.

Note agora, como

$$P\left(\mathbf{c}|\,\mathbf{y}\right) = \frac{P\left(\mathbf{y}|\,\mathbf{c}\right)P\left(\mathbf{c}\right)}{P\left(\mathbf{y}\right)} \tag{2.1}$$

onde $P\left(\mathbf{y}\right) = \sum_{\mathbf{c}\in C} P\left(\mathbf{y}|\,\mathbf{c}\right)P\left(\mathbf{c}\right)$, que todo decodificador MAP depende das probabilidades condicionais $P\left(y_j|\,x_i\right)$ e da distribuição de probabilidades a priori $P\left(\mathbf{c}\right)$ de C.

Definição 2.2 *Um decodificador de canal* $a : \mathcal{Y}^N \to C$ *que satisfaz a condição*

$$P\left(\mathbf{y}|\,a\left(\mathbf{y}\right)\right) = \max\left\{P\left(\mathbf{y}|\,\mathbf{c}\right) : \mathbf{c} \in C\right\}$$

é dito um ***decodificador de máxima verossimilhança*** (*ML - Maximum Likelihood Probability*).

Os decodificadores ML só dependem do canal. No caso em que a distribuição de probabilidades a priori do código é uniforme, segue de (2.1) que todo decodificador MAP é também um decodificador ML e vice-versa.

Considere agora um canal q-ário simétrico com probabilidade de erro p (como no Exemplo 1.5). Nestas condições, se $\mathbf{y} = (y_1, \ldots, y_N)$ e $\mathbf{c} = (c_1, \ldots, c_N)$ são elementos

de $\mathcal{X}^N$, então

$$\begin{aligned} P(\mathbf{y}|\,\mathbf{c}) &= P(y_1|\,c_1)\cdot\ldots\cdot P(y_N|\,c_N) \\ &= (1-p)^{|\{i:y_i=c_i\}|}\left(\frac{p}{q-1}\right)^{|\{i:y_i\neq c_i\}|} \\ &= (1-p)^{N-|\{i:y_i\neq c_i\}|}\left(\frac{p}{q-1}\right)^{|\{i:y_i\neq c_i\}|} \\ &= (1-p)^{N}\left(\frac{p}{(1-p)(q-1)}\right)^{|\{i:y_i\neq c_i\}|}. \end{aligned} \tag{2.2}$$

Sendo assim, para $\mathbf{c}' = (c'_1, \ldots, c'_N)$,

$$P(\mathbf{y}|\,\mathbf{c}) \geq P(\mathbf{y}|\,\mathbf{c}') \Leftrightarrow |\{i : y_i \neq c_i\}| \leq |\{i : y_i \neq c'_i\}|,$$

onde $|X|$ denota a cardinalidade do conjunto finito X. Isto motiva a seguinte definição.

Definição 2.3 *A função $d_H : \mathcal{X}^N \times \mathcal{X}^N \to \mathbb{R}_+$ definida por*

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

$\mathbf{x} = (x_1, \ldots, x_N)$ e $\mathbf{y} = (y_1, \ldots, y_N)$, é a chamada ***distância de Hamming****.*

Não é difícil verificar que a distância de Hamming é uma métrica (ver por exemplo [30]). Os comentários acima mostram que os decodificadores ML em um canal q-ário simétrico também podem ser caracterizados pela condição

$$d_H(\mathbf{y}, a(\mathbf{y})) = \min\{d_H(\mathbf{y}, \mathbf{c}) : \mathbf{c} \in C\}.$$

Neste caso diremos que a é um *decodificador por máxima proximidade* (NN - Nearest Neighbor Decoder). É claro que:

Teorema 2.2 *Em um canal q-ário simétrico todo decodificador ML ou MAP de um código com distribuição de probabilidades a priori uniforme é também um decodificador NN e vice-versa.*

2.2 Códigos Lineares

Até aqui nenhuma restrição foi imposta sobre os alfabetos de entrada e de saída do canal discreto. Consequentemente os únicos possíveis algoritmos de decodificação são os de força bruta: comparamos a palavra recebida $\mathbf{y}$ com todas as palavras-código e decodificamos esta como sendo a palavra-código mais provável. É claro, os algoritmos de força bruta são computacionalmente inviáveis (salvo para dimensões baixas). A saída usual adotada para otimizar este problema consiste em agregar algum tipo de estrutura algébrica aos alfabetos de entrada e saída (que pode ser de grupo, anel ou corpo). A primeira destas estruturas comumente adotado é a de corpo finito.

Seja $\mathbb{F}_q$ um corpo finito contendo q elementos e $(\mathcal{X}, P(\mathcal{Y}|\mathcal{X}), \mathcal{Y})$ um canal discreto tal que $\mathcal{X} = \mathcal{Y} = \mathbb{F}_q$, que chamaremos simplesmente de *canal discreto sobre* $\mathbb{F}_q$. Neste caso um (M, N) código é um subconjunto C de $\mathbb{F}_q^N$ contendo M elementos e todo decodificador de C será uma aplicação sobrejetora do tipo $a : \mathbb{F}_q^N \to C$. A distância de Hamming $d_H(\mathbf{x}, \mathbf{y})$ pode ser escrita em função do *peso de Hamming* $w_H(\mathbf{x}) = |\{i : x_i \neq 0\}|$:

$$d_H(\mathbf{x}, \mathbf{y}) = w_H(\mathbf{x} - \mathbf{y}).$$

Estamos interessados particularmente na classe dos códigos lineares.

Definição 2.4 *Um* $[N; k]_q$ ***código linear*** *é um* $\left(q^k, N\right)$ *código de* $\mathbb{F}_q^N$ *que é um subespaço linear e de dimensão* k*.*

A *distância mínima de Hamming* $d_H(C)$,

$$d_H(C) := \min\{d_H(\mathbf{c}_1, \mathbf{c}_2) : \mathbf{c}_1, \mathbf{c}_2 \in C, \mathbf{c}_1 \neq \mathbf{c}_2\},$$

e o *raio de empacotamento* $R_H(C)$,

$$R_H(C) := \max\{r : B_H(\mathbf{c}_1; r) \cap B_H(\mathbf{c}_2; r) = \varnothing \text{ para todo } \mathbf{c}_1, \mathbf{c}_2 \in C,\ \mathbf{c}_1 \neq \mathbf{c}_2\},$$

onde $B_H(\mathbf{c}; r) := \left\{\mathbf{y} \in \mathbb{F}_q^N : d_H(\mathbf{y}, \mathbf{c}) \leq r\right\}$ é a *bola* de centro $\mathbf{c}$ e raio r, são os parâmetros básicos de um $[N; k]_q$ código linear C. Para os *espaços métricos de*

Hamming $\left(\mathbb{F}_q^N, d_H\right)$ o raio de empacotamento é dado em função da distância mínima (veja por exemplo [30]):

$$R_H(C) = \left[\frac{d_H(C) - 1}{2}\right]$$

onde $[x]$ é por definição a *parte inteira de* x. Outro parâmetro básico de um $[N; k]_q$ código linear C é o seu *espectro de pesos* $A_0(C), \ldots, A_N(C)$ com

$$A_i(C) := |\{\mathbf{c} \in C : w_H(\mathbf{c}) = i\}|.$$

Temos dois bons motivos para considerar a classe dos códigos lineares. Primeiro, se C é um $[N; k]_q$ código linear, então podemos descrever todos os elementos de C a partir de qualquer base de C: se $\{\mathbf{g}_1, \ldots, \mathbf{g}_k\}$ é uma base de C e $\mathbf{c} \in C$, então existem únicos escalares $u_1, \ldots, u_k \in \mathbb{F}_q$ tal que $\mathbf{c} = u_1\mathbf{g}_1 + \ldots + u_k\mathbf{g}_k$. Em outras palavras, $C = \left\{\mathbf{u}G : \mathbf{u} \in \mathbb{F}_q^k\right\}$ onde $G = (g_{ij})$ é a matriz dada pelos vetores $\mathbf{g}_i = (g_{i1}, \ldots, g_{in})$, $1 \leq i \leq k$, a chamada *matriz geradora* de C. O segundo motivo está relacionado com a existência de um esquema simples de decodificação (simples de descrever matematicamente!). Para tanto basta considerar qualquer matriz H de verificação de paridade de C. Uma matriz H de posto $N - k$ e ordem $(N - k) \times N$ tal que $H\mathbf{c} = \mathbf{0}$ para todo $\mathbf{c} \in C$ é dita uma *matriz de verificação de paridade* de C. Neste texto $\mathbf{0} = (0, \ldots, 0)$ sempre denotará o vetor nulo. Uma matriz de verificação de paridade H não só oferece um mecanismo para detectar erros ($\mathbf{c} \in C$ se, e somente se, $H\mathbf{c} = \mathbf{0}$), como também pode ser usada para definir um decodificador ML para o canal q-ário simétrico: se recebemos $\mathbf{y}$, definimos $a(\mathbf{y})$ como sendo a palavra-código $\mathbf{c}$ tal que $\mathbf{c} = \mathbf{y} - \mathbf{z}_0$ onde $\mathbf{z}_0$ é o vetor de menor peso tal que $H\mathbf{z}_0 = H\mathbf{y}$. Para determinarmos a solução ótima $\mathbf{z}_0$ precisamos aparentemente calcular as q^k possíveis soluções $\mathbf{z}$ de $H\mathbf{z} = H\mathbf{y}$. A procura por algoritmos de decodificação por máxima verossimilhança que realizem esta tarefa com um número de etapas inferior a q^k é um dos principais problemas em Teoria dos Códigos (ver por exemplo [13] e [36]). É possível que exista um algoritmo de decodificação que determine a resposta em tempo polinomial? Infelizmente[1] o problema geral da decodificação por máxima verossimilhança para canais q-ários simétricos é um problema do tipo NP-completo (ver [4]).

[1] A não ser que P=NP.

Exemplo 2.1 *Considere o* $[7;4]_2$ *código binário de Hamming* $\mathcal{H}(3)$ *dado pela matriz de verificação de paridade*

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

Como todo par de colunas de H *é linearmente independente e existem três colunas linearmente dependentes, concluímos que a distância mínima de Hamming de* $\mathcal{H}(3)$ *é igual a* 3*. Consequentemente, o seu raio de empacotamento é igual a* 1*. Como* $|B_H(\mathbf{c};r)|$ *não depende do centro* $\mathbf{c}$ *e como* $|B_H(\mathbf{0};1)| = 8$*, concluímos que*

$$\bigcup_{\mathbf{c}\in C} B_H(\mathbf{c};1) = \mathbb{F}_2^7,$$

ou seja, se $\mathbf{y} \in \mathbb{F}_2^7$ *e* $\mathbf{y} \notin \mathcal{H}(3)$*, então* $\mathbf{y} = \mathbf{c} + \mathbf{e}_i$ *para algum* $\mathbf{c} \in C$ *e para algum vetor canônico* $\mathbf{e}_i$*,* $1 \leq i \leq 7$*. Denotando as colunas de* H *por* $\mathbf{h}_1, \ldots, \mathbf{h}_7$*, teremos que* $H\mathbf{y} = \mathbf{0}$ *ou* $H\mathbf{y} = \mathbf{h}_i$ *para algum* $1 \leq i \leq 7$*. Com isto fica fácil descrever o decodificador NN:* $a_H(\mathbf{y}) = \mathbf{y}$ *se* $H\mathbf{y} = \mathbf{0}$*;* $a_H(\mathbf{y}) = \mathbf{y} - \mathbf{e}_i$ *se* $H\mathbf{y} = i$*-ésimo coluna de* H*.*

Fechamos o exemplo usando $\mathcal{H}(3)$ *para simular a transmissão da imagem "Hello World" por um canal binário simétrico, considerando o seguinte codificador de canal*:

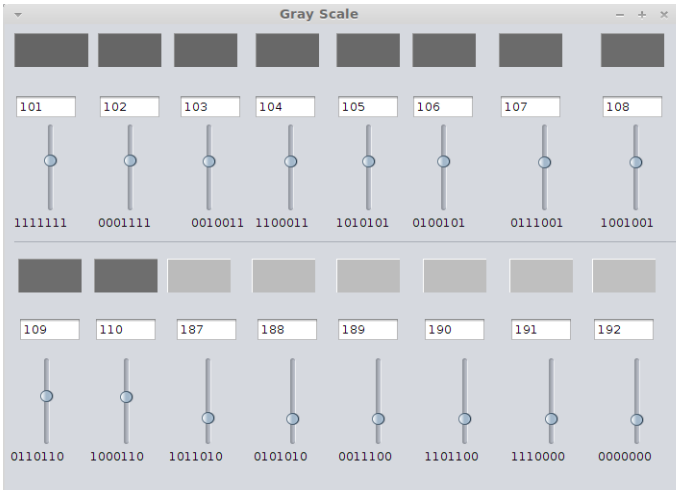

Figura 2.1: Palavras-código de $\mathcal{H}(3)$ e tons de cinza (ou valores de RGB) associados.

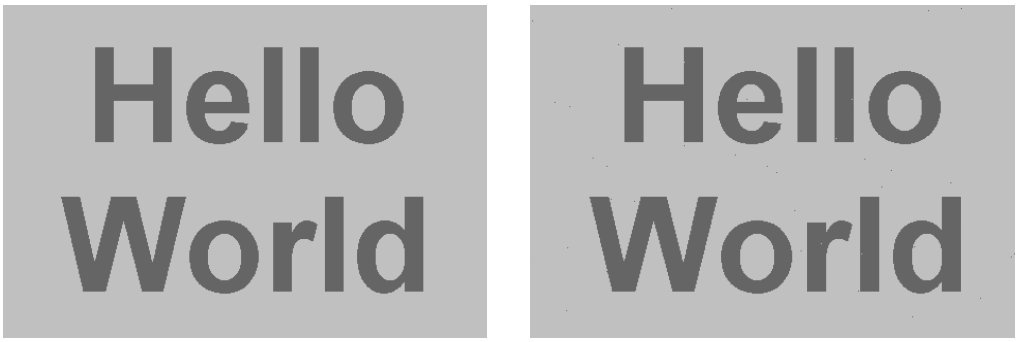

Figura 2.2: Imagem original e simulação com $p = 0.005$ respectivamente.

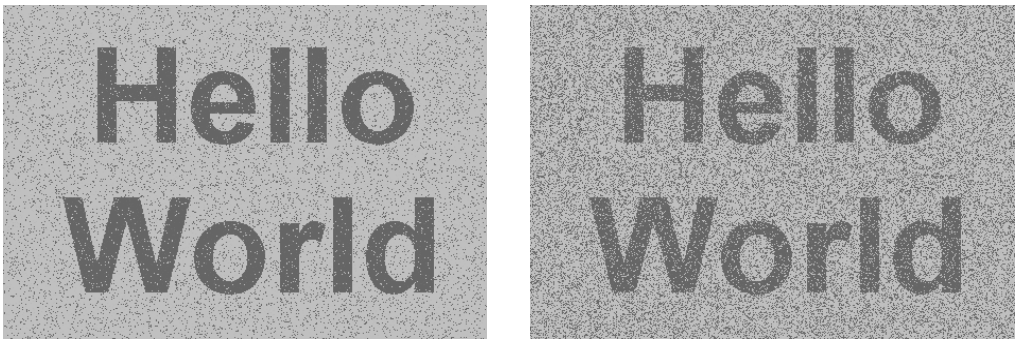

Figura 2.3: Simulação com $p = 0.1$ e $p = 0.2$ respectivamente.

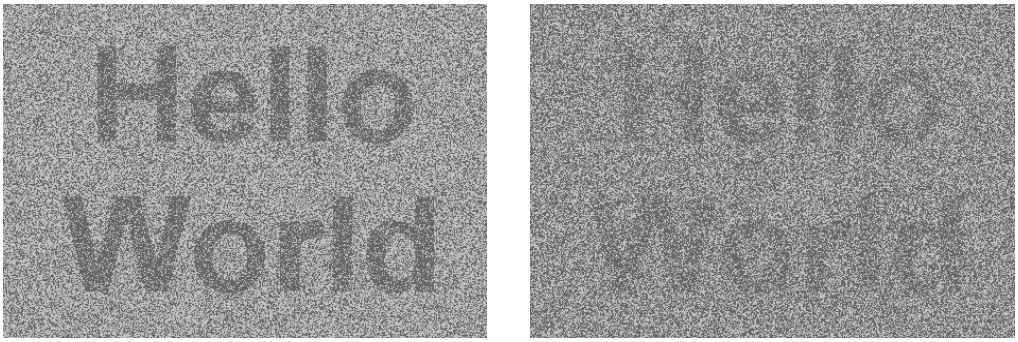

Figura 2.4: Simulação com $p = 0.3$ e $p = 0.4$ respectivamente.

Neste trabalho estamos interessados nos esquemas de decodificação que minimizam (ou pelo menos otimizam) a perda esperada relativa aos erros de decodificação com valor. A estrutura linear entra num primeiro momento para simplificar a for-

mulação da perda esperada.

No atual contexto não há distinção entre os diferentes erros de decodificação: supondo que a distribuição de probabilidades a priori do $[N;k]_q$ código linear C é uniforme e sendo $a : \mathbb{F}_q^N \to C$ um decodificador, considerando a aplicação $\nu_{(0,1)} : C \times C \to \mathbb{R}_+$ dada por

$$\nu_{(0,1)}(\mathbf{c},\mathbf{c}') = \begin{cases} 0 & \text{se} \quad \mathbf{c}=\mathbf{c}' \\ 1 & \text{caso} \quad \text{contrário} \end{cases},$$

teremos que

$$P_e(C) = \frac{1}{M}\sum_{\mathbf{c}\in C}\sum_{\mathbf{y}\in\mathbb{F}_q^N} \nu_{(0,1)}(a(\mathbf{y}),\mathbf{c})\, P(\mathbf{y}|\,\mathbf{c}). \tag{2.3}$$

A função $\nu_{(0,1)}$ é a *função indicadora* que somente detecta os erros de decodificação, não fazendo distinção entre os diferentes tipos de erros.

Uma outra característica de $P_e(C)$ que merece destaque: a expressão em (2.3) não leva em conta como o espaço de informações é mergulhado no código, ou seja, o codificador de canal é irrelevante para $P_e(C)$. Como estamos assumindo que a nossa fonte de informações é o espaço $\mathbb{F}_q^k$, o codificador de canal é uma aplicação bijetora do tipo $f : \mathbb{F}_q^k \to C$. Assumindo que $f(\mathbf{0}) = \mathbf{0}$ (poderíamos exigir mais, supondo que f é linear; toda matriz geradora G pode ser interpretada com sendo um codificador de canal que satisfaz esta condição) e considerando a função auxiliar $\widetilde{\nu}_{(0,1)} : \mathbb{F}_q^k \times \mathbb{F}_q^k \to \mathbb{R}_+$ dada por

$$\widetilde{\nu}_{(0,1)}(\mathbf{u},\mathbf{u}') = \begin{cases} 0 & \text{se} \quad \mathbf{u}=\mathbf{u}' \\ 1 & \text{caso} \quad \text{contrário} \end{cases}, \tag{2.4}$$

podemos reescrever $P_e(C)$ pondo

$$P_e(C) = \frac{1}{M}\sum_{\mathbf{c}\in C}\sum_{\mathbf{y}\in\mathbb{F}_q^N} \widetilde{\nu}_{(0,1)}\left(f^{-1}(a(\mathbf{y})), f^{-1}(\mathbf{c})\right) P(\mathbf{y}|\,\mathbf{c}), \tag{2.5}$$

evidenciando finalmente o codificador de canal. No entanto, como

$$\widetilde{\nu}_{(0,1)}\left(f^{-1}(\mathbf{c}), f^{-1}(\mathbf{c}')\right) = \nu_{(0,1)}(\mathbf{c},\mathbf{c}') \tag{2.6}$$

para todo $\mathbf{c} \in C$, concluímos que $P_e(C)$ não depende do codificador de canal f.

Exemplo 2.2 *Seja* $C = \mathbb{F}_2^4$ *o* $[4;4]_2$ *código binário. Considere o decodificador* a_H*, dado por* $a_H(\mathbf{y}) = \mathbf{y}$ *para todo* $\mathbf{y} \in \mathbb{F}_2^4$*, e os codificadores de canal dados na Figura 2.5 abaixo.*

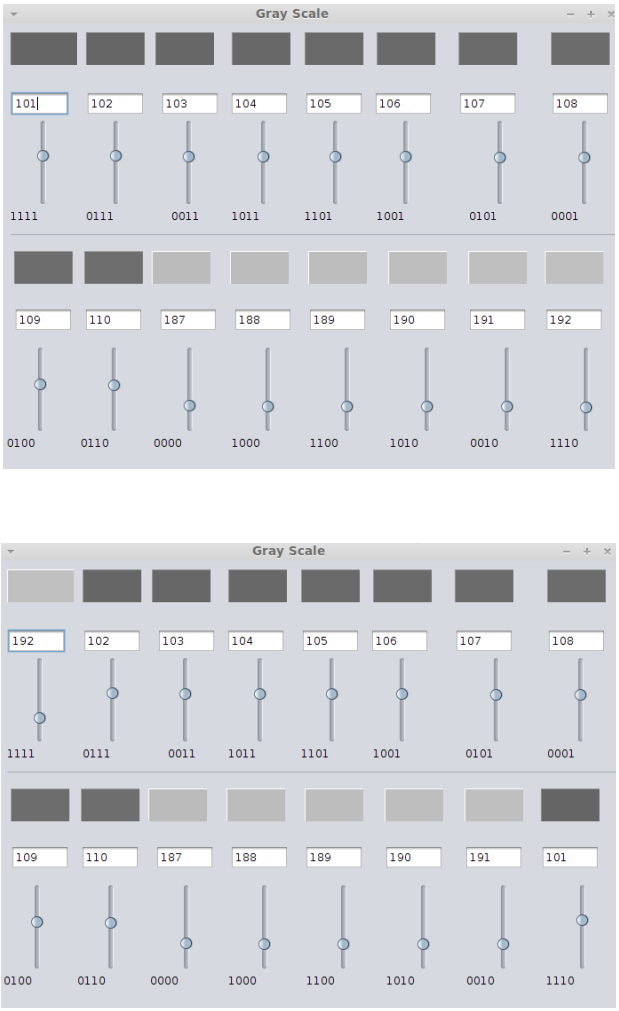

Figura 2.5: Codificadores de canal.

Simulando a transmissão da imagem "Hello World" pelo canal binário simétrico com probabilidade de erro $p = 0.1$*, obtemos diferentes percepções das imagens decodificadas em relação a cada um dos dois codificadores de canal, apesar de ambas as imagens terem aproximadamente o mesmo número de erros de decodificação*:

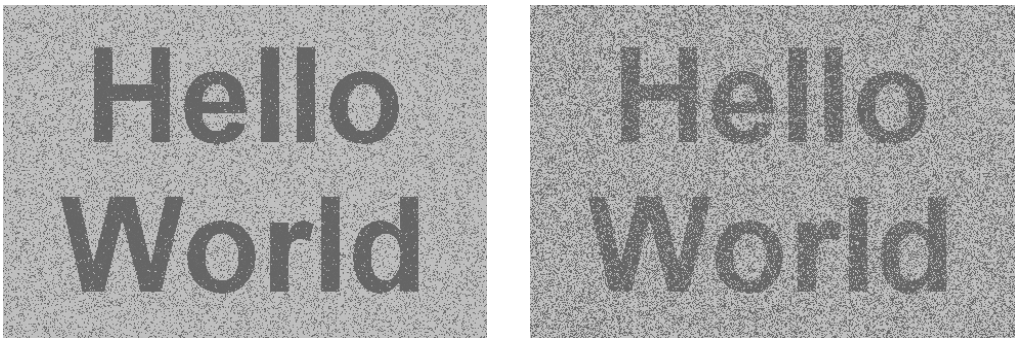

Figura 2.6: Simulação com $p = 0.1$.

Estas imagens ilustram a dependência do decodificador (que neste caso é único) com o codificador de canal. O número total de possíveis codificadores de canal aqui é igual a

$$16! = 2092\,2789888000.$$

2.3 Funções de Valor e Perdas Esperadas

Fica evidente agora que se considerarmos em (2.5) uma função $\nu : \mathbb{F}_q^k \times \mathbb{F}_q^k \to \mathbb{R}_+$ distinta da função auxiliar $\widetilde{\nu}_{(0,1)}$, então (2.5) passará a depender do codificador de canal $f : \mathbb{F}_q^k \to C$. É exatamente nesta nova formulação que estamos interessados. Esta é uma das principais contribuições deste trabalho, que passamos a descrever.

Primeiramente precisamos nos convencer de que em algumas situações é razoável assumirmos valores distintos para os diferentes tipos de erros de decodificação. Talvez o melhor exemplo para este tipo de situação seja o da transmissão de imagens digitais. Por exemplo, considere a imagem "Hello World", em escala de cinza, dada na Figura 1. Neste caso é razoável impormos que os valores dos erros de decodificação por tons mais próximos sejam menores do que os valores dos erros de decodificação por tons mais distantes.

Definição 2.5 *Uma* ***função de valor*** *sobre $\mathbb{F}_q^k \times \mathbb{F}_q^k$ é uma aplicação ν que associa a cada par $(\mathbf{u}, \mathbf{u}')$ de elementos de $\mathbb{F}_q^k \times \mathbb{F}_q^k$ um número real não-negativo, ou seja,*

uma aplicação

$$\nu : \mathbb{F}_q^k \times \mathbb{F}_q^k \to \mathbb{R}_+.$$

Se $f : \mathbb{F}_q^k \to C$ *é um codificador de canal, então a aplicação*

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

definida por $\nu_f(\mathbf{c}, \mathbf{c}') := \nu(f^{-1}(\mathbf{c}), f^{-1}(\mathbf{c}'))$ *é a chamada* ***função de valor associada a*** f.

Algumas condições naturais podem ser impostas sobre uma função de valor. Primeiramente, baseados na expressão em (2.5), onde $\widetilde{\nu}_{(0,1)}\left(f^{-1}\left(a\left(\mathbf{y}\right)\right), f^{-1}\left(\mathbf{c}\right)\right)$ indica se houve ou não perda (erro) no processo de decodificação, podemos impor que

$$\nu(\mathbf{u}, \mathbf{u}) = 0$$

para todo $\mathbf{u} \in \mathbb{F}_q^k$. Uma função de valor ν que satisfizer está condição será dita *razoável*. Uma segunda condição pode ser assumida baseada na seguinte observação: trocar preto por branco (onde o preto tem significado) ou branco por preto (onde o branco tem significado) são erros, segundo a nossa percepção, equivalentes. Neste sentido parece natural impormos que

$$\nu(\mathbf{u}, \mathbf{v}) = \nu(\mathbf{v}, \mathbf{u}).$$

Neste trabalho vamos considerar somente a classe das funções de valor $\nu : \mathbb{F}_q^k \times \mathbb{F}_q^k \to \mathbb{R}_+$ que são invariantes por translações, ou seja,

$$\nu\left(\mathbf{u} + \mathbf{w}, \mathbf{v} + \mathbf{w}\right) = \nu\left(\mathbf{u}, \mathbf{v}\right)$$

para todo $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{F}_q^k$. Com esta hipótese podemos identificar a função de valor ν com a função $\widehat{\nu} : \mathbb{F}_q^k \to \mathbb{R}_+$ dada por

$$\widehat{\nu}\left(\mathbf{u}\right) := \nu\left(\mathbf{u}, \mathbf{0}\right),$$

já que $\nu\left(\mathbf{u}, \mathbf{v}\right) = \nu\left(\mathbf{u} - \mathbf{v}, \mathbf{0}\right)$. Assumindo que todo codificador de canal $f : \mathbb{F}_q^k \to C$ é linear, podemos escrever

$$\nu_f = \nu \circ f^{-1}.$$

A hipótese de que a função de valor é invariante por translações não é totalmente compatível com a ideia de valor do erro de decodificação. Por exemplo, se considerarmos o espaço de informações $\mathbb{F}_2^2 = \{\mathbf{u}_0, \mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ e assumirmos $\nu(\mathbf{u}_i) = i$, com $\mathbf{u}_0 := 00$, teremos que

$$\mathbf{u}_1 - \mathbf{u}_2 = (\mathbf{u}_1 - \mathbf{u}_1) - (\mathbf{u}_2 - \mathbf{u}_1) = \mathbf{u}_0 - \mathbf{u}_3,$$

o que mostra que a diferença entre informações com valores próximos pode ser igual a diferença entre informações com valores mais afastados. Neste sentido perdemos alguma informação quando associamos a $\mathbf{u}_1 - \mathbf{u}_2$ o mesmo valor de $\mathbf{u}_0 - \mathbf{u}_3$.

Como daqui em diante todas as funções de valores consideradas serão invariantes por translações:

Definição 2.6 *Uma* ***função de valor invariante por translações*** *para $\mathbb{F}_q^k$ é uma aplicação ν que associa a cada elemento de $\mathbb{F}_q^k$ um número real não-negativo, ou seja, uma aplicação*

$$\nu : \mathbb{F}_q^k \to \mathbb{R}_+.$$

Se $f : \mathbb{F}_q^k \to C$ é um codificador de canal, então a aplicação

$$\nu_f : C \to \mathbb{R}_+$$

definida por $\nu_f := \nu \circ f^{-1}$ é a chamada ***função de valor associada a*** *f.*

Como neste texto trataremos exclusivamente de funções de valor invariantes por translações, nos referiremos a estas simplesmente como **funções de valor**.

Classicamente o desempenho de um par (C, a), onde C é um código e $a : \mathbb{F}_q^N \to C$ é um decodificador, é medido pela probabilidade de erro de decodificação $P_e(C)$, e como já sabemos, esta medida não depende do codificador de canal $f : \mathbb{F}_q^k \to C$. Isto significa que as informações são igualmente importantes, sendo todas as informações igualmente protegidas: o valor do erro de decodificação ($= 1$) não depende do erro.

A situação é bem mais delicada quando consideramos uma função de valor, já que neste caso temos que considerar também o codificador de canal. A expressão em (2.5) para $P_e(C)$ é a chave para definirmos a quantidade que medirá o desempenho do sistema (C, f, a, ν), onde C é um $[N; k]_q$ código linear, $f : \mathbb{F}_q^k \to C$ é um codificador de canal, $a : \mathbb{F}_q^N \to C$ é um decodificador e $\nu : \mathbb{F}_q^k \to \mathbb{R}_+$ é uma função de valor.

Definição 2.7 *Fixe um canal discreto sobre $\mathbb{F}_q$ e seja (C, f, a, ν) uma quádrupla tal que C é um $[N;k]_q$ código linear, $f : \mathbb{F}_q^k \to C$ é um codificador de canal, $a : \mathbb{F}_q^N \to C$ é um decodificador e $\nu : \mathbb{F}_q^k \to \mathbb{R}_+$ é uma função de valor. Seja $\{P(\mathbf{c}) : \mathbf{c} \in C\}$ a distribuição a priori de C. Definimos a* ***perda esperada total*** *$\mathbb{E}_C(f, a, \nu)$ de C relativa a tripla (f, a, ν) como sendo*

$$\mathbb{E}_C(f, a, \nu) = \sum_{\mathbf{y} \in \mathbb{F}_q^N} \mathbb{E}_{\mathbf{y}}(f, a, \nu) P(\mathbf{y}) \tag{2.7}$$

onde

$$\mathbb{E}_{\mathbf{y}}(f, a, \nu) = \sum_{\mathbf{c} \in C} \nu_f(a(\mathbf{y}) - \mathbf{c}) P(\mathbf{c}|\,\mathbf{y})$$

é a ***perda esperada*** *em $\mathbf{y}$ e ν_f é a função de valor associada a f.*

Podemos escrever (2.7) em função dos valores $\nu_f(\tau)$, $\tau \in C$. De fato, supondo que $C = \{\mathbf{c}_1, \ldots, \mathbf{c}_M\}$, temos que

$$\begin{aligned}
\mathbb{E}_C(f, a, \nu) &= \sum_{i=1}^{M} \sum_{\mathbf{y} \in a^{-1}(\mathbf{c}_i)} \mathbb{E}_{\mathbf{y}}(f, a, \nu) P(\mathbf{y}) \\
&= \sum_{i=1}^{M} \sum_{\mathbf{y} \in a^{-1}(\mathbf{c}_i)} \sum_{j=1}^{M} \nu_f(a(\mathbf{y}) - \mathbf{c}_j) P(\mathbf{c}_j|\,\mathbf{y}) P(\mathbf{y}) \\
&= \sum_{i=1}^{M} \sum_{\mathbf{y} \in a^{-1}(\mathbf{c}_i)} \sum_{j=1}^{M} \nu_f(\mathbf{c}_i - \mathbf{c}_j) P(\mathbf{c}_j|\,\mathbf{y}) P(\mathbf{y}).
\end{aligned}$$

Para cada $1 \leq i \leq M$ seja $\mathbf{c}_l^i \in C$ tal que $\mathbf{c}_i - \mathbf{c}_l^i = \mathbf{c}_l$. Nestas condições

$$\begin{aligned}
\mathbb{E}_C(f, a, \nu) &= \sum_{i=1}^{M} \sum_{l=1}^{M} \sum_{\mathbf{y} \in a^{-1}(\mathbf{c}_i)} \nu_f\left(\mathbf{c}_i - \mathbf{c}_l^i\right) P\left(\mathbf{c}_l^i\middle|\,\mathbf{y}\right) P(\mathbf{y}) \\
&= \sum_{i=1}^{M} \sum_{l=1}^{M} \sum_{\mathbf{y} \in a^{-1}(\mathbf{c}_i)} \nu_f(\mathbf{c}_l) P(\mathbf{c}_i - \mathbf{c}_l|\,\mathbf{y}) P(\mathbf{y}) \\
&= \sum_{l=1}^{M} \sum_{\mathbf{y} \in \mathbb{F}_q^N} \nu_f(\mathbf{c}_l) P(a(\mathbf{y}) - \mathbf{c}_l|\,\mathbf{y}) P(\mathbf{y}) \\
&= \sum_{l=1}^{M} \left(\sum_{\mathbf{y} \in \mathbb{F}_q^N} P(a(\mathbf{y}) - \mathbf{c}_l|\,\mathbf{y}) P(\mathbf{y}) \right) \nu_f(\mathbf{c}_l).
\end{aligned}$$

Consequentemente:

Proposição 2.1 *Temos que*

$$\mathbb{E}_C(f, a, \nu) = \sum_{\tau \in C} H_a(\tau)\, \nu_f(\tau) \tag{2.8}$$

com

$$H_a(\tau) = \sum_{\mathbf{y} \in \mathbb{F}_q^N} P(a(\mathbf{y}) - \tau | \mathbf{y})\, P(\mathbf{y}). \tag{2.9}$$

Como $P(\mathbf{y}|\mathbf{c})\, P(\mathbf{c}) = P(\mathbf{c}|\mathbf{y})\, P(\mathbf{y})$*, também podemos escrever*

$$\mathbb{E}_C(f, a, \nu) = \sum_{\tau \in C} G_a(\tau)\, \nu_f(\tau) \tag{2.10}$$

com

$$G_a(\tau) = \sum_{\mathbf{y} \in \mathbb{F}_q^N} P(\mathbf{y}| a(\mathbf{y}) - \tau)\, P(a(\mathbf{y}) - \tau). \tag{2.11}$$

A probabilidade de erro de decodificação $P_e(C)$ é minimizada com os decodificadores MAP, independentemente do codificador de canal. Dado um código C e uma função de valor ν, nosso problema agora consiste em buscar por pares (f^*, a^*) de codificadores e decodificadores que minimizam a perda esperada total.

Definição 2.8 *Seja C um código linear, ν uma função de valor, f^* um codificador e a^* um decodificador. Diremos que (f^*, a^*) é (f^*, a^*)-**Bayes** relativo a função de valor ν se*

$$\mathbb{E}_{\mathbf{y}}(f^*, a^*, \nu) = \min_{(f,a)} \mathbb{E}_{\mathbf{y}}(f, a, \nu)$$

para todo $\mathbf{y} \in \mathbb{F}_q^N$, onde o mínimo é tomado sobre o conjunto de todos os pares (f, a) de codificadores e decodificadores de C.

Determinar os pares (f^*, a^*)-Bayes é uma tarefa difícil. Podemos relaxar este problema fixando o codificador ou o decodificador.

Definição 2.9 *(i) Seja C um código linear, ν uma função de valor, a um decodificador e f^* um codificador. Diremos que f^* é um* ***codificador de Bayes*** *relativo ao decodificador a e a função de valor ν se*

$$\mathbb{E}_{\mathbf{y}}\left(f^*, a, \nu\right) = \min_{f} \mathbb{E}_{\mathbf{y}}\left(f, a, \nu\right)$$

para todo $\mathbf{y} \in \mathbb{F}_q^N$, onde o mínimo é tomado sobre o conjunto de todos os codificadores de C. (ii) Seja C um código linear, ν uma função de valor, a^* um decodificador e f um codificador. Diremos que a^* é um ***decodificador de Bayes*** *relativo ao codificador f e a função de valor ν se*

$$\mathbb{E}_{\mathbf{y}}\left(f, a^*, \nu\right) = \min_{a} \mathbb{E}_{\mathbf{y}}\left(f, a, \nu\right)$$

para todo $\mathbf{y} \in \mathbb{F}_q^N$, onde o mínimo é tomado sobre o conjunto de todos os decodificadores de C.

Para um $[N;k]_q$ código linear temos um total de $q^k!$ possíveis codificadores de canal. Por exemplo, para o código binário de Hamming $\mathcal{H}(3)$ existem $16! = 20922789888000$ codificadores de canal. O próximo teorema nos diz quais dos $2^k!$ codificadores de canal são de fato codificadores de Bayes.

Teorema 2.3 *Seja $C = \{\tau_1, \ldots, \tau_M\}$ um $[N;k]_q$ código linear, $a : \mathbb{F}_q^N \to C$ um decodificador e $\nu : \mathbb{F}_q^k \to \mathbb{R}_+$ uma função de valor. Assuma que*

$$G_a\left(\tau_1\right) \geq G_a\left(\tau_2\right) \geq \ldots \geq G_a\left(\tau_M\right)$$

com $G_a(\tau)$ dado em (2.11). Temos que $f : \mathbb{F}_q^k \to C$ é um codificador de Bayes relativo ao decodificador a e a função de valor ν se, e somente se,

$$\nu_f\left(\tau_1\right) \leq \nu_f\left(\tau_2\right) \leq \ldots \leq \nu_f\left(\tau_M\right).$$

O Teorema 2.3 é uma consequência do seguinte fato elementar: se $x \geq y \geq 0$ e $0 \leq a \leq b$, então

$$ax + by \leq bx + ay.$$

Temos ainda um quarto problema de otimização: fixado o espaço de informações $\mathbb{F}_q^k$, a dimensão N e a função de valor ν para $\mathbb{F}_q^k$, determinar as triplas (C, f, a), de $[N; k]_q$ códigos lineares, codificadores e decodificadores que minimizam a perda esperada total. Nesta caso é razoável considerar a notação $\mathbb{E}(C, f, a)$ para a perda esperada total.

2.4 O Caso Clássico: Função de Valor 0-1

Vamos fazer uma pausa e rever as idéias estabelecidas na Seção 2.3 considerando apenas a função de valor $\widetilde{\nu}_{(0,1)}$ definida em (2.4).

Seja $\nu := \widetilde{\nu}_{(0,1)}$ e assuma nesta seção que todo codificador de canal $f : \mathbb{F}_q^k \to C$ é *razoável*, ou seja, $f(\mathbf{0}) = \mathbf{0}$. Nestas condições temos que

$$\nu_f(\mathbf{c}) = \begin{cases} 0 & \text{se} \quad \mathbf{c} = \mathbf{0} \\ 1 & \text{se} \quad \mathbf{c} \neq \mathbf{0} \end{cases},$$

e daí que a perda esperada total relativa a tripla $\left(f, a, \widetilde{\nu}_{(0,1)}\right)$ coincide com a probabilidade de erro de decodificação de C:

$$\mathbb{E}_C(f, a, \nu) = P_e(C).$$

Consequentemente, como $P_e(C)$ não depende do codificador de canal, todo par (f, a^*), onde a^* é um decodificador MAP, é (f, a^*)-Bayes e todo codificador de canal f é um codificador de Bayes para C. Mais ainda, os decodificadores de Bayes são exatamente os decodificadores MAP de C. Temos ainda, no caso particular dos canais q-ários simétricos, que (que é exatamente o Teorema 2.2 no contexto de função de valor):

Teorema 2.4 *Em um canal q-ário simétrico todo decodificador de Bayes de um código com distribuição de probabilidades a priori uniforme é também um decodificador NN e vice-versa.*

Considere agora o quarto problema de otimização estabelecido no final da seção anterior: fixados N e k, determinar as triplas (C, f, a), de $[N; k]_q$ códigos lineares,

codificadores e decodificadores, que minimizam a perda esperada total $\mathbb{E}_C(f, a, \nu)$. Este problema parece ser ainda mais complexo que os demais, mesmo quando consideramos a função de valor $\widetilde{\nu}_{(0,1)}$ dada em (2.4). Vejamos um exemplo.

Exemplo 2.3 *Não podemos usar a distância mínima como parâmetro no problema acima. Sejam*

$$C_1 = \{0000000, 0001110, 1111110, 1110000\}$$

e

$$C_2 = \{0000000, 0011110, 1111000, 1100110\}$$

dois $[7;2]_2$ *códigos binários. Temos que* $d_H(C_1) = 3$ *e* $d_H(C_2) = 4$. *Assumindo que a transmissão é sobre um canal binário simétrico com probabilidade de erro* $p = 0.2$ *e considerando decodificadores ML, é possível mostrar*[2] *que* $P_e(C_1) \approx 0.1972$ *e* $P_e(C_2) \approx 0.2218$. *Conclusão: uma maior distância mínima não implica necessariamente em uma menor probabilidade de erro.*

Os códigos com máxima distância mínima de Hamming são relevantes quando consideramos os t-decodificadores. Neste caso o problema da minimização da perda esperada total $\mathbb{E}_C\left(f, a, \nu_{(0,1)}\right)$ deve ser substituído pelo problema da minimização da probabilidade de erro de decodificação $\widetilde{P}_e(C)$ referente aos t-decodificadores.

Seja C um $[N;k]_q$ código linear e $t := R_H(C)$ o seu raio de empacotamento. Um *t-decodificador* para C é um decodificador a que satisfaz a condição

$$a(\mathbf{y}) = \begin{cases} \mathbf{c} & \text{se } \mathbf{y} \in B_H(\mathbf{c}; t) \text{ para algum } \mathbf{c} \in C \\ ? & \text{caso contrário} \end{cases}.$$

Em outras palavras, um t-decodificador é um decodificador que é NN sobre a união $\cup_{\mathbf{c}\in C} B_H(\mathbf{c}; t)$ e constante, igual a ?, fora desta união. Quando $a(\mathbf{y}) = ?$ o decodificador simplesmente declara que houve um *erro de decodificação*. Sendo assim, a probabilidade $\widetilde{P}_e(\mathbf{c})$ do t-decodificador entregar para o destino uma palavra-código distinta da palavra transmitida $\mathbf{c}$ ou declarar um erro de decodificação é igual a

$$\widetilde{P}_e(\mathbf{c}) = 1 - \sum_{\mathbf{y}\in B_H(\mathbf{c};t)} P(\mathbf{y}|\,\mathbf{c}).$$

[2]Este exemplo foi cedido cordialmente por Leandro Cruvinel, atualmente aluno do Programa de Doutorado em Matemática da UNICAMP.

Assumindo que o canal é q-ário simétrico e que a distribuição de probabilidades a priori de C é uniforme, concluímos que a probabilidade de erro de decodificação é igual a

$$\widetilde{P}_e(C) = 1 - \sum_{\mathbf{y} \in B_H(\mathbf{c};t)} P(\mathbf{y}|\,\mathbf{c}),$$

já que neste caso $\widetilde{P}_e(\mathbf{c})$ não depende de $\mathbf{c}$. Explicitamente,

$$\widetilde{P}_e(C) = 1 - \sum_{i=0}^{t} \binom{N}{i} (q-1)^i (1-p)^N \left(\frac{p}{(1-p)(q-1)}\right)^i.$$

Fica fácil ver agora que:

Teorema 2.5 *Sejam C_1 e C_2 dois $[N;k]_q$ códigos lineares com distribuições de probabilidades a priori uniformes. Para o canal q-ário simétrico*

$$\widetilde{P}_e(C_1) \leq \widetilde{P}_e(C_2) \Leftrightarrow d_H(C_1) \geq d_H(C_2).$$

Sendo assim, o problema da minimização da probabilidade de erro de decodificação $\widetilde{P}_e(C)$ é equivalente ao problema da determinção dos $[N;k]_q$ códigos lineares com máxima distância mínima de Hamming.

Encerramos esta seção comentando brevemente o Teorema de Wyner. Em 1965 Wyner (ver [43]) mostrou que é sempre possível realizar uma *transmissão confiável* com t-decodificadores, ou seja, com $\widetilde{P}_e(C)$ tão pequeno quanto se queira, sempre que a taxa de informação é inferior a uma certa constante $\widetilde{\mathcal{C}}$, que Wyner chamou de *capacidade relativa aos t-decodificadores*. Apesar de Wyner não estabelecer uma expressão exata para $\widetilde{\mathcal{C}}$, alguns limitantes inferiores e superiores foram determinados. Por exemplo, para o canal binário simétrico Wyner mostrou que (ver Figura 2.7)

$$1 - H(2p) \leq \widetilde{\mathcal{C}} \leq 1 - H\left(\frac{1}{2} - \frac{1}{2}\sqrt{1-4p}\right),$$

onde p é a probabilidade de erro do canal e $H(x) := -x\log_2 x - (1-x)\log_2(1-x)$.

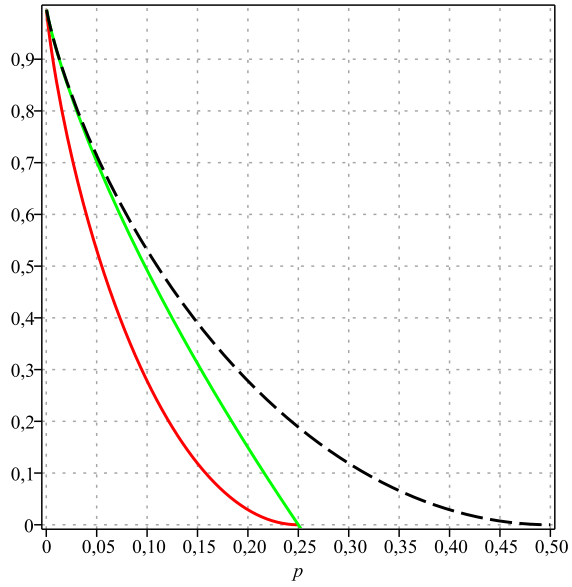

Figura 2.7: Limitante inferior (vermelho); limitante superior (verde); capacidade de Shannon (preto).

2.5 O Teorema de Shannon (Fraco) para $\mathbb{E}_C(f, a, \nu)$

Seja C um código (linear ou não) com distribuição de probabilidades a priori uniforme e $P_{eM}(C)$ a probabilidade de erro de decodificação em relação a um decodificador ML. Como o t-decodificador não explora todo o potencial de correção do código, é fácil ver que

$$\widetilde{P}_e(C) \geq P_{eM}(C).$$

Segue agora do Teorema de Shannon (Teorema 1.2) que $\widetilde{\mathcal{C}} \leq \mathcal{C}$. Wyner mostrou na realidade que $\widetilde{\mathcal{C}} < \mathcal{C}$ (ver [43]). Mostraremos agora que a capacidade do canal de transmitir informações com valor é limitada inferiormente por $\mathcal{C}$. É claro, precisamos estabelecer precisamente o conceito de taxa confiável para informações com valor.

No caso clássico, uma taxa de informação R é dita confiável quando para todo $\delta > 0$, existe um código C com taxa de informação R tal que $P_e(C) < \delta$. É claro, neste caso estamos considerando as funções de valor dadas como em (2.4). Agora queremos caracterizar as taxas de informações R tal que para todo $\delta > 0$ exista

uma quádrupla (C, f, a, ν) de código, codificador de canal, decodificador e função de valor, com C sendo um $[N;k]_q$ código linear com $R = \frac{k}{N}$, tal que $\mathbb{E}_C(f, a, \nu) < \delta$.

Identificando o espaço das funções de valor com o primeiro octante

$$\mathbb{R}_+^{q^k} := \left\{ \left(x_1, \ldots, x_{q^k}\right) \in \mathbb{R}^{q^k} : x_i \geq 0 \right\}$$

de $\mathbb{R}^{q^k}$ fica fácil ver de (2.8) que

$$\lim_{\lambda \to 0^+} \mathbb{E}_C(f, a, \lambda \cdot \nu) = 0$$

para toda função de valor ν, independentemente da taxa R. Em outras palavras, dado um código C, uma função de valor ν e $\delta > 0$, existe $\lambda > 0$ tal que $\mathbb{E}_C(f, a, \lambda \cdot \nu) < \delta$. Nos argumentos acima estamos ignorando completamente o fato elementar de que

$$\mathbb{E}_C(f, a, \lambda \cdot \nu) = \lambda \cdot \mathbb{E}_C(f, a, \nu),$$

ou seja, $\mathbb{E}_C(f, a, \lambda \cdot \nu)$ difere de $\mathbb{E}_C(f, a, \nu)$ por uma constante multiplicativa. O problema fica interessante se fixarmos um representante para cada classe

$$[\nu] := \{\lambda \cdot \nu : \lambda \in \mathbb{R}_+\}, \tag{2.12}$$

todos com uma mesma "medida".

Seja C um $[N;k]_q$ código linear. Se duas funções de valor $\nu_1, \nu_2 : \mathbb{F}_q^k \to \mathbb{R}_+$ diferem por uma constante multiplicativa $\lambda > 0$, $\nu_1 = \lambda \cdot \nu_2$, então as perdas esperadas totais de C relativas às funções de valor ν_1 e ν_2 diferem pela mesma constante multiplicativa λ e consequentemente diremos que ν_1 e ν_2 são *equivalentes*. Temos que $[\nu]$, definida em (2.12), é a *classe de equivalência* da função de valor ν. O *representante canônico* da classe $[\nu]$ é definido como sendo a função de valor $\mu \in [\nu]$ tal que $\|\mu\|_\infty = 1$, onde $\|\cdot\|_\infty$ é a *norma do máximo*

$$\|\mu\|_\infty = \max\left\{\mu(\mathbf{u}) : \mathbf{u} \in \mathbb{F}_q^k\right\}.$$

Podemos identificar as classes de equivalência $[\nu]$ com os seus respectivos representantes canônicos e desta forma identificar o espaço de todas as classes de equivalência $[\nu]$ com as faces do cubo

$$[0,1]^{q^k} = [0,1] \times \ldots \times [0,1] \ \ (q^k \text{ vezes})$$

que não estão contidas nos planos coordenados $x_i = 0$, $1 \leq i \leq q^k$.

Vamos denotar o conjunto dos representantes canônicos μ que são razoáveis, ou seja, tal que $\mu(\mathbf{0}) = 0$, por $\mathcal{V}_0$ e as faces do cubo $[0,1]^{q^k}$, exceto as faces contidas nos planos coordenados, por $\mathcal{V}$ (veja Figura 2.8).

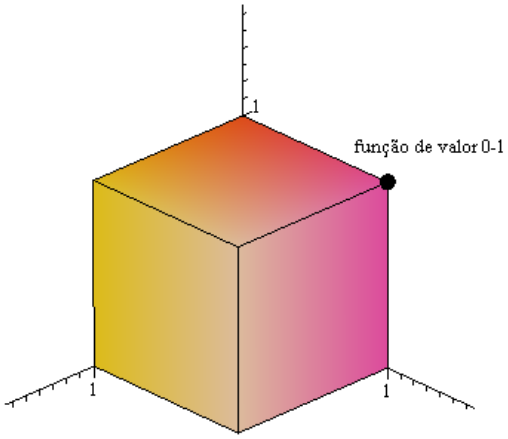

Figura 2.8: $\mathcal{V}$ e a função de valor 0-1.

Agora estamos prontos para estabelecer o conceito de taxa confiável para informações com valor.

Definição 2.10 *Uma taxa de informação R para espaços de informações com valor é dita* ***confiável*** *se para todo $\delta > 0$ existe um quádrupla (C, f, a, μ) de código, codificador de canal, decodificador e função de valor, onde C é um código com taxa de informação R, f é razoável e $\mu \in \mathcal{V}_0$, tal que $\mathbb{E}_C(f, a, \mu) < \delta$.*

Naturalmente estendemos a noção de capacidade:

Definição 2.11 *A máxima taxa de informação confiável para espaços de informações com valor para um canal discreto é a chamada capacidade do canal de transmistir informações com valor.*

Mostraremos agora que a capacidade de um canal binário simétrico de transmitir informações com valor é limitada inferiormente pela sua capacidade do canal. Usaremos o fato de que o Teorema 1.2 continua valendo para códigos binários lineares se o canal é binário simétrico (ver [25]).

Teorema 2.6 *Seja $\mathcal{C}$ a capacidade de um canal binário simétrico. Se $\widehat{\mathcal{C}}$ é a capacidade deste canal de transmitir informações com valor, então*

$$\widehat{\mathcal{C}} \geq \mathcal{C}.$$

Demonstração Seja $R \leq \mathcal{C}$ e $\delta > 0$. Como $R \leq \mathcal{C}$ existe um $[N;k]_2$ código linear C com $\frac{k}{N} = R$ tal que $P_e(C) < \delta$ para todo decodificador MAP $a : \mathbb{F}_2^N \to C$. Podemos reescrever (2.7) pondo

$$\mathbb{E}_C(f, a, \mu) = \sum_{\mathbf{c} \in C} \left(\sum_{\mathbf{y} \in \mathbb{F}_2^N} \mu_f(a(\mathbf{y}) - \mathbf{c}) P(\mathbf{y}|\,\mathbf{c}) \right) P(\mathbf{c}).$$

Supondo que μ e f são razoáveis, teremos que

$$\mathbb{E}_C(f, a, \mu) = \sum_{\mathbf{c} \in C} \left(\sum_{\mathbf{y} \notin a^{-1}(\mathbf{c})} \mu_f(a(\mathbf{y}) - \mathbf{c}) P(\mathbf{y}|\,\mathbf{c}) \right) P(\mathbf{c}).$$

Além disso, se $\mu(\mathbf{u}) \leq 1$ para todo $\mathbf{u} \in \mathbb{F}_q^k$ temos que

$$\mathbb{E}_C(f, a, \mu) \leq \sum_{\mathbf{c} \in C} \left(\sum_{\mathbf{y} \notin a^{-1}(\mathbf{c})} P(\mathbf{y}|\,\mathbf{c}) \right) P(\mathbf{c}),$$

ou seja,

$$\mathbb{E}_C(f, a, \mu) \leq P_e(C).$$

Disto segue que $\mathbb{E}_C(f, a, \mu) < \delta$, donde concluímos que R é também uma taxa de informação confiável para espaços de informações com valor. Consequentemente, $\widehat{\mathcal{C}} \geq \mathcal{C}$. □

2.6 Decodificadores de Bayes que não são ML

Já sabemos que os decodificadores ML minimizam a probabilidade de erro de decodificação $P_e(C)$ quando a distribuição de probabilidades a priori é uniforme (Teorema 2.1). Na linguagem de função de valor, os decodificadores ML são exatamente os decodificadores de Bayes relativos a função de valor 0-1 (veja a seção 2.4). É de se esperar que, em geral, nem todos os decodificadores ML sejam decodificadores de Bayes para outras funções de valor. Os resultados desta seção confirmam essa impressão.

Como o codificador de canal $f : \mathbb{F}_q^k \to C$ não será relevante nesta seção, denotaremos a perda esperada total simplesmente por $\mathbb{E}_C(a, \nu)$:

$$\mathbb{E}_C(a, \nu) := \mathbb{E}_C(f, a, \nu)$$

com f fixo. Também não faremos distinção entre a noção de função de valor e função de valor associada, escrevendo simplesmente ν ao invés de $\nu \circ f^{-1}$.

Teorema 2.7 *Seja C um $[N; k]_q$ código linear com distribuição de probabilidades a priori uniforme e $a : \mathbb{F}_q^N \to C$ um decodificador ML. Se $P(\mathbf{x}|\,\mathbf{x}) > P(\mathbf{y}|\,\mathbf{x})$ para todo $\mathbf{y} \neq \mathbf{x}$, então existe um decodificador $b : \mathbb{F}_q^N \to C$ e funções de valor $\nu_1, \nu_2 : C \to \mathbb{R}_+$ tais que*

$$\mathbb{E}_C(b, \nu_1) < \mathbb{E}_C(a, \nu_1) \tag{2.13}$$

e

$$\mathbb{E}_C(b, \nu_2) > \mathbb{E}_C(a, \nu_2). \tag{2.14}$$

Demonstração Seja C um $[N; k]_q$ código linear, $a : \mathbb{F}_q^N \to C$ um decodificador ML para C e $\nu_1 : C \to \mathbb{R}_+$ a função de valor definida como

$$\nu_1(\tau) = \begin{cases} 1 & \text{se} \quad \tau = \mathbf{0} \\ 0 & \text{se} \quad \tau \neq \mathbf{0} \end{cases}.$$

Agora sejam $\mathbf{c}_1, \mathbf{c}_2 \in C$, com $\mathbf{c}_1 \neq \mathbf{c}_2$, e defina $b := b_{a,\mathbf{c}_1,\mathbf{c}_2} : \mathbb{F}_q^N \to C$ pondo

$$b(\mathbf{y}) = \begin{cases} \mathbf{c}_1 & \text{se} \quad \mathbf{y} = \mathbf{c}_2 \\ \mathbf{c}_2 & \text{se} \quad \mathbf{y} = \mathbf{c}_1 \\ a(\mathbf{y}) & \text{caso contrário} \end{cases}.$$

Temos claramente que b não é um decodificador ML para C.

Considerando um canal discreto sobre $\mathbb{F}_q$ e considerando a função de valor ν_1 temos que

$$\begin{aligned}\mathbb{E}_C(a,\nu_1)-\mathbb{E}_C(b,\nu_1) &= \sum_{\tau\in C}\left(\sum_{\mathbf{y}\in\mathbb{F}_q^N} P(\mathbf{y}|\, a(\mathbf{y})-\tau)\, P(a(\mathbf{y})-\tau)\right)\nu_1(\tau)\\ &\quad -\sum_{\tau\in C}\left(\sum_{\mathbf{y}\in\mathbb{F}_q^N} P(\mathbf{y}|\, b(\mathbf{y})-\tau)\, P(b(\mathbf{y})-\tau)\right)\nu_1(\tau)\\ &= \sum_{\mathbf{y}\in\mathbb{F}_q^N} P(\mathbf{y}|\, a(\mathbf{y}))\, P(a(\mathbf{y})) - \sum_{\mathbf{y}\in\mathbb{F}_q^N} P(\mathbf{y}|\, b(\mathbf{y}))\, P(b(\mathbf{y}))\,.\end{aligned}$$

Como $a(\mathbf{y})=b(\mathbf{y})$ se $\mathbf{y}\neq\mathbf{c}_1,\mathbf{c}_2$ temos agora que

$$\begin{aligned}\mathbb{E}_C(a,\nu_1)-\mathbb{E}_C(b,\nu_1) &= P(\mathbf{c}_1|\, a(\mathbf{c}_1))\, P(a(\mathbf{c}_1)) + P(\mathbf{c}_2|\, a(\mathbf{c}_2))\, P(a(\mathbf{c}_2))\\ &\quad -P(\mathbf{c}_1|\, b(\mathbf{c}_1))\, P(b(\mathbf{c}_1)) - P(\mathbf{c}_2|\, b(\mathbf{c}_2))\, P(b(\mathbf{c}_2))\\ &= P(\mathbf{c}_1|\,\mathbf{c}_1)\, P(\mathbf{c}_1) + P(\mathbf{c}_2|\,\mathbf{c}_2)\, P(\mathbf{c}_2)\\ &\quad -P(\mathbf{c}_1|\,\mathbf{c}_2)\, P(\mathbf{c}_2) - P(\mathbf{c}_2|\,\mathbf{c}_1)\, P(\mathbf{c}_1)\,.\end{aligned}$$

Como por hipótese a distribuição de probabilidades a priori de C é uniforme e assumindo que $P(\mathbf{c}|\,\mathbf{c}) > P(\mathbf{y}|\,\mathbf{c})$ para todo $\mathbf{y}\neq\mathbf{c}$, concluímos que

$$\mathbb{E}_C(a,\nu_1)-\mathbb{E}_C(b,\nu_1)>0$$

já que

$$P(\mathbf{c}_1|\,\mathbf{c}_1) > P(\mathbf{c}_1|\,\mathbf{c}_2) \text{ e } P(\mathbf{c}_2|\,\mathbf{c}_2) > P(\mathbf{c}_2|\,\mathbf{c}_1)\,.$$

Daí que

$$\mathbb{E}_C(a,\nu_1) > \mathbb{E}_C(b,\nu_1)\,.$$

Para finalizar basta considerar ν_2 como sendo a função de valor 0-1. De fato, como b não é um decodificador ML e como $\mathbb{E}_C(g,\nu_2)$ é a probabilidade de erro $P_e(C)$ de C em relação ao decodificador g, concluímos que

$$\mathbb{E}_C(b,\nu_2) > \mathbb{E}_C(a,\nu_2)\,,$$

já que a é um decodificador ML. □

O Teorema 2.7 assegura que existem funções de valor para as quais os decodificadores ML não são necessariamente decodificadores de Bayes. Impondo algumas condições adicionais é possível estender este teorema sobre a classe dos códigos binários. Mais especificamente:

Teorema 2.8 *Seja C um $[N;k]_2$ código binário e $a : \mathbb{F}_2^N \to C$ um decodificador tal que*

$$\sum_{\mathbf{y}\in a^{-1}(\mathbf{0})} \left(P\left(\mathbf{y}\middle|\,\mathbf{0}\right)P\left(\mathbf{0}\right) - P\left(\mathbf{y}\middle|\,\mathbf{c}\right)P\left(\mathbf{c}\right)\right) \neq \sum_{\mathbf{y}\in a^{-1}(\mathbf{c})} \left(P\left(\mathbf{y}\middle|\,\mathbf{0}\right)P\left(\mathbf{0}\right) - P\left(\mathbf{y}\middle|\,\mathbf{c}\right)P\left(\mathbf{c}\right)\right) \tag{2.15}$$

para algum $\mathbf{c} \in C$. Então existe um decodificador $b : \mathbb{F}_2^N \to C$ e funções de valor $\nu_1, \nu_2 : C \to \mathbb{R}_+$ tal que

$$\left(\mathbb{E}_C\left(a, \nu_1\right) - \mathbb{E}_C\left(b, \nu_1\right)\right) \cdot \left(\mathbb{E}_C\left(a, \nu_2\right) - \mathbb{E}_C\left(b, \nu_2\right)\right) < 0.$$

Demonstração Seja C um $[N;k]_2$ código binário e $a : \mathbb{F}_2^N \to C$ um decodificador para C. Dado $\mathbf{c}_1 \in C$ definimos a função de valor $\nu_1 : C \to \mathbb{R}_+$ pondo

$$\nu_{_1}\left(\tau\right) = \begin{cases} 1 & \text{se} \quad \tau = \mathbf{c}_1 \\ 0 & \text{se} \quad \tau \neq \mathbf{c}_1 \end{cases}.$$

Agora dado $\mathbf{c}_2 \in C$ definimos o decodificador $b := b_{a,\mathbf{c}_1,\mathbf{c}_2} : \mathbb{F}_2^N \to C$ pondo

$$b\left(\mathbf{y}\right) = \begin{cases} \mathbf{c}_1 & \text{se} \quad \mathbf{y} \in a^{-1}\left(\mathbf{c}_2\right) \\ \mathbf{c}_2 & \text{se} \quad \mathbf{y} \in a^{-1}\left(\mathbf{c}_1\right) \\ a\left(\mathbf{y}\right) & \text{caso contrário} \end{cases}.$$

Agora seja $\nu_2 : C \to \mathbb{R}_+$ a função de valor dada por $\nu_2\left(\mathbf{c}_2\right) = 1$ e $\nu_2\left(\tau\right) = 0$ para todo $\tau \neq \mathbf{c}_2$.

Considerando um canal discreto sobre $\mathbb{F}_2$ e considerando a função de valor ν_1

temos que

$$\begin{aligned}\mathbb{E}_C(a,\nu_1)-\mathbb{E}_C(b,\nu_1) &= \sum_{\tau\in C}\left(\sum_{\mathbf{y}\in\mathbb{F}_2^N} P(\mathbf{y}|\, a(\mathbf{y})-\tau)\, P(a(\mathbf{y})-\tau)\right)\nu_1(\tau)\\ &\quad -\sum_{\tau\in C}\left(\sum_{\mathbf{y}\in\mathbb{F}_2^N} P(\mathbf{y}|\, b(\mathbf{y})-\tau)\, P(b(\mathbf{y})-\tau)\right)\nu_1(\tau)\\ &= \sum_{\mathbf{y}\in\mathbb{F}_2^N} P(\mathbf{y}|\, a(\mathbf{y})-\mathbf{c}_1)\, P(a(\mathbf{y})-\mathbf{c}_1)\\ &\quad -\sum_{\mathbf{y}\in\mathbb{F}_2^N} P(\mathbf{y}|\, b(\mathbf{y})-\mathbf{c}_1)\, P(b(\mathbf{y})-\mathbf{c}_1).\end{aligned}$$

Como $a(\mathbf{y})=b(\mathbf{y})$ para todo $\mathbf{y}\notin a^{-1}(\mathbf{c}_1)\cup a^{-1}(\mathbf{c}_2)$ obtemos que

$$\begin{aligned}\mathbb{E}_C(a,\nu_1)-\mathbb{E}_C(b,\nu_1) &= \sum_{\mathbf{y}\in a^{-1}(\mathbf{c}_1)} P(\mathbf{y}|\, a(\mathbf{y})-\mathbf{c}_1)\, P(a(\mathbf{y})-\mathbf{c}_1)\\ &\quad -\sum_{\mathbf{y}\in a^{-1}(\mathbf{c}_1)} P(\mathbf{y}|\, b(\mathbf{y})-\mathbf{c}_1)\, P(b(\mathbf{y})-\mathbf{c}_1)\\ &\quad +\sum_{\mathbf{y}\in a^{-1}(\mathbf{c}_2)} P(\mathbf{y}|\, a(\mathbf{y})-\mathbf{c}_1)\, P(a(\mathbf{y})-\mathbf{c}_1)\\ &\quad -\sum_{\mathbf{y}\in a^{-1}(\mathbf{c}_2)} P(\mathbf{y}|\, b(\mathbf{y})-\mathbf{c}_1)\, P(b(\mathbf{y})-\mathbf{c}_1).\end{aligned}$$

A definição do decodificador b assegura que

$$\sum_{\mathbf{y}\in a^{-1}(\mathbf{c}_1)} P(\mathbf{y}|\, b(\mathbf{y})-\mathbf{c}_1)\, P(b(\mathbf{y})-\mathbf{c}_1)=\sum_{\mathbf{y}\in a^{-1}(\mathbf{c}_1)} P(\mathbf{y}|\,\mathbf{c}_2-\mathbf{c}_1)\, P(\mathbf{c}_2-\mathbf{c}_1)$$

e

$$\sum_{\mathbf{y}\in a^{-1}(\mathbf{c}_2)} P(\mathbf{y}|\, b(\mathbf{y})-\mathbf{c}_1)\, P(b(\mathbf{y})-\mathbf{c}_1)=\sum_{\mathbf{y}\in a^{-1}(\mathbf{c}_2)} P(\mathbf{y}|\,\mathbf{0})\, P(\mathbf{0}),$$

donde segue que

$$\begin{aligned}\mathbb{E}_C(a,\nu_1)-\mathbb{E}_C(b,\nu_1) &= \sum_{\mathbf{y}\in a^{-1}(\mathbf{c}_1)} P(\mathbf{y}|\,\mathbf{0})\,P(\mathbf{0}) \\ &\quad - \sum_{\mathbf{y}\in a^{-1}(\mathbf{c}_1)} P(\mathbf{y}|\,\mathbf{c}_2-\mathbf{c}_1)\,P(\mathbf{c}_2-\mathbf{c}_1) \\ &\quad + \sum_{\mathbf{y}\in a^{-1}(\mathbf{c}_2)} P(\mathbf{y}|\,\mathbf{c}_2-\mathbf{c}_1)\,P(\mathbf{c}_2-\mathbf{c}_1) \\ &\quad - \sum_{\mathbf{y}\in a^{-1}(\mathbf{c}_2)} P(\mathbf{y}|\,\mathbf{0})\,P(\mathbf{0})\,.\end{aligned}$$

De forma similar, considerando agora a função de valor ν_2, temos que

$$\begin{aligned}\mathbb{E}_C(a,\nu_2)-\mathbb{E}_C(b,\nu_2) &= \sum_{\mathbf{y}\in a^{-1}(\mathbf{c}_1)} P(\mathbf{y}|\,\mathbf{c}_1-\mathbf{c}_2)\,P(\mathbf{c}_1-\mathbf{c}_2) \\ &\quad - \sum_{\mathbf{y}\in a^{-1}(\mathbf{c}_1)} P(\mathbf{y}|\,\mathbf{0})\,P(\mathbf{0}) \\ &\quad + \sum_{\mathbf{y}\in a^{-1}(\mathbf{c}_2)} P(\mathbf{y}|\,\mathbf{0})\,P(\mathbf{0}) \\ &\quad - \sum_{\mathbf{y}\in a^{-1}(\mathbf{c}_2)} P(\mathbf{y}|\,\mathbf{c}_1-\mathbf{c}_2)\,P(\mathbf{c}_1-\mathbf{c}_2)\,.\end{aligned}$$

Como estamos supondo que C é um código binário, temos que $\mathbf{c}_1-\mathbf{c}_2=\mathbf{c}_2-\mathbf{c}_1$. Com isto podemos concluir que

$$(\mathbb{E}_C(a,\nu_1)-\mathbb{E}_C(b,\nu_1))\cdot(\mathbb{E}_C(a,\nu_2)-\mathbb{E}_C(b,\nu_2))\leq 0.$$

Finalmente, tomando $\mathbf{c}_1=\mathbf{0}$ e $\mathbf{c}_2=\mathbf{c}$, segue de (2.15) que

$$(\mathbb{E}_C(a,\nu_1)-\mathbb{E}_C(b,\nu_1))\cdot(\mathbb{E}_C(a,\nu_2)-\mathbb{E}_C(b,\nu_2))<0.$$

□

Observe agora que a condição (2.15) é de fato satisfeita se a distribuição de probabilidades a priori de C é uniforme e $a:\mathbb{F}_2^N\to C$ é um decodificador ML,

assumindo que o canal satisfaz a condição razoável de que $P(\mathbf{c}|\,\mathbf{c}) > P(\mathbf{y}|\,\mathbf{c})$ para todo $\mathbf{y} \neq \mathbf{c}$: se a é um decodificador ML, então

$$P(\mathbf{y}|\,\mathbf{0}) - P(\mathbf{y}|\,\mathbf{c}) \geq 0$$

para todo $\mathbf{y} \in a^{-1}(\mathbf{0})$ e

$$P(\mathbf{y}|\,\mathbf{0}) - P(\mathbf{y}|\,\mathbf{c}) \leq 0$$

para todo $\mathbf{y} \in a^{-1}(\mathbf{c})$. Como $P(\mathbf{0}|\,\mathbf{0}) > P(\mathbf{0}|\,\mathbf{c})$ e $P(\mathbf{c}|\,\mathbf{0}) < P(\mathbf{c}|\,\mathbf{c})$, concluímos que

$$\sum_{\mathbf{y}\in a^{-1}(\mathbf{0})} \left(P(\mathbf{y}|\,\mathbf{0})\,P(\mathbf{0}) - P(\mathbf{y}|\,\mathbf{c})\,P(\mathbf{c})\right) > \sum_{\mathbf{y}\in a^{-1}(\mathbf{c})} \left(P(\mathbf{y}|\,\mathbf{0})\,P(\mathbf{0}) - P(\mathbf{y}|\,\mathbf{c})\,P(\mathbf{c})\right).$$

2.7 A Perda Esperada como Funcional Linear

Na Seção 2.5 identificamos o espaço das funções de valor com o primeiro octante $\mathbb{R}_+^{q^k}$ de $\mathbb{R}^{q^k}$. Com esta identificação temos que as funções de valor empregadas nas demonstrações dos Teoremas 2.7 e 2.8 correspondem aos vértices do cubo $[0,1]^{q^k}$. Veremos agora que estes resultados continuam valendo para muitas outras funções de valor. Vamos restringir o espaço das funções de valor ao conjunto

$$\mathcal{V} = \{\nu : \nu \text{ é uma função de valor com } \|\nu\|_\infty = 1\}$$

das faces de $[0,1]^{q^k}$ que não estão contidas nos planos coordenados, exatamente como na Seção 2.5.

Assumindo que o codificador de canal $f : \mathbb{F}_q^k \to C$ está fixo e considerando a notação $\mathbb{E}_C(a,\nu)$ para a perda esperada total (como na Seção 2.6), podemos comparar pares de decodificadores $a, b : \mathbb{F}_q^N \to C$ relativos a uma dada função de valor ν considerando a *diferença*

$$\mathbb{E}_C(a,b,\nu) := \mathbb{E}_C(a,\nu) - \mathbb{E}_C(b,\nu)$$

entre as perdas esperadas totais relativas a estes decodificadores.

Definição 2.12 *Dados um $[N;k]_q$ código linear C, uma função de valor $\nu \in \mathcal{V}$ e decodificadores $a, b : \mathbb{F}_q^N \to C$, diremos que a é **melhor** do que b, relativo a função de valor ν, se*

$$\mathbb{E}_C(a, b, \nu) < 0,$$

ou seja, se $\mathbb{E}_C(a, \nu) < \mathbb{E}_C(b, \nu)$.

Seja $\mathcal{D}(C)$ o conjunto de todos os possíveis decodificadores $a : \mathbb{F}_q^N \to C$. A diferença entre perdas esperadas $\mathbb{E}_C(a, b, \nu)$ induz naturalmente uma relação de ordem sobre o conjunto das classes

$$[a]_\nu := \{b \in \mathcal{D}(C) : \mathbb{E}_C(a, b, \nu) = 0\}$$

formadas pelos decodificadores $b : \mathbb{F}_q^N \to C$ tal que $\mathbb{E}_C(b, \nu) = \mathbb{E}_C(a, \nu)$: diremos que $[a]_\nu$ é *menor ou igual* que $[b]_\nu$, e escrevemos

$$[a]_\nu \preceq [b]_\nu,$$

se $\mathbb{E}_C(a, b, \nu) \leq 0$. Note agora que $\preceq$ é uma relação de ordem total. Mais ainda, o elemento minimal desta ordem é exatamente a classe dos decodificadores de Bayes de C relativos a ν.

Nas considerações acima fixamos uma função de valor ν e consideramos a diferença das perdas esperadas $\mathbb{E}_C(a, b, \nu)$ para analisar a performance dos decodificadores a e b relativos a ν. Agora vamos considerar o problema dual: fixado dois decodificadores a e b de um código C, queremos determinar o conjunto das funções de valor ν para o qual a é melhor do que b e vice-versa. Faremos isto interpretando a diferença $\mathbb{E}_C(a, \nu)$ como sendo um funcional linear.

Temos de (2.10) que

$$\mathbb{E}_C(a, \nu) = \sum_{\tau \in C} G_a(\tau)\, \nu(\tau)$$

com

$$G_a(\tau) = \sum_{\mathbf{y} \in \mathbb{F}_q^N} P(\mathbf{y} |\, a(\mathbf{y}) - \tau)\, P(a(\mathbf{y}) - \tau).$$

Consequentemente,

$$\mathbb{E}_C(a,b,\nu) = \sum_{\tau \in C} T_{(a,b)}(\tau)\,\nu(\tau)$$

com

$$T_{(a,b)}(\tau) := G_a(\tau) - G_b(\tau).$$

Pondo $M = q^k$ e rotulando C como $C = \{\mathbf{c}_1, \ldots, \mathbf{c}_M\}$, podemos identificar cada função de valor ν com a M-upla $(\nu(\mathbf{c}_1), \ldots, \nu(\mathbf{c}_M))$ e consequentemente interpretar a diferença das perdas esperadas $\mathbb{E}_C(a,b,\cdot)$ como sendo a restrição do funcional linear

$$E : \mathbb{R}^M \to \mathbb{R},$$

dado por $E_{(a,b)}(x_1, \ldots, x_M) = T_{(a,b)}(\mathbf{c}_1) \cdot x_1 + \ldots + T_{(a,b)}(\mathbf{c}_M) \cdot x_M$, ao primeiro octante $\mathbb{R}^M_+$:

$$\mathbb{E}_C(a,b,\cdot) = E_{(a,b)}\big|_{\mathbb{R}^M_+}.$$

Sendo assim, se $E_{(a,b)}$ é um operador não nulo e $K_{(a,b)}$ é o seu núcleo, então $K_{(a,b)}$ divide $\mathbb{R}^M$ em dois semi-espaços abertos. A saber,

$$\mathbb{R}^M = E^-_{(a,b)} \cup K_{(a,b)} \cup E^+_{(a,b)}$$

com

$$E^-_{(a,b)} := \left\{(x_1, \ldots, x_M) : E_{(a,b)}(x_1, \ldots, x_M) < 0\right\}$$

e

$$E^+_{(a,b)} := \left\{(x_1, \ldots, x_M) : E_{(a,b)}(x_1, \ldots, x_M) > 0\right\}.$$

Como

$$\eta = \left(T_{(a,b)}(\mathbf{c}_1), \ldots, T_{(a,b)}(\mathbf{c}_M)\right),$$

que é ortogonal ao hiperplano $K_{(a,b)}$, aponta exatamente na direção dos pontos $\mathbf{x} \in \mathbb{R}^M$ tal que $E(\mathbf{x}) > 0$, concluímos que $K_{(a,b)}$ intercepta $\mathcal{V}$ não trivialmente se, e só se, pelo menos duas coordenadas de η têm sinais trocados.

Em resumo, pondo

$$\mathcal{V}^-(a,b) := \mathcal{V} \cap E^-_{(a,b)}$$

e

$$\mathcal{V}^+(a,b) := \mathcal{V} \cap E^+_{(a,b)},$$

temos que:

Teorema 2.9 *Seja C um $[N;k]_q$ código linear e $a, b \in \mathcal{D}(C)$ dois decodificadores. Ambos os conjuntos $\mathcal{V}^-(a,b)$ e $\mathcal{V}^+(a,b)$ são não vazios se, e somente se, existem $\mathbf{c}, \mathbf{c}' \in C$ tal que*

$$T_{(a,b)}(\mathbf{c}) \cdot T_{(a,b)}(\mathbf{c}') < 0.$$

Fica fácil ver agora que as desigualdades em (2.13) e (2.14) são satisfeitas para uma infinidade de funções de valor.

2.8 $\mathbb{E}_C(a,b,\nu)$ para Canais q-ários Simétricos

Considere agora um canal q-ário simétrico com probabilidade de erro p e C um $[N;k]_q$ código linear. Assumindo que $M := |C|$ $(= q^k)$, segue de (2.2) que

$$\begin{aligned} G_a(\tau) &= \sum_{\mathbf{y}\in\mathbb{F}_q^N} P(\mathbf{y}|\, a(\mathbf{y}) - \tau)\, P(a(\mathbf{y}) - \tau) \\ &= \sum_{\mathbf{y}\in\mathbb{F}_q^N} (1-p)^N \left(\frac{p}{(1-p)(q-1)}\right)^{d_H(\mathbf{y},a(\mathbf{y})-\tau)} \frac{1}{M} \\ &= \frac{(1-p)^N}{M} \sum_{\mathbf{y}\in\mathbb{F}_q^N} \left(\frac{p}{(1-p)(q-1)}\right)^{d_H(\mathbf{y},a(\mathbf{y})-\tau)}, \end{aligned}$$

ou seja,

$$G_a(\tau) = \frac{(1-p)^N}{M} \sum_{\mathbf{y}\in\mathbb{F}_q^N} s^{d_H(\mathbf{y},a(\mathbf{y})-\tau)} \tag{2.16}$$

com $s = \frac{p}{(1-p)(q-1)}$. Consequentemente,

$$T_{(a,b)}(\tau) = \frac{(1-p)^N}{M} \sum_{\mathbf{y}\in\mathbb{F}_q^N} \left(s^{d_H(\mathbf{y},a(\mathbf{y})-\tau)} - s^{d_H(\mathbf{y},b(\mathbf{y})-\tau)}\right). \tag{2.17}$$

Este será exatamente o cenário onde serão desenvolvidos os resultados do próximo capítulo. O único resultado desta seção, similar aos Teoremas 2.7 e 2.8, segue nesta direção.

Teorema 2.10 *Seja C um $[N;k]_2$ código binário sobre um canal binário simétrico e $\widetilde{\mathbf{y}} = (1,1,\ldots,1)$. Assuma que a distribuição de probabilidades a priori de C é uniforme e que $\widetilde{\mathbf{y}} \notin C$. Se*

$$d_H(\widetilde{\mathbf{y}}, \mathbf{c}_1) = d_H(\widetilde{\mathbf{y}}, \mathbf{c}_2) = \min\{d_H(\widetilde{\mathbf{y}}, \mathbf{c}) : \mathbf{c} \in C\}$$

para algum par $\mathbf{c}_1, \mathbf{c}_2 \in C$ com $\mathbf{c}_1 \neq \mathbf{c}_2$, então existem decodificadores a e b do tipo NN tal que os conjuntos $\mathcal{V}^-(a,b)$ e $\mathcal{V}^+(a,b)$ são ambos não vazios.

Demonstração Seja $a : \mathbb{F}_2^N \to C$ um decodificador NN para C e assuma que $a(\widetilde{\mathbf{y}}) = \mathbf{c}_1$. Agora defina $b : \mathbb{F}_2^N \to C$ como sendo

$$b(\mathbf{y}) = \begin{cases} a(\mathbf{y}) & \text{se} \quad \mathbf{y} \neq \widetilde{\mathbf{y}} \\ \mathbf{c}_2 & \text{se} \quad \mathbf{y} = \widetilde{\mathbf{y}} \end{cases}.$$

Por construção temos que b é também um decodificador NN.

Afirmamos agora que sobre um canal binário simétrico $T_{(a,b)}(\mathbf{c}_1) \cdot T_{(a,b)}(\mathbf{c}_2) < 0$. De fato, como $a(\mathbf{y}) = b(\mathbf{y})$ para todo $\mathbf{y} \neq \widetilde{\mathbf{y}}$, segue de (2.17) que

$$T_{(a,b)}(\tau) = \frac{(1-p)^N}{M}\left(s^{d_H(\widetilde{\mathbf{y}}, a(\widetilde{\mathbf{y}})-\tau)} - s^{d_H(\widetilde{\mathbf{y}}, b(\widetilde{\mathbf{y}})-\tau)}\right).$$

Como $a(\widetilde{\mathbf{y}}) = \mathbf{c}_1$ e $b(\widetilde{\mathbf{y}}) = \mathbf{c}_2$, temos que

$$T_{(a,b)}(\mathbf{c}_1) = \frac{(1-p)^N}{M}\left(s^{d_H(\widetilde{\mathbf{y}}, \mathbf{0})} - s^{d_H(\widetilde{\mathbf{y}}, \mathbf{c}_2-\mathbf{c}_1)}\right)$$

e

$$T_{(a,b)}(\mathbf{c}_2) = \frac{(1-p)^N}{M}\left(s^{d_H(\widetilde{\mathbf{y}}, \mathbf{c}_1-\mathbf{c}_2)} - s^{d_H(\widetilde{\mathbf{y}}, \mathbf{0})}\right).$$

Nos resta analisar os expoentes $d_H(\widetilde{\mathbf{y}}, \mathbf{c}_1 - \mathbf{c}_2)$, $d_H(\widetilde{\mathbf{y}}, \mathbf{c}_2 - \mathbf{c}_1)$ e $d_H(\widetilde{\mathbf{y}}, \mathbf{0})$. Começamos observando que

$$d_H(\widetilde{\mathbf{y}}, \mathbf{c}) = N - w_H(\mathbf{c})$$

em $\mathbb{F}_2^N$. Segue daí que $d_H(\widetilde{\mathbf{y}}, \mathbf{c}) = N$ se, e só se, $\mathbf{c} = \mathbf{0}$. Como $\mathbf{c}_1 - \mathbf{c}_2 \neq \mathbf{0}$, concluímos que

$$d_H(\widetilde{\mathbf{y}}, \mathbf{c}_1 - \mathbf{c}_2) = d_H(\widetilde{\mathbf{y}}, \mathbf{c}_2 - \mathbf{c}_1) < N.$$

Consequentemente, como $d_H(\widetilde{\mathbf{y}}, \mathbf{0}) = N$,

$$T_{(a,b)}(\mathbf{c}_1) < 0 < T_{(a,b)}(\mathbf{c}_2).$$

O resultado segue agora do Teorema 2.9. □

2.9 Códigos Binários de Hamming e de Golay e os seus Codificadores de Bayes

Nesta seção todos os canais considerados são do tipo binário simétrico com probabilidade de erro p. Queremos determinar os codificadores de Bayes dos códigos binários de Hamming e de Golay referentes aos seus decodificadores ML.

Começamos com os *códigos binários de Hamming*. Estes códigos podem ser caracterizados simplesmente pelos parâmetros $\left[2^k - 1; 2^k - 1 - k; d_H = 3\right]_2$ (se dois códigos possuem estes parâmetros, então eles são equivalentes). Não é difícil mostrar que todo código binário de Hamming é perfeito. Vamos denotar qualquer um destes códigos por $\mathcal{H}(k)$. Os códigos binários de Hamming foram introduzidos em 1959 por Richard Hamming em [17].

O espectro de pesos de $\mathcal{H}(k)$ também é bem conhecido na literatura (ver [19]):

$$A_j(\mathcal{H}(k)) = 2^{-k}\left[\binom{2^k-1}{j} + \left(2^k-1\right)\left(\sum_{i=0}^{j}(-1)^i\binom{2^{k-1}}{i}\binom{2^k-1-2^{k-1}}{j-i}\right)\right].$$

Sabemos da Seção 2.8 que sobre um canal binário simétrico

$$G_{a_H}(\tau) = \frac{(1-p)^{2^k-1}}{M}\sum_{\mathbf{y}\in\mathbb{F}_2^{2^k-1}} s^{d_H(\mathbf{y},a_H(\mathbf{y})-\tau)}$$

onde $s := \frac{p}{1-p}$ e $M := |\mathcal{H}(k)|$. Como $\mathcal{H}(k)$ é perfeito e seu raio de empacotamento é igual a 1, se a_H é um decodificador ML de $\mathcal{H}(k)$, então

$$\mathbf{y} - a_H(\mathbf{y}) = \begin{cases} \mathbf{e}_i & \text{para algum } 1 \le i \le 2^k - 1 \text{ se } \mathbf{y} \notin \mathcal{H}(k) \\ \mathbf{0} & \text{caso contrário} \end{cases}$$

(ver o Exemplo 2.1). Isto assegura que

$$d_H(\mathbf{y}, a_H(\mathbf{y}) - \tau) = \begin{cases} w_H(\mathbf{e}_i + \tau) & \text{para algum } 1 \leq i \leq 2^k - 1 \text{ se } \mathbf{y} \notin \mathcal{H}(k) \\ w_H(\tau) & \text{caso contrário} \end{cases}$$

Consequentemente, como

$$\sum_{\mathbf{y} \in \mathbb{F}_2^{2^k-1}} s^{d_H(\mathbf{y}, a_H(\mathbf{y}) - \tau)} = \sum_{\mathbf{y} \in \mathcal{H}(k)} s^{d_H(\mathbf{y}, a_H(\mathbf{y}) - \tau)} + \sum_{\mathbf{y} \notin \mathcal{H}(k)} s^{d_H(\mathbf{y}, a_H(\mathbf{y}) - \tau)},$$

temos que

$$\sum_{\mathbf{y} \in \mathbb{F}_2^{2^k-1}} s^{d_H(\mathbf{y}, a_H(\mathbf{y}) - \tau)} = \sum_{\mathbf{y} \in \mathcal{H}(k)} s^{w_H(\tau)} + M \cdot \sum_{i \notin supp(\tau)} s^{w_H(\mathbf{e}_i + \tau)} + M \cdot \sum_{i \in supp(\tau)} s^{w_H(\mathbf{e}_i + \tau)}.$$

Agora como

$$\sum_{i \notin supp(\tau)} s^{w_H(\mathbf{e}_i + \tau)} = \left(2^k - 1 - w_H(\tau)\right) s^{w_H(\tau)+1}$$

e

$$\sum_{i \in supp(\tau)} s^{w_H(\mathbf{e}_i + \tau)} = w_H(\tau)\, s^{w_H(\tau)-1},$$

concluímos que

$$G_{a_H}(\tau) = (1-p)^{2^k-1} \left(s^{w_H(\tau)} + \left(2^k - 1 - w_H(\tau)\right) s^{w_H(\tau)+1} + w_H(\tau)\, s^{w_H(\tau)-1} \right).$$

Note agora que $G_{a_H}(\tau)$ depende somente de $w_H(\tau)$. Em outras palavras:

Teorema 2.11 *$G_{a_H}(\tau_1) = G_{a_H}(\tau_2)$ para todo $\tau_1, \tau_2 \in \mathcal{H}(k)$ tal que $w_H(\tau_1) = w_H(\tau_2)$.*

Agora estamos prontos para caracterizar os codificadores de Bayes de $\mathcal{H}(k)$ para a_H sobre um canal binário simétrico.

Teorema 2.12 *Suponha que $\mathcal{H}(k) = \{\mathbf{c}_1, \ldots, \mathbf{c}_M\}$ e que*

$$w_H(\mathbf{c}_1) \leq \ldots \leq w_H(\mathbf{c}_M).$$

Então, sobre um canal binário simétrico,

$$G_{a_H}(\mathbf{c}_1) \geq \ldots \geq G_{a_H}(\mathbf{c}_M).$$

Demonstração Suponha que $w_H(\mathbf{c}_i) = n$ e $w_H(\mathbf{c}_{i+1}) = n+m$ com $m \geq 0$. Já sabemos que

$$G_{a_H}(\mathbf{c}_i) = (1-p)^N \left(s^n + (N-n)\, s^{n+1} + ns^{n-1}\right).$$

onde $N := 2^k - 1$. Afirmamos agora que

$$s^n + (N-n)\, s^{n+1} + ns^{n-1} \geq s^{n+1} + (N-n-1)\, s^{n+2} + (n+1)\, s^n.$$

De fato, como

$$(N-n-1)\, s^{n+1} + ns^{n-1} \geq 0,$$

então $(N-n)\, s^{n+1} + ns^{n-1} \geq s^{n+1}$. Somando s^n em ambos os lados da última desigualdade obtemos que

$$s^n + (N-n)\, s^{n+1} + ns^{n-1} \geq s^n (s+1).$$

Assumindo que $0 < s < 1$ e multiplicarndo a desigualdade acima por $s-1$, concluímos que

$$s^{n+1} + (N-n-1)\, s^{n+2} + (n+1)\, s^n \leq s^n + (N-n)\, s^{n+1} + ns^{n-1}.$$

Consequentemente,

$$s^{n+m} + (N-n-m)\, s^{n+m+1} + (n+m)\, s^{n+m-1} \leq s^n + (N-n)\, s^{n+1} + ns^{n-1}$$

e daí que

$$G_{a_H}(\mathbf{c}_{i+1}) \leq G_{a_H}(\mathbf{c}_i)$$

já que $(1-p)^N > 0$. □

Segue do Teorema 2.3 que:

Teorema 2.13 *Se* $\mathcal{H}(k) = \{\mathbf{c}_1, \ldots, \mathbf{c}_M\}$ *com*

$$w_H(\mathbf{c}_1) \leq \ldots \leq w_H(\mathbf{c}_M),$$

então $f : \mathbb{F}_2^{2^k-1-k} \to \mathcal{H}(k)$ *é um codificador de Bayes para* a_H *e uma função de valor* ν *se, e somente se,*

$$\nu_f(\mathbf{c}_1) \leq \ldots \leq \nu_f(\mathbf{c}_M).$$

Agora seja $\mathcal{G}_{23}$ o $[23; 12; d_H = 7]_2$ *código binário de Golay* dado pela matriz geradora

$$G = [I_{12}|\, A]$$

onde I_{12} é a matriz identidade de ordem 12 e

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

Temos que $\mathcal{G}_{23}$ é um código perfeito com raio de empacotamento igual a 3 e seu espectro de pesos é dado por

$$A_0 = A_{23} = 1, A_7 = A_{16} = 253, A_8 = A_{15} = 506, A_{11} = A_{12} = 1288.$$

O código binário de Golay $\mathcal{G}_{23}$ foi introduzido em 1949 por Marcel Golay em [15].

Como o raio de empacotamento de $\mathcal{G}_{23}$ é igual a 3 e $\mathcal{G}_{23}$ é perfeito, se a_H é um decodificador ML para $\mathcal{G}_{23}$, então

$$\mathbf{y} - a_H(\mathbf{y}) = \begin{cases} \mathbf{e}_i \text{ ou } \mathbf{e}_i + \mathbf{e}_j \text{ ou } \mathbf{e}_i + \mathbf{e}_j + \mathbf{e}_k & \text{se } \mathbf{y} \notin \mathcal{G}_{23} \\ \mathbf{0} & \text{caso contrário} \end{cases}.$$

Sendo assim, sobre um canal binário simétrico, temos que:

$$G_{a_H}(\tau) = (1-p)^{23}\left(s^{w_H(\tau)} + \sum_{i=1}^{3}\binom{23-w_H(\tau)}{i}s^{w_H(\tau)+i} + \sum_{i=1}^{3}\binom{w_H(\tau)}{i}s^{w_H(\tau)-i}\right.$$

$$\left. + \sum_{i=0}^{1}\binom{23-w_H(\tau)}{i+1}w_H(\tau)\,s^{w_H(\tau)+i} + (23-w_H(\tau))\binom{w_H(\tau)}{2}s^{w_H(\tau)-1}\right).$$

Novamente temos que $G_{a_H}(\tau)$ só depende de $w_H(\tau)$, ou seja:

Teorema 2.14 *$G_{a_H}(\tau_1) = G_{a_H}(\tau_2)$ para todo $\tau_1, \tau_2 \in \mathcal{G}_{23}$ tal que $w_H(\tau_1) = w_H(\tau_2)$.*

Os gráficos das funções $G_{a_H}(\tau)$ para cada um dos possíveis pesos de $\mathcal{G}_{23}$ estão ilustrados na Figura 2.9 abaixo.

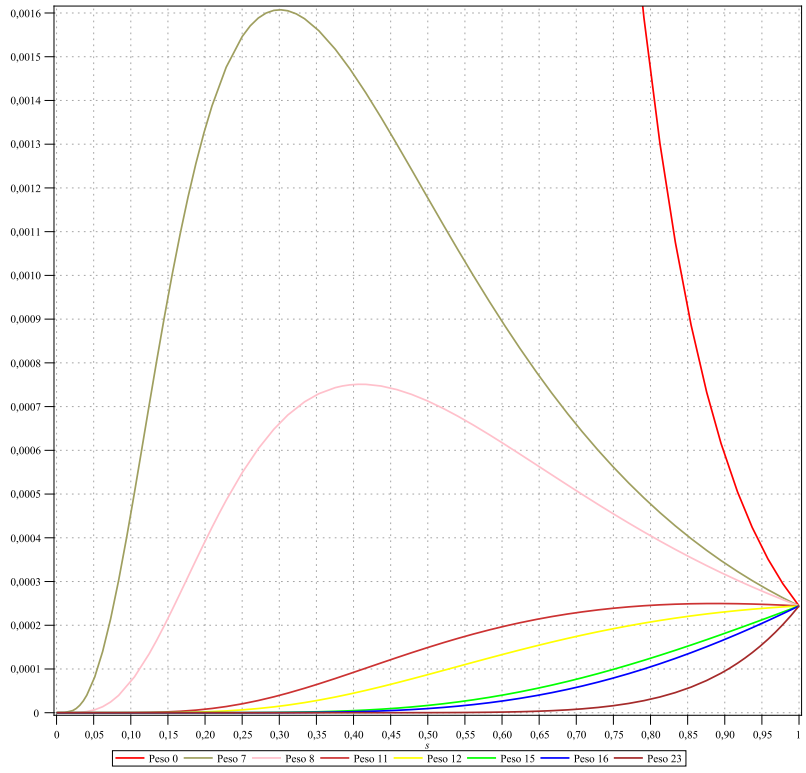

Figura 2.9: Os gráficos das funções $G_{a_H}(\tau)$ para cada um dos possíveis pesos $0, 7, 8, 11, 12, 15, 16, 23$ do código binário de Golay $\mathcal{G}_{23}$.

Como para todo $\mathbf{c}, \mathbf{c}' \in \mathcal{G}_{23}$ com $\mathbf{c} \neq \mathbf{c}'$ vale que $G_{a_H}(\mathbf{c}) = G_{a_H}(\mathbf{c}')$ se, e só se, $s = 0, 1$, segue do Teorema 2.3 que:

Teorema 2.15 *Se* $\mathcal{G}_{23} = \{\mathbf{c}_1, \ldots, \mathbf{c}_{2^{12}}\}$ *com*

$$w_H(\mathbf{c}_1) \leq \ldots \leq w_H(\mathbf{c}_{2^{12}}),$$

então $f : \mathbb{F}_2^{12} \to \mathcal{G}_{23}$ *é um codificador de Bayes para* a_H *e uma função de valor* ν *se, e somente se,*

$$\nu_f(\mathbf{c}_1) \leq \ldots \leq \nu_f(\mathbf{c}_{2^{12}}).$$

2.10 Funções de Valor para Canais Contínuos

O conceito de perda esperada total para canais discretos pode ser naturalmente adaptado para canais contínuos. Seremos simplistas na nossa formulação, considerando apenas os canais contínuos que são discretos no tempo (ver Definição 1.5). Também não explicitaremos os codificadores de canal.

Seja $S = \{\mathbf{s}_1, \ldots, \mathbf{s}_M\}$ uma (M, N) constelação de sinais sobre o espaço Euclidiano $\mathbb{R}^N$ e $a : \mathbb{R}^N \to S$ um decodificador para S. O desempenho do par (S, a) relativo a um canal contínuo $(\mathcal{X}, p(\mathcal{Y}|\mathcal{X}), \mathcal{Y})$ é medido pela probabilidade de erro de decodificação

$$P_e(S) = \sum_{\mathbf{s} \in S} P_e(\mathbf{s}) P(\mathbf{s})$$

onde

$$P_e(\mathbf{s}) = 1 - \int_{\mathbf{y}: a(\mathbf{y}) = \mathbf{s}} p(\mathbf{y}|\mathbf{s}) \, d\mathbf{y}$$

é a probabilidade do decodificador entregar ao destino um sinal distinto do sinal transmitido $\mathbf{s}$.

Adaptando as Definições 2.1 e 2.2 para canais contínuos, temos também que os decodificadores MAP minimizam a probabilidade de erro $P_e(S)$: para tanto basta reescrever $P_e(S)$ como

$$P_e(S) = \int_{\mathbb{R}^N} P_e(\mathbf{y}) P(\mathbf{y}) \, dy$$

onde $P_e(\mathbf{y}) = 1 - p(a(\mathbf{y})|\,\mathbf{y})$ é a probabilidade de termos transmitido um sinal distinto do sinal entregue pelo decodificador e $P(\mathbf{y}) = \sum_{\mathbf{s}\in S} p(\mathbf{y}|\,\mathbf{s})\,P(\mathbf{s})$. No caso em que a distribuição de probabilidades a priori $P(\mathbf{s})$ é uniforme, vale também que todo decodificador ML é um decodificador MAP e vice-versa.

Considere agora um canal Gaussiano sem memória (como no Exemplo 1.7), ou seja, um canal com funções de densidade de probabilidade dadas por

$$p(y|\,x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(y-x)^2}{2\sigma^2}\right]$$

tal que

$$p(y_1 \ldots y_N|\,x_1 \ldots x_N) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\sum_{i=1}^{N} \frac{(y_i - x_i)^2}{2\sigma^2}\right].$$

Neste caso temos que

$$\ln p(y_1 \ldots y_N|\,x_1 \ldots x_N) = -\frac{N}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - x_i)^2,$$

e daí fica fácil ver que $p(\mathbf{y}|\,\mathbf{x}) \geq p(\mathbf{y}|\,\mathbf{x}')$ se, e só se, a distância Euclidiana entre $\mathbf{y}$ e $\mathbf{x}$ é menor ou igual do que a distância Euclidiana entre $\mathbf{y}$ e $\mathbf{x}'$. Assim, se $S = \{\mathbf{s}_1, \ldots, \mathbf{s}_M\}$ é uma (M, N) constelação de sinais e $a : \mathbb{R}^N \to S$ é um decodificador ML, então as imagens inversas de a são *regiões de Voronoi* para S: se $d_E(\mathbf{x}, \mathbf{y})$ denota a distância Euclidiana entre $\mathbf{x}$ e $\mathbf{y}$, então

$$a^{-1}(\mathbf{s}_i) = \{\mathbf{y} : d_E(\mathbf{y}, \mathbf{s}_i) \leq d_E(\mathbf{y}, \mathbf{s}_j) \text{ para todo } j \neq i\}.$$

Esta é exatamente a versão do Teorema 2.2 para canais contínuos.

Agora seja $S = \{\mathbf{s}_1, \ldots, \mathbf{s}_M\}$ uma (M, N) constelação de sinais e

$$\Delta S = \{\mathbf{s}_i - \mathbf{s}_j : 1 \leq i, j \leq M\}$$

o *conjunto das diferenças* de S. Queremos determinar uma expressão para $P_e(S)$ similar a expressão (2.3) para $P_e(C)$. Como $a(\mathbf{y}) - \mathbf{s}_i$ não nescessariamente pertence a S, $\nu(a(\mathbf{y}) - \mathbf{s}_i)$ só faz sentido se assumirmos que ao menos ΔS seja o domínio de ν. Desta forma, considerando a aplicação $\nu : \Delta S \to \mathbb{R}_+$ dada por

$$\nu(\mathbf{s}_i - \mathbf{s}_j) = \begin{cases} 0 & \text{se} \quad i = j \\ 1 & \text{se} \quad i \neq j \end{cases},$$

podemos reescrever $P_e(S)$ pondo

$$P_e(S) = \int_{\mathbb{R}^N} \left(\sum_{i=1}^{M} \nu(a(\mathbf{y}) - \mathbf{s}_i)\, p(\mathbf{s}_i|\, \mathbf{y}) \right) P(\mathbf{y})\, d\mathbf{y}.$$

Esta é a chave para adaptarmos a Definição 2.7 para canais contínuos. Começamos com a noção de função de valor para constelações de sinais.

Definição 2.13 *Uma* ***função de valor*** *para uma constelação de sinais S é uma aplicação do tipo*

$$\nu : \Delta S \to \mathbb{R}_+.$$

Consequentemente:

Definição 2.14 *Fixe um canal contínuo sobre $\mathbb{R}$ e seja (S, a, ν) uma tripla onde S é uma (M, N) constelação de sinais, $a : \mathbb{R}^N \to S$ é um decodificador e $\nu : \Delta S \to \mathbb{R}_+$ é uma função de valor. Seja $\{P(\mathbf{s}) : \mathbf{s} \in S\}$ a distribuição de probabilidades a priori de S. Definimos a* ***perda esperada total*** *$\mathbb{E}_S(a, \nu)$ de S relativa a dupla (a, ν) como sendo a média*

$$\mathbb{E}_S(a, \nu) = \int_{\mathbb{R}^N} \mathbb{E}_{\mathbf{y}}(a, \nu)\, P(\mathbf{y})\, d\mathbf{y} \tag{2.18}$$

onde

$$\mathbb{E}_{\mathbf{y}}(a, \nu) = \sum_{\mathbf{s} \in S} \nu(a(\mathbf{y}) - \mathbf{s})\, p(\mathbf{s}|\, \mathbf{y})$$

é a ***perda esperada*** *em $\mathbf{y}$.*

Seguindo os passos do caso discreto, podemos caracterizar a perda esperada total em (2.18), identificando as funções de valor como variáveis, como sendo a restrição de um funcional linear. Vejamos os detalhes.

Colocando $S = \{\mathbf{s}_1, \ldots, \mathbf{s}_M\}$, temos que

$$\begin{aligned} \mathbb{E}_S(a, \nu) &= \int_{\mathbb{R}^N} \left[\sum_{i=1}^{M} \nu(a(\mathbf{y}) - \mathbf{s}_i)\, p(\mathbf{s}_i|\, \mathbf{y}) \right] P(\mathbf{y})\, d\mathbf{y} \\ &= \sum_{i=1}^{M} \int_{\mathbb{R}^N} \nu(a(\mathbf{y}) - \mathbf{s}_i)\, p(\mathbf{s}_i|\, \mathbf{y})\, P(\mathbf{y})\, d\mathbf{y}. \end{aligned}$$

Como $p(\mathbf{s}_i|\,\mathbf{y})\,P(\mathbf{y}) = p(\mathbf{y}|\,\mathbf{s}_i)\,P(\mathbf{s}_i)$, segue que

$$\begin{aligned}\mathbb{E}_S(a,\nu) &= \sum_{i=1}^{M}\int_{\mathbb{R}^N}\nu(a(\mathbf{y})-\mathbf{s}_i)\,p(\mathbf{y}|\,\mathbf{s}_i)\,P(\mathbf{s}_i)\,d\mathbf{y}\\ &= \sum_{i=1}^{M}\sum_{j=1}^{M}\int_{a^{-1}(\mathbf{s}_j)}\nu(\mathbf{s}_j-\mathbf{s}_i)\,p(\mathbf{y}|\,\mathbf{s}_i)\,P(\mathbf{s}_i)\,d\mathbf{y}\\ &= \sum_{i=1}^{M}\sum_{j=1}^{M}\nu(\mathbf{s}_j-\mathbf{s}_i)\int_{a^{-1}(\mathbf{s}_j)}p(\mathbf{y}|\,\mathbf{s}_i)\,P(\mathbf{s}_i)\,d\mathbf{y}.\end{aligned}$$

Obtemos assim que

$$\mathbb{E}_S(a,\nu) = \sum_{\tau\in\Delta S} G_a(\tau)\,\nu(\tau) \tag{2.19}$$

com

$$G_a(\tau) = \sum_{i,j:\mathbf{s}_j-\mathbf{s}_i=\tau}\int_{a^{-1}(\mathbf{s}_j)}p(\mathbf{y}|\,\mathbf{s}_i)\,P(\mathbf{s}_i)\,d\mathbf{y}.$$

Definindo $\mathbb{E}_S(a,b,\nu)$ como sendo a diferença entre $\mathbb{E}_S(a,\nu)$ e $\mathbb{E}_S(b,\nu)$,

$$\mathbb{E}_S(a,b,\nu) := \mathbb{E}_S(a,\nu) - \mathbb{E}_S(b,\nu),$$

segue de (2.19) que

$$\mathbb{E}_S(a,b,\nu) = \sum_{\tau\in\Delta S} T_{(a,b)}(\tau)\,\nu(\tau)$$

com

$$T_{(a,b)}(\tau) := G_a(\tau) - G_b(\tau)$$

Assumindo que $|\Delta S| = s$ e rotulando ΔS como $\Delta S = \{\tau_1,\ldots,\tau_s\}$, podemos identificar cada função de valor ν com a s-ulpa $(\nu(\tau_1),\ldots,\nu(\tau_s))$ e consequentemente interpretar a diferença $\mathbb{E}_S(a,b,\cdot)$ como sendo a restrição do funcional linear

$$E:\mathbb{R}^s\to\mathbb{R},$$

dado por $E_{(a,b)}(x_1,\ldots,x_s) = T_{(a,b)}(\tau_1)\cdot x_1 + \ldots + T_{(a,b)}(\tau_s)\cdot x_s$, ao primeiro octante $\mathbb{R}^s_+$:

$$\mathbb{E}_S(a,b,\cdot) = E_{(a,b)}\big|_{\mathbb{R}^S_+}.$$

Sendo assim, se $E_{(a,b)}$ é um operador não nulo e $K_{(a,b)}$ é o seu núcleo, então $K_{(a,b)}$ divide $\mathbb{R}^s$ em dois semi-espaços abertos. A saber,

$$\mathbb{R}^s = E^-_{(a,b)} \cup K_{(a,b)} \cup E^+_{(a,b)}$$

com

$$E^-_{(a,b)} := \left\{(x_1, \ldots, x_s) : E_{(a,b)}(x_1, \ldots, x_s) < 0\right\}$$

e

$$E^+_{(a,b)} := \left\{(x_1, \ldots, x_s) : E_{(a,b)}(x_1, \ldots, x_s) > 0\right\}.$$

Como

$$\eta = \left(T_{(a,b)}(\tau_1), \ldots, T_{(a,b)}(\tau_s)\right)$$

aponta exatamente na direção dos pontos $\mathbf{x} \in \mathbb{R}^s$ tal que $E(\mathbf{x}) > 0$, concluímos que $K_{(a,b)}$ intercepta $\mathcal{V}$, o conjunto das funções de valor, não trivialmente se, e só se, pelo menos duas coordenadas de η têm sinais trocados.

Em resumo, pondo

$$\mathcal{V}^-(a,b) := \{\nu \in \mathcal{V} : \mathbb{E}_S(a,b,\nu) < 0\}$$

e

$$\mathcal{V}^+(a,b) := \{\nu \in \mathcal{V} : \mathbb{E}_S(a,b,\nu) > 0\},$$

temos que:

Teorema 2.16 *Seja S uma constelação de sinais e a, b dois decodificadores para S. Ambos os conjuntos $\mathcal{V}^-(a,b)$ e $\mathcal{V}^+(a,b)$ são não vazios se, e somente se, existem $\tau, \tau' \in \Delta S$ tal que*

$$T_{(a,b)}(\tau) \cdot T_{(a,b)}(\tau') < 0.$$

Estabelecido o Teorema 2.16, mostraremos agora que os decodificadores ML para canais Gaussianos, determinados pelas regiões de Voronoi, não são nescessariamente os melhores decodificadores. Isto será uma consequência do seguinte resultado:

Teorema 2.17 *Seja* $S = \{\mathbf{s}_1, \ldots, \mathbf{s}_M\}$ *uma* (M, N) *constelação de sinais tal que para algum* $\tau \in \Delta S$ *exista um único par* $(X, Y) \in S \times S$ *tal que* $Y - X = \tau$. *Considere um decodificador* $a : \mathbb{R}^N \to S$ *com regiões de decisão com interiores não vazios. Então existe um decodificador* $b : \mathbb{R}^N \to S$ *tal que ambos os conjuntos* $\mathcal{V}^- (a, b)$ *e* $\mathcal{V}^+ (a, b)$ *são não vazios.*

Demonstração Assuma inicialmente que $(X, Y) = (\mathbf{s}_i, \mathbf{s}_j)$ é a única solução da equação $Y - X = \tau$ em $S \times S$. Sejam $a^{-1}(\mathbf{s}_1), \ldots, a^{-1}(\mathbf{s}_M)$ as regiões de decisão do decodificador $a : \mathbb{R}^N \to S$. Considere uma partição

$$\{R(\mathbf{s}_i), R(\mathbf{s}_j)\}$$

de $a^{-1}(\mathbf{s}_i) \cup a^{-1}(\mathbf{s}_j)$, distinta da partição $\{a^{-1}(\mathbf{s}_i), a^{-1}(\mathbf{s}_j)\}$, tal que $R(\mathbf{s}_i) = a^{-1}(\mathbf{s}_i) \cup S_j$ para algum conjunto aberto $S_j \subseteq a^{-1}(\mathbf{s}_j)$ tal que $\mathbf{s}_j \notin S_j$. É claro que tal partição existe já que as regiões de decisão de a possuem interior não vazio. Sob estas condições temos que $R(\mathbf{s}_j) = a^{-1}(\mathbf{s}_j) - S_j$.

Agora seja

$$b : \mathbb{R}^N \to S$$

o decodificador determinado pelas regiões de decisão

$$\left\{a^{-1}(\mathbf{s}_1), \ldots, \widehat{a^{-1}(\mathbf{s}_i)}, \ldots, \widehat{a^{-1}(\mathbf{s}_j)}, \ldots, a^{-1}(\mathbf{s}_M)\right\} \cup \{R(\mathbf{s}_i), R(\mathbf{s}_j)\}.$$

Como $(X, Y) = (\mathbf{s}_i, \mathbf{s}_j)$ é a única solução da equação $Y - X = \tau$ em $S \times S$, então $(X, Y) = (\mathbf{s}_j, \mathbf{s}_i)$ é também a única solução da equação $Y - X = -\tau$ em $S \times S$. Disto segue que

$$G_a(\mathbf{s}_j - \mathbf{s}_i) = \int_{a^{-1}(\mathbf{s}_j)} p(\mathbf{y}|\,\mathbf{s}_i)\, P(\mathbf{s}_i)\, d\mathbf{y},$$

$$G_b(\mathbf{s}_j - \mathbf{s}_i) = \int_{a^{-1}(\mathbf{s}_j) - S_j} p(\mathbf{y}|\,\mathbf{s}_i)\, P(\mathbf{s}_i)\, d\mathbf{y},$$

$$G_a(\mathbf{s}_i - \mathbf{s}_j) = \int_{a^{-1}(\mathbf{s}_i)} p(\mathbf{y}|\,\mathbf{s}_j)\, P(\mathbf{s}_j)\, d\mathbf{y}$$

e

$$G_b(\mathbf{s}_i - \mathbf{s}_j) = \int_{a^{-1}(\mathbf{s}_i) \cup S_j} p(\mathbf{y}|\,\mathbf{s}_j)\, P(\mathbf{s}_j)\, d\mathbf{y}.$$

Consequentemente,

$$T_{(a,b)}(\mathbf{s}_i - \mathbf{s}_i) = \int_{S_j} p(\mathbf{y}|\,\mathbf{s}_i)\, P(\mathbf{s}_i)\, d\mathbf{y} > 0$$

e

$$T_{(a,b)}(\mathbf{s}_i - \mathbf{s}_j) = -\int_{S_j} p(\mathbf{y}|\,\mathbf{s}_j)\, P(\mathbf{s}_j)\, d\mathbf{y} < 0,$$

e o resultado segue do Teorema 2.16. □

Restringindo o Teorema 2.17 para canais Gaussianos:

Corolário 2.1 *Se S é uma constelação de sinais que satisfaz a condição do Teorema 2.17 e a_{ML} é um decodificador ML para S, então existem funções de valor para as quais o decodificador a_{ML} não minimiza a perda esperada total.*

Seja $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_M\}$ uma constelação de sinais sobre $\mathbb{R}^2$. Diremos que S é do tipo M-PSK (Phase-Shift Keying) se $\mathbf{v}_1, \ldots, \mathbf{v}_M$ são vértices de um polígono regular P_M de M lados centrado na origem $\mathbf{0}$.

Corolário 2.2 *Se S é um constelação de sinais do tipo M-PSK com M ímpar e a_{ML} é um decodificador ML para S, então existem funções de valor para as quais o decodificador a_{ML} não minimiza a perda esperada total.*

Demonstração Seja P_M um polígono regular com $M = 2n + 1$ lados centrado na origem. Rotule os vértices de P_M com $\mathbf{v}_1, \ldots, \mathbf{v}_M$ (nesta ordem e no sentido horário). Suponha também que P_M está inscrito no círculo S_1 de raio 1. Nestas condições, as mediatrizes dos lados de P_M são exatamente as retas $\overline{\mathbf{0}\mathbf{v}_i}$ que passam pela origem $\mathbf{0}$ e pelos vértices $\mathbf{v}_i$. Isto assegura que $-\mathbf{v}_i$ é um ponto sobre S_1 localizado entre dois vértices consecutivos de P_M. Se $-\mathbf{v}_i$ está localizado entre os vértices $\mathbf{v}_{l_i}$ e $\mathbf{v}_{l_i+1}$, então o comprimento de $\mathbf{v}_j - \mathbf{v}_i$ é máximo somente quando $\mathbf{v}_j = \mathbf{v}_{l_i}$ ou $\mathbf{v}_j = \mathbf{v}_{l_i+1}$. Como $\mathbf{v}_{l_i} - \mathbf{v}_i$ não é paralelo a $\mathbf{v}_{l_j} - \mathbf{v}_j$ para todo $j \neq i$, concluímos que a equação $Y - X = \mathbf{v}_{l_i} - \mathbf{v}_i$ admite uma única solução (X, Y) em $\{\mathbf{v}_1, \ldots, \mathbf{v}_M\} \times \{\mathbf{v}_1, \ldots, \mathbf{v}_M\}$. □

Se S é uma constelação do tipo M-PSK com M par e $\mathbf{v}, \mathbf{v}' \in S$, então $-\mathbf{v}$ e $-\mathbf{v}'$ também pertencem a S. Isto assegura que a equação $Y - X = \mathbf{v} - \mathbf{v}'$ admite duas soluções distintas em $S \times S$: $(X, Y) = (\mathbf{v}', \mathbf{v})$ e $(X, Y) = (-\mathbf{v}, -\mathbf{v}')$. Como $\mathbf{v}, \mathbf{v}'$ são arbitrários, concluímos que a equação $Y - X = \tau$ sempre admite mais de uma solução em $S \times S$ independentemente de $\tau \in \Delta S$. Portanto as constelações do tipo M-PSK com M par não satisfazem a condição do Teorema 2.17.

Capítulo 3

Métricas Poset

Neste capítulo apresentamos a segunda contribuição desta tese: a viabilidade da transmissão de informações com valor sobre um canal q-ário simétrico considerando uma família de decodificadores que são compatíveis com a idéia de informação com valor, no sentido de que as informações com maiores valores são protegidas por vizinhanças de informações com valores similares. Faremos isto considerando a família das *métricas poset* e os decodificadores por máxima proximidade relativos a estas métricas.

3.1 Métricas Poset

As métricas poset foram introduzidas em 1995 por Brualdi, Graves e Lawrence em [6] para generalizar um problema de Niederreiter de 1987 formulado em [29]. Vejamos os detalhes.

Começamos estabelecendo o conceito de *ideal* para conjuntos parcialmente ordenados. Seja $[N] := \{1, 2, \ldots, N\}$ e $P = ([N], \leq)$ uma ordem parcial sobre $[N]$, que também chamaremos de *poset* (uma abreviação para ***p**artially **o**rdered **set***). Um subconjunto I de P é dito um *ideal* de P se para todo $y \in I$ e $x \leq y$ tivermos que $x \in I$. Dado um subconjunto X de $[N]$, o menor ideal de P contendo X será denotado por $\langle X \rangle$ e será chamado de *ideal gerado* por X. Diremos que dois elementos x, y em P são *comparáveis* se $x \leq y$ ou $y \leq x$. Uma ordem parcial P tal

que todo par de elementos é comparável é dita uma *ordem total* ou *ordem cadeia*. Um ideal I de P é dito uma *cadeia* em P se I com a ordem induzida de P é uma ordem total. Se nenhum par de elementos de P é comparável, diremos então que P é uma *anti-cadeia* ou uma *ordem de Hamming*. Como referência para ordens parciais, indicamos o livro de Richard Stanley [40].

Agora seja

$$H = \{\mathbf{h}_i : 1 \leq i \leq N\}$$

um *sistema* sobre $\mathbb{F}_q^m$ rotulado por uma ordem parcial $P = ([N], \leq)$. Defina $\rho_P(H)$ como sendo o menor inteiro positivo d tal que existe um ideal I de P contendo d elementos com $\{\mathbf{h}_i : i \in I\}$ linearmente dependente. Se não existe tal ideal, definimos $\rho_P(H)$ como sendo $N+1$. O *problema de Brualdi, Graves e Lawrence* consiste em determinar o número

$$\rho_q(P; m) = \max_H \rho_P(H).$$

O problema de Brualdi *et al.* corresponde ao problema de Niederreiter quando P é uma união disjunta de cadeias de comprimento $n_1, \ldots, n_s$.

Agora podemos estabelecer o conceito de *métrica poset* (ver [6]):

Definição 3.1 *Dados* $\mathbf{x} = (x_1, \ldots, x_N)$ *e* $\mathbf{y} = (y_1, \ldots, y_N)$ *em* $\mathbb{F}_q^N$ *e um poset* $P = ([N], \leq)$*, definimos a* P***-distância*** $d_P(\mathbf{x}, \mathbf{y})$ *entre* $\mathbf{x}$ *e* $\mathbf{y}$ *pondo*

$$d_P(\mathbf{x}, \mathbf{y}) = |\langle i : x_i \neq y_i \rangle|.$$

O P***-peso*** *de* $\mathbf{x}$ *é o número* $w_P(\mathbf{x}) = d_P(\mathbf{0}, \mathbf{x})$.

Definindo a *P-distância mínima* $d_P(C)$ de um código C como sendo

$$d_P(C) = \min\{d_P(\mathbf{c}, \mathbf{c}') : \mathbf{c}, \mathbf{c}' \in C, \mathbf{c} \neq \mathbf{c}'\}$$

e supondo que H é uma matriz de verificação de paridade de C, concluímos que $\rho_P(H)$ coincide com $d_P(C)$. Consequentemente, $\rho_q(P; m)$ coincide com a maior P-distância mínima dada por um $[N; N-m]_q$ código linear. Note agora que se P é a ordem anti-cadeia, então d_P coincide com a métrica de Hamming d_H e, consequentemente, $\rho_q(P; m)$ coincide com a clássica função

$$K_q[N, N-m] := \max\left\{d_H(C) : C \text{ é um } [N; N-m]_q \text{ código}\right\},$$

amplamente estudada em combinatória (ver por exemplo [7]).

Assim como nos espaços de Hamming, definimos a *P-bola* de centro $\mathbf{x}$ e raio r como sendo o conjunto

$$B_P(\mathbf{x}; r) = \left\{\mathbf{y} \in \mathbb{F}_q^N : d_P(\mathbf{x}, \mathbf{y}) \leq r\right\}.$$

O *P-raio de empacotamento* $R_P(C)$ de um código C é definido como sendo o maior raio r tal que as P-bolas de raio r com centro nos elementos de C são duas a duas disjuntas. O código C é dito *P-perfeito* se as P-bolas de raio $R_P(C)$ cobrem todo o espaço $\mathbb{F}_q^N$, ou seja,

$$\bigcup_{\mathbf{c}\in C} B_P(\mathbf{c}; R_P(C)) = \mathbb{F}_q^N.$$

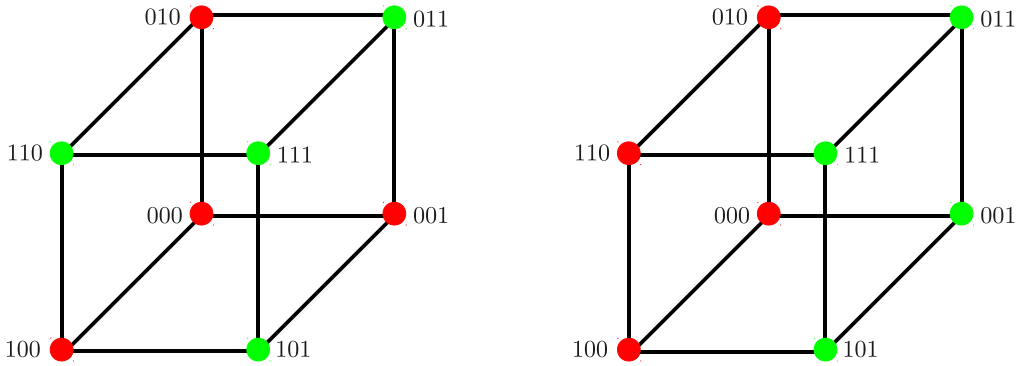

Figura 3.1: À esquerda, representação geométrica das H-bolas, H dado pela ordem de Hamming, de raio 1 centradas em 000 (vermelho) e 111 (verde), respectivamente. À direita, representação geométrica das P-bolas, P dado por $1 < 2 < 3$, de raio 2 centradas em 000 (vermelho) e 111 (verde), respectivamente. Com isto temos que $C = \{000, 111\}$ é H-perfeito e P-perfeito com $R_H(C) = 1$ e $R_P(C) = 2$.

As P-métricas foram introduzidas por Brualdi, Graves e Lawrence em [6]. Desde a sua introdução em 1995 muitas contribuições foram estabelecidas para a *Teoria dos Códigos Poset*. Os trabalhos sobre existência de códigos perfeitos (ver [6], [20], [22]), identidades do tipo MacWilliams (ver [1], [24]), dualidade de Wei (ver [28]), códigos MDS (ver [21]), grupos de isometrias (ver [31]) e formas sistemáticas (ver [12], [33]) são exemplos destas contribuições. Algumas famílias particulares de métricas poset

recebem mais atenção, caso esse das métricas de Niederreiter-Rosenbloom-Tsfasman (ver [3], [32], [33], [37]), que têm potêncial aplicação em problemas envolvendo canais de desvanecimento (*fading channels*) (ver [37], [41]).

3.2 Decodificadores P-NN

Temos dois bons motivos para considerar as métricas poset no contexto de informações com valor: primeiro, a baixa complexidade dos decodificadores por máxima proximidade relativos a algumas destas métricas (ver [33], por exemplo); segundo, a compatibilade dos decodificadores por máxima proximidade com a idéia de informação com valor, no sentido de que as informações com maiores valores são cercadas por vizinhanças de informações com valores similares.

Começamos restringindo o espaço dos decodificadores $\mathcal{D}(C)$ para a classe dos decodificadores por máxima proximidade relativos às métricas poset.

Definição 3.2 *Seja C um $[N;k]_q$ código linear e $P = ([N], \leq)$ um poset. Diremos que $a \in \mathcal{D}(C)$ é um P-**decodificador por máxima proximidade** (ou simplesmente um decodificador P-NN), se*

$$d_P(\mathbf{y}, a(\mathbf{y})) = \min\{d_P(\mathbf{y}, \mathbf{c}) : \mathbf{c} \in C\}$$

para todo $\mathbf{y} \in \mathbb{F}_q^N$. Neste caso escrevemos a_P para indicar que o decodificador a é do tipo P-NN. No caso particular em que P é uma anti-cadeia, escreveremos a_H e diremos que a é do tipo H-NN.

As questões de compatibilidade e complexidade ainda não serão exploradas neste capítulo. Antes precisamos responder a seguinte questão:

> *Dados dois posets P e Q sobre $[N]$, exite um $[N;k]_q$ código linear C e decodificadores a_P e a_Q de C tais que $\mathcal{V}^+(a_P, a_Q)$ e $\mathcal{V}^-(a_P, a_Q)$ são ambos não vazios?*

Uma resposta positiva para a questão acima significa que todo decodificador P-NN é relevante, dependendo do código e da função de valor em consideração.

Como mencionamos na introdução deste capítulo, daqui em diante todos os nossos esforços estarão voltados para os canais q-ários simétricos. Sendo assim, assumindo que o codificador de canal está fixo, se C é um $[N;k]_q$ código linear, a_P e a_Q são decodificadores P-NN e Q-NN de C e p é a probabilidade de erro do canal, segue de (2.17) que

$$\mathbb{E}_C\left(a_P, a_Q, \nu\right) = \sum_{\tau \in C} T_{\left(a_P, a_Q\right)}\left(\tau\right) \nu\left(\tau\right)$$

com

$$T_{\left(a_P, a_Q\right)}\left(\tau\right) = \frac{(1-p)^N}{M} \sum_{\mathbf{y} \in \mathbb{F}_q^N} \left(s^{d_H(\mathbf{y}, a_P(\mathbf{y}) - \tau)} - s^{d_H\left(\mathbf{y}, a_Q(\mathbf{y}) - \tau\right)} \right)$$

onde $M := |C|$ e $s := \frac{p}{(1-p)(q-1)}$.

Encerramos esta seção apresentando um exemplo modesto. Reservaremos o Capítulo 4 para um exemplo mais elaborado.

Exemplo 3.1 *Queremos calcular a perda esperada do* $[3;1]_2$ *código binário*

$$C = \{\mathbf{c}_0 = 000, \mathbf{c}_1 = 111\}$$

para cada uma das seguintes ordens: *a ordem anti-cadeia* P_1; *a ordem cadeia* P_2 *dada pelas relações* $1 < 2 < 3$; *e a ordem* P_3 *dada pelas relações* $1 < 2$ *e* $3 < 2$.

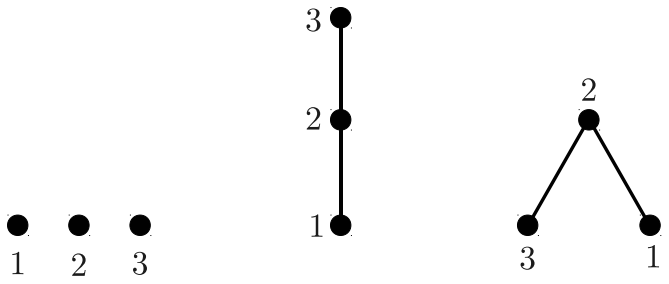

Figura 3.2: Os diagramas de Hasse de P_1, P_2 e P_3, respectivamente.

Começamos observando que C *é perfeito para as três métricas* $d_{P_1}, d_{P_2}, d_{P_3}$: *de fato, como*

$$B_{P_1}\left(\mathbf{c}_0; 1\right) = \{000, 100, 010, 001\} \;\; e \;\; B_{P_1}\left(\mathbf{c}_1; 1\right) = \{111, 110, 101, 011\},$$

$$B_{P_2}(\mathbf{c}_0;2) = \{000,100,010,110\} \;\; e \; B_{P_2}(\mathbf{c}_1;2) = \{111,011,101,001\},$$

$$B_{P_3}(\mathbf{c}_0;2) = \{000,100,001,101\} \;\; e \; B_{P_3}(\mathbf{c}_1;2) = \{111,011,110,010\},$$

segue que

$$B_{P_1}(\mathbf{c}_0;1) \cup B_{P_1}(\mathbf{c}_1;1) = B_{P_2}(\mathbf{c}_0;2) \cup B_{P_2}(\mathbf{c}_1;2) = B_{P_3}(\mathbf{c}_0;2) \cup B_{P_3}(\mathbf{c}_1;2) = \mathbb{F}_2^3$$

com

$$B_{P_1}(\mathbf{c}_0;1) \cap B_{P_1}(\mathbf{c}_1;1) = B_{P_2}(\mathbf{c}_0;2) \cap B_{P_2}(\mathbf{c}_1;2) = B_{P_3}(\mathbf{c}_0;2) \cap B_{P_3}(\mathbf{c}_1;2) = \varnothing.$$

O fato de C ser perfeito implica que os decodificadores $a_{P_1}, a_{P_2}, a_{P_3}$ são únicos para C. Agora podemos calcular as perdas esperadas:

$$\mathbb{E}_C(a_{P_1},\nu) = \frac{(1-p)^3}{2}\left((2+6s)\,\nu(\mathbf{c}_0) + \left(6s^2+2s^3\right)\nu(\mathbf{c}_1)\right),$$

$$\mathbb{E}_C(a_{P_2},\nu) = \frac{(1-p)^3}{2}\left(\left(2+4s+2s^2\right)\nu(\mathbf{c}_0) + \left(2s+4s^2+2s^3\right)\nu(\mathbf{c}_1)\right),$$

$$\mathbb{E}_C(a_{P_3},\nu) = \frac{(1-p)^3}{2}\left(\left(2+4s+2s^2\right)\nu(\mathbf{c}_0) + \left(2s+4s^2+2s^3\right)\nu(\mathbf{c}_1)\right).$$

Consequentemente,

$$\mathbb{E}_C(a_{P_1},a_{P_2},\nu) = \mathbb{E}_C(a_{P_1},a_{P_3},\nu) = \frac{(1-p)^3}{2}\left(\left(2s-2s^2\right)\nu(\mathbf{c}_0) + \left(2s^2-2s\right)\nu(\mathbf{c}_1)\right)$$

e

$$\mathbb{E}_C(a_{P_1},a_{P_2},\nu) = 0.$$

Temos assim que

$$\mathcal{V}^+(a_{P_1},a_{P_2}) = \mathcal{V}^+(a_{P_1},a_{P_3}) = \{\nu \in \mathcal{V} : \nu(\mathbf{c}_0) > \nu(\mathbf{c}_1)\},$$

$$\mathcal{V}^-(a_{P_1},a_{P_2}) = \mathcal{V}^-(a_{P_1},a_{P_3}) = \{\nu \in \mathcal{V} : \nu(\mathbf{c}_0) < \nu(\mathbf{c}_1)\}$$

e

$$\mathcal{V}^+(a_{P_2},a_{P_3}) = \mathcal{V}^-(a_{P_2},a_{P_3}) = \varnothing.$$

3.3 Códigos de Brualdi-Graves-Lawrence

No capítulo anterior mostramos que para todo código linear C sempre existe um decodificador b e uma função de valor ν para os quais nenhum decodificador ML é melhor do que b em relação a função de valor ν (veja o Teorema 2.7). Agora exibiremos uma família de posets satisfazendo a seguinte propriedade: para quaisquer dois posets P e Q dessa família, sempre existe um código linear C e decodificadores a_P e a_Q de C tais que $\mathcal{V}^+(a_P, a_Q)$ e $\mathcal{V}^-(a_P, a_Q)$ são ambos não vazios.

Começamos com as definições básicas. O *poset dual* $P^* = ([N], \leq^*)$ de um poset $P = ([N], \leq)$ é definido pela relação de ordem oposta: $x \leq^* y \Leftrightarrow y \leq x$. Os ideais de P^* serão chamados de *filtros*. Dado um filtro não trivial I e um subconjunto J não vazio e próprio de I, definimos

$$I_J^+ := \{i \in I - J : i > j \text{ para algum } j \in J\}$$

e

$$I_J^- := \{i \in I - J : i < j \text{ para algum } j \in J\}.$$

Diremos que um filtro I é *J-decomponível* se

$$I = I_J^+ \cup J \cup I_J^-$$

é uma partição de I com ambos os conjuntos I_J^+ e I_J^- não vazios. Se existe um filtro I de P que é J-decomponível, diremos então que P é *(I, J)-decomponível*.

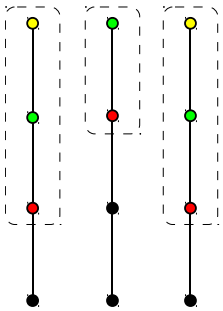

Figura 3.3: Um filtro J-decomponível com J =“vértices verdes”, I_J^+ =“vértices amarelos” e I_J^- =“vértices vermelhos”.

Seja $\{\mathbf{e}_i : 1 \leq i \leq n\}$ a base canônica de $\mathbb{F}_q^N$. Para cada subconjunto não vazio $X \subseteq [N]$ seja

$$C_X := span\{\mathbf{e}_i : i \in X\},$$

o espaço gerado por $\{\mathbf{e}_i : i \in X\}$. Em outras palavras, C_X é o espaço coordenado com suporte em X. A projeção de $\mathbf{y} \in \mathbb{F}_q^N$ sobre C_X será denotada por $\mathbf{y}_X$, ou seja, se $\mathbf{y} = \sum_{i=1}^N y_i\mathbf{e}_i$ então

$$\mathbf{y}_X = \sum_{i \in X} \mathbf{y}_i e_i.$$

Seja P a ordem cadeia dada pelas relações $1 < 2 < \ldots < N$ e considere o filtro $I = \{N-k+1, \ldots, N\}$. Nestas condições temos que C_I é P-perfeito com raio de empacotamento $R_P(C) = N - k$: de fato, se $\mathbf{c} = (0, \ldots, 0, c_{N-k+1}, \ldots, c_N) \in C$, então

$$B_P(\mathbf{c}; N-k) = \{(x_1, \ldots, x_{n-k}, c_{N-k+1}, \ldots, c_N) : \ x_i \in \mathbb{F}_q, 1 \leq i \leq n-k\};$$

daí que $\{B_P(\mathbf{c}; N-k) : \mathbf{c} \in C\}$ cobre $\mathbb{F}_q^N$; é fácil ver agora que

$$B_P(\mathbf{c}; N-k) \cap B_P(\mathbf{c}'; N-k) = \varnothing$$

para todo par $\mathbf{c} \neq \mathbf{c}' \in C$. Este fato foi estabelecido primeiramente por Brualdi, Graves e Lawrence em [6]. Para estes códigos C_I temos que:

Teorema 3.1 *Seja P a ordem cadeia dada pelas relações $1 < 2 < \ldots < N$, H a ordem de Hamming e $I = \{N-k+1, \ldots, N\}$. Temos então que $\mathbb{E}_{C_I}(a_H, a_P, \nu) = 0$ para toda função de valor ν.*

Demonstração É suficiente mostrar que $a_H(\mathbf{y}) = a_P(\mathbf{y})$ para todo $\mathbf{y} \in \mathbb{F}_q^N$. Começamos observando que se $\mathbf{y} = (y_1, \ldots, y_N)$, então

$$\mathbf{y} = \mathbf{x} + \mathbf{c}_\mathbf{y}$$

com $\mathbf{x} = (y_1, \ldots, y_{N-k}, 0, \ldots, 0) \in \mathbb{F}_q^N$ e $\mathbf{c}_\mathbf{y} = (0, \ldots, 0, y_{N-k+1}, \ldots, y_N) \in C_I$. Note que $\mathbf{c}_\mathbf{y} = \mathbf{y}_I$. Assim, se $\mathbf{c} \in C_I$, então

$$d_H(\mathbf{y}, \mathbf{c}) = w_H(\mathbf{x}) + d_H(\mathbf{c}_\mathbf{y}, \mathbf{c})$$

e

$$d_P(\mathbf{y},\mathbf{c}) = \begin{cases} d_P(\mathbf{c_y},\mathbf{c}) & \text{se} \quad \mathbf{c_y} \neq \mathbf{c} \\ d_P(\mathbf{x},\mathbf{0}) & \text{se} \quad \mathbf{c_y} = \mathbf{c} \end{cases}.$$

Isto mostra que as distâncias $d_H(\mathbf{y},\mathbf{c})$ e $d_P(\mathbf{y},\mathbf{c})$ dependem somente das distâncias $d_H(\mathbf{c_y},\mathbf{c})$ e $d_P(\mathbf{c_y},\mathbf{c})$, respectivamente. Consequentemente, se a_H e a_P são decodificadores H-NN e P-NN de C, então

$$a_H(\mathbf{y}) = \mathbf{c_y} = a_P(\mathbf{y})$$

para todo $\mathbf{y} \in \mathbb{F}_q^N$. □

Os argumentos na demonstração do Teorema 3.1 continuam valendo se P é uma ordem qualquer e I é um filtro de P. Se pretendemos construir decodificadores P-NN que sejam distintos do decodificador H-NN precisamos "*abrir buracos em* I". Essa é a razão para definirmos os códigos BGL.

Definição 3.3 *Seja I um filtro J-decomponível de $P = ([N], \leq)$, com $|I| = K$ e $|J| = k$. Definimos o* ***código BGL*** *(Brualdi-Graves-Lawrence) $C_{(I,J)}$ como sendo o espaço coordenado C_{I-J}, ou seja,*

$$C_{(I,J)} := span\{\mathbf{e}_i : i \in I - J\}.$$

Nestas condições temos que $C_{(I,J)}$ é um $[N; K-k]_q$ código linear.

Usaremos o conjunto

$$d_P\left(\mathbf{y}, C_{(I,J)}\right) := \left\{\mathbf{c} \in C_{(I,J)} : d_P(\mathbf{c},\mathbf{y}) \leq d_P(\mathbf{c}',\mathbf{y}) \text{ para todo } \mathbf{c}' \in C\right\}$$

para determinar os possíveis decodificadores P-NN de $C_{(I,J)}$. Denotaremos o complementar de um subconjunto X de $[N]$ por X^c.

Teorema 3.2 *Seja $P = ([N], \leq)$ uma ordem (I,J)-decomponível e $H = ([N], \leq)$ a ordem de Hamming. Para o $[N; |I| - |J|]_q$ código BGL $C_{(I,J)}$ existe um decodificador P-NN a_P e $\tau, \tau' \in C_{(I,J)}$ tais que*

$$T_{(a_H,a_P)}(\tau) \cdot T_{(a_H,a_P)}(\tau') < 0$$

para todo $0 < s < 1$. Consequentemente, $\mathcal{V}^+(a_H,a_P)$ e $\mathcal{V}^-(a_H,a_P)$ são ambos não vazios.

Demonstração Começamos observando que todo vetor $\mathbf{y}$ de $\mathbb{F}_q^N$ pode ser decomposto como

$$\mathbf{y} = \mathbf{y}_{I^c} + \mathbf{y}_J + \mathbf{y}_{I-J}$$

onde $\mathbf{y}_{I^c}$, $\mathbf{y}_J$ e $\mathbf{y}_{I-J}$ são as projeções de $\mathbf{y}$ nos espaços coordenados C_{I^c}, C_J e $C_{(I,J)}$, respectivamente. Desta decomposição segue que

$$d_H(\mathbf{y}, \mathbf{c}) = w_H(\mathbf{y}_{I^c}) + w_H(\mathbf{y}_J) + d_H(\mathbf{y}_{I-J}, \mathbf{c})$$

para todo $\mathbf{c} \in C_{(I,J)}$. Isto mostra que a distância $d_H(\mathbf{y}, \mathbf{c})$ depende somente da distância $d_H(\mathbf{y}_{I-J}, \mathbf{c})$. Consequentemente,

$$d_H\left(\mathbf{y}, C_{(I,J)}\right) = d_H\left(\mathbf{y}_{I-J}, C_{(I,J)}\right).$$

Como $d_H\left(\mathbf{y}_{I-J}, C_{(I,J)}\right) = \{\mathbf{y}_{I-J}\}$, segue que $d_H\left(\mathbf{y}, C_{(I,J)}\right) = \{\mathbf{y}_{I-J}\}$. Sendo assim, se a_H é um decodificador H-NN para $C_{(I,J)}$, então

$$a_H(\mathbf{y}) = \mathbf{y}_{I-J} \tag{3.1}$$

para todo $\mathbf{y} \in \mathbb{F}_q^N$.

Afirmamos agora que

$$a_P(\mathbf{y}) = \mathbf{y}_{I-J}$$

para todo $\mathbf{y} \in C_{J^c}$ (note que $\mathbf{y} \in C_{J^c}$ se, e só se, $\mathbf{y}_J = 0$). Definindo $supp(\mathbf{x}) = \{i : x_i \neq 0\}$ como sendo o *suporte* de $\mathbf{x} = (x_1, \ldots, x_N)$, começamos observando que

$$d_P(\mathbf{y}, \mathbf{c}) = \begin{cases} |\langle supp(\mathbf{y}_{I^c}) \cup supp(\mathbf{y}_{I-J} - \mathbf{c})\rangle| & \text{se} \quad \mathbf{c} \neq \mathbf{y}_{I-J} \\ |\langle supp(\mathbf{y}_{I^c})\rangle| & \text{se} \quad \mathbf{c} = \mathbf{y}_{I-J} \end{cases}.$$

Isto mostra que $d_P(\mathbf{y}, \mathbf{c})$ depende somente do suporte $supp(\mathbf{y}_{I-J} - \mathbf{c})$, donde segue que

$$d_P\left(\mathbf{y}, C_{(I,J)}\right) = d_P\left(\mathbf{y}_{I-J}, C_{(I,J)}\right).$$

Como $d_P\left(\mathbf{y}_{I-J}, C_{(I,J)}\right) = \{\mathbf{y}_{I-J}\}$, temos então que $d_P\left(\mathbf{y}, C_{(I,J)}\right) = \{\mathbf{y}_{I-J}\}$. Consequentemente, se a_P é um decodificador P-NN de $C_{(I,J)}$ e $\mathbf{y}$ é tal que $\mathbf{y}_J = \mathbf{0}$, então

$$a_P(\mathbf{y}) = \mathbf{y}_{I-J}. \tag{3.2}$$

Segue de (3.1) e (3.2) que

$$a_P(\mathbf{y}) = a_H(\mathbf{y})$$

para todo $\mathbf{y} \in C_{J^c}$.

Agora vamos usar o "*buraco* J" de $C_{(I,J)}$ para mostrar que $d_P\left(\mathbf{y}, C_{(I,J)}\right)$ é não trivial se $\mathbf{y}_J \neq 0$. Começamos observando que

$$d_P(\mathbf{y}, \mathbf{c}) \geq |\langle supp(\mathbf{y}_J)\rangle|$$

para todo $\mathbf{c} \in C_{(I,J)}$. Note agora, pondo $J' = supp(\mathbf{y}_J)$, que

$$d_P(\mathbf{y}, \mathbf{c}) = |\langle J'\rangle|$$

para todo $\mathbf{c} = \mathbf{z} + \mathbf{y}_{I-J}$ com $\mathbf{z} \in C_{I_{J'}^-}$. Consequentemente,

$$d_P\left(\mathbf{y}, C_{(I,J)}\right) = \left\{\mathbf{c} \in C_{(I,J)} : \mathbf{c} = \mathbf{z} + \mathbf{y}_{I-J}, \mathbf{z} \in C_{I_{J'}^-}\right\}.$$

Para cada escolha em $d_P\left(\mathbf{y}, C_{(I,J)}\right)$ temos um possível decodificador P-NN definido em $\mathbf{y}$. Como $\mathbf{y}_{I-J} \in d_P\left(\mathbf{y}, C_{(I,J)}\right)$ para todo $\mathbf{y} \in \mathbb{F}_q^N$, segue que o decodificador H-NN é também um decodificador P-NN e neste caso a diferença das perdas esperadas é igual a zero, exatamente como no Teorema 3.1. O fato de $d_P\left(\mathbf{y}, C_{(I,J)}\right)$ ser não trivial se $\mathbf{y}_J \neq 0$ gera a possibilidade de mudarmos esse cenário.

Seja

$$\widetilde{\mathbf{y}} = \mathbf{x}_{I^c} + \mathbf{e}_J$$

com $\mathbf{x}_{I^c} \in C_{I^c}$ e $\mathbf{e}_J := \sum_{j\in J} \mathbf{e}_j$. Por construção temos que

$$d_P(\widetilde{\mathbf{y}}, \mathbf{c}) > d_P(\widetilde{\mathbf{y}}, \mathbf{c}')$$

para todo $\mathbf{c} \in C_{(I,J)}$ com $\mathbf{c}_{I_J^+} \neq \mathbf{0}$ e para todo $\mathbf{c}' \in C_{I_J^-}$, já que $d_P(\widetilde{\mathbf{y}}, \mathbf{c}') = d_P(\widetilde{\mathbf{y}}, \mathbf{0})$ e

$$\langle supp(\widetilde{\mathbf{y}})\rangle \subsetneq \left\langle supp(\widetilde{\mathbf{y}}) \cup supp\left(\mathbf{c}_{I_J^+}\right)\right\rangle \subseteq \langle supp(\widetilde{\mathbf{y}} - \mathbf{c})\rangle.$$

Como $d_P(\widetilde{\mathbf{y}}, \mathbf{c}')$ não depende de $\mathbf{c}'$, segue que $d_P\left(\widetilde{\mathbf{y}}, C_{(I,J)}\right) = C_{I_J^-}$. Fixando

$$\mathbf{0} \neq \widetilde{\mathbf{c}} \in C_{I_J^-}$$

e definindo

$$a_P(\mathbf{y}) = \begin{cases} \mathbf{y}_{I-J} & \text{se} \quad \mathbf{y} \neq \widetilde{\mathbf{y}} \\ \widetilde{\mathbf{c}} & \text{se} \quad \mathbf{y} = \widetilde{\mathbf{y}} \end{cases}, \tag{3.3}$$

teremos, por construção, que a_P é um decodificador P-NN distinto do decodificador H-NN a_H: $a_P(\widetilde{\mathbf{y}}) \neq \mathbf{0}$.

Tendo os decodificadores a_H e a_P em mãos, nos resta determinar duas palavras-código τ_1 e τ_2 que satisfaçam a condição do Teorema 2.9. Sejam

$$\tau_1 = \sum_{i \in I_J^+} \mathbf{e}_i \in C_{(I,J)}$$

e

$$\tau_2 = \widetilde{\mathbf{c}}.$$

Como $a_H(\widetilde{\mathbf{y}}) = \mathbf{0}$ e $a_P(\widetilde{\mathbf{y}}) = \widetilde{\mathbf{c}}$, temos que

$$n_1 := d_H(\widetilde{\mathbf{y}}, a_H(\widetilde{\mathbf{y}}) - \tau_1) = d_H(\widetilde{\mathbf{y}}, \mathbf{0} - \tau_1) = w_H(\widetilde{\mathbf{y}}) + \left|I_J^+\right|,$$

$$m_1 := d_H(\widetilde{\mathbf{y}}, a_P(\widetilde{\mathbf{y}}) - \tau_1) = d_H(\widetilde{\mathbf{y}}, \widetilde{\mathbf{c}} - \tau_1) = w_H(\widetilde{\mathbf{y}}) + w_H(\widetilde{\mathbf{c}}) + \left|I_J^+\right|,$$

$$n_2 := d_H(\widetilde{\mathbf{y}}, a_H(\widetilde{\mathbf{y}}) - \tau_2) = d_H(\widetilde{\mathbf{y}}, \mathbf{0} - \tau_2) = w_H(\widetilde{\mathbf{y}}) + w_H(\widetilde{\mathbf{c}})$$

e

$$m_2 := d_H(\widetilde{\mathbf{y}}, a_P(\widetilde{\mathbf{y}}) - \tau_2) = d_H(\widetilde{\mathbf{y}}, \widetilde{\mathbf{c}} - \tau_2) = w_H(\widetilde{\mathbf{y}})$$

e, consequentemente, $n_1 < m_1$ e $n_2 > m_2$. Como $a_H(\mathbf{y}) = a_P(\mathbf{y})$ para todo $\mathbf{y} \neq \widetilde{\mathbf{y}}$, segue de (2.17) que

$$T_{(a_H,a_P)}(\tau_1) = \frac{(1-p)^N}{M}\left(s^{n_1} - s^{m_1}\right)$$

e

$$T_{(a_H,a_P)}(\tau_2) = \frac{(1-p)^N}{M}\left(s^{n_2} - s^{m_2}\right).$$

O resultado segue agora do fato de que

$$T_{(a_H,a_P)}(\tau_2) < 0 < T_{(a_H,a_P)}(\tau_1)$$

para todo $0 < s < 1$. □

O cálculo de $T_{(a_H,a_P)}(\tau)$, em geral, é demasiadamente longo e complexo. Para um código BGL temos uma expressão explícita:

Corolário 3.1 *Seja $P = ([N], \leq)$ uma ordem (I, J)-decomponível. Para o código BGL $C_{(I,J)}$ e para os decodificadores P-NN dados em (3.3) temos que*

$$T_{(a_H,a_P)}(\tau) = \frac{(1-p)^N}{M}\left(s^{d_H(\widetilde{\mathbf{y}},-\tau)} - s^{d_H(\widetilde{\mathbf{y}},\widetilde{\mathbf{c}}-\tau)}\right)$$

para todo $\tau \in C_{(I,J)}$.

Destacamos agora uma importante família de posets que são (I, J)-decomponíveis. Se $P = ([N], \leq)$ é uma união disjunta de n cadeias de comprimento m, neste caso $N = n \cdot m$, diremos que P é uma (n, m)-*ordem de Niederreiter-Rosenbloom-Tsfasman* ou, simplesmente, uma (n, m) *NRT-ordem*. Se $m = 1$, então a NRT-ordem coincide com a ordem de Hamming formada por n elementos. Se $n = 1$, então a NRT-ordem se reduz a uma cadeia de comprimento m. As métricas poset induzidas pelas ordens de Niederreiter-Rosenbloom-Tsfasman foram introduzidas independentemente por Niederreiter em 1987 (ver [29]) e por Rosenbloom e Tsfasman em 1997 (ver [37]). A relevância das métricas poset NRT é justificada pela sua aplicação no estudo de certos canais de desvanecimento (ver [41]).

Note agora que toda NRT-ordem com $m \geq 4$ é (I, J)-decomponível. De fato, supondo que $1 < 2 < \ldots < m$ é uma das cadeias de P e tomando $J = \{m-1\}$, teremos que $I = \{m-2, m-1, m\}$ é um filtro J-decomponível de P: I é um filtro próprio de P com $I_J^- = \{m-2\}$ e $I_J^+ = \{m\}$ ambos não vazios. Consequentemente:

Corolário 3.2 *Se P é uma (n, m) NRT-ordem com $m \geq 4$ e H é a ordem de Hamming sobre $[nm]$, então existe um $[nm; k]_q$ código linear C e $\tau, \tau' \in C$ tais que*

$$T_{(a_H,a_P)}(\tau) \cdot T_{(a_H,a_P)}(\tau') < 0$$

para algum par a_H e a_P de decodificadores H-NN e P-NN de C, respectivamente. Consequentemente, $\mathcal{V}^+(a_H, a_P)$ e $\mathcal{V}^-(a_H, a_P)$ são ambos não vazios.

Em particular, temos que o filtro $I = \{2, 3, \ldots, m\}$ da cadeia $1 < 2 < \ldots < m$, com $m \geq 4$, é J-decomponível para todo $J = \{3, , \ldots, m-k+1\}$ com $2 \leq k \leq m-2$. Considerando os códigos BGL $C_{(I,J)}$ associados a cada um dos $m-3$ possíveis conjuntos J, teremos que:

Corolário 3.3 *Se P é uma $(1, m)$ NRT-ordem com $m \geq 4$ e H é a ordem de Hamming sobre $[m]$, então para todo $2 \leq k \leq m-2$ existe um $[m;k]_q$ código linear C e decodificadores a_H e a_P H-NN e P-NN de C, respectivamente, tais que $\mathcal{V}^+(a_H, a_P)$ e $\mathcal{V}^-(a_H, a_P)$ são ambos não vazios.*

Na demonstração do Teorema 3.2 mostramos que se $a : \mathbb{F}_q^N \to C_{(I,J)}$ é um decodificador H-NN para o código BGL $C_{(I,J)}$, então $a(\mathbf{y}) = \mathbf{y}_{I-J}$ para todo $\mathbf{y} \in \mathbb{F}_q^N$. Na sequência, verificamos que a também era um decodificador P-NN para $C_{(I,J)}$. Desta forma, se P e Q são duas ordens (I, J)-decomponíveis, podemos repetir os argumentos da prova do Teorema 3.2, mantendo a definição de a_P e pondo $a_Q := a_H$, e concluir que:

Teorema 3.3 *Se P e Q são duas ordens (I, J)-decomponíveis sobre $[N]$, então existe um $[N;k]_q$ código linear C e decodificadores P-NN e Q-NN tais que $\mathcal{V}^+(a_P, a_Q)$ e $\mathcal{V}^-(a_P, a_Q)$ são ambos não vazios.*

3.4 Raio de Empacotamento

Para a maioria dos exemplos o cálculo exato da perda esperada é demasiadamente longo e complexo. Neste caso podemos considerar o raio de empacotamento como parâmetro para medir o desempenho de um decodificador NN, mesmo sabendo que este parâmetro não é totalmente confiável: mesmo que $\mathbf{y}$ não pertença a nenhuma das bolas $B_P(\mathbf{c}; R_P(C))$ com $\mathbf{c} \in C$, ainda é possível que exista uma única palavra-código $\mathbf{c} \in C$ tal que $d_P(\mathbf{y}, \mathbf{c}) \leq d_P(\mathbf{y}, \mathbf{c}')$ para todo $\mathbf{c}' \in C$.

No caso da métrica de Hamming o raio de empacotamento é dado em função da distância mínima do código (veja a Seção 2.2):

$$R_H(C) = \left\lfloor \frac{d_H(C) - 1}{2} \right\rfloor.$$

O raio de empacotamento também é dado em função distância mínima do código se P é uma ordem cadeia (ver [33], Theorem 5):

$$R_P(C) = d_P(C) - 1. \tag{3.4}$$

Neste caso vale também a recíproca: se $R_P(C) = d_P(C) - 1$ para todo código C, então P é uma cadeia (ver [34]).

Nem sempre o raio de empacotamento de um código é dado em função da sua distância mínima. Antes de exemplificarmos este fato estabeleceremos um resultado auxiliar.

Teorema 3.4 *Seja C um código linear e para cada $\mathbf{c} \in C$ defina*

$$R_{\mathbf{c}} = \max\{r : B_P(\mathbf{0}; r) \cap B_P(\mathbf{c}; r) = \varnothing\}.$$

Nestas condições $R_P(C) = \min\{R_{\mathbf{c}} : \mathbf{c} \in C\}$.

Demonstração Como a P-métrica é invariante por translações,

$$B_P(\mathbf{c}; r) \cap B_P(\mathbf{c}'; r) = \varnothing \Leftrightarrow B_P(\mathbf{0}; r) \cap B_P(\mathbf{c} - \mathbf{c}'; r) = \varnothing.$$

Isto assegura que

$$\max\{r : B_P(\mathbf{c}; r) \cap B_P(\mathbf{c}'; r) = \varnothing\} = R_{\mathbf{c}-\mathbf{c}'}.$$

Agora, como C é linear, $\mathbf{c} - \mathbf{c}' \in C$ para todo par $\mathbf{c}, \mathbf{c}' \in C$. Isto mostra que $R_{\mathbf{c}}$ é suficiente para o cálculo do raio de empacotamento $R_P(C)$. □

Exemplo 3.2 *Neste exemplo mostraremos que nem sempre o raio de empacotamento é determinado pelos vetores de peso mínimo. Seja P a $(6, 4)$ NRT-ordem dada pelas relações*

$$1 < 2 < 3 < 4$$
$$5 < 6 < 7 < 8$$
$$\vdots$$
$$21 < 22 < 23 < 24.$$

e C o $[24; 2]_2$ código binário gerado pelos vetores

$$\mathbf{c}_1 = \{1, 2, 3, 4, 5\} \text{ e } \mathbf{c}_2 = \{9, 10, 11, 13, 17, 21\}$$

(*estamos identificando os vetores de $\mathbb{F}_2^{24}$ com os seus suportes*).

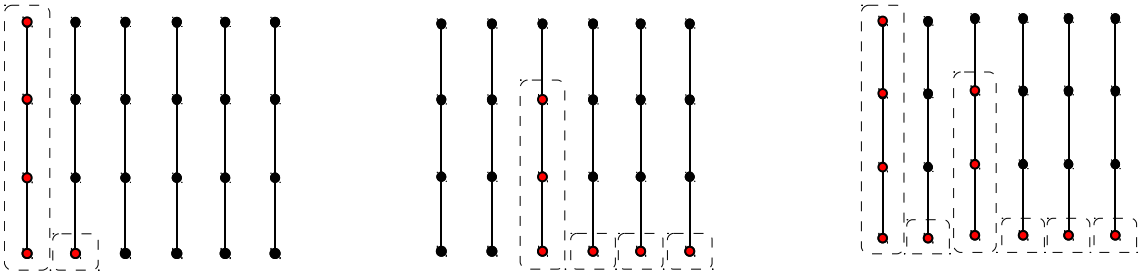

Figura 3.4: Representações dos suportes de $\mathbf{c}_1$, $\mathbf{c}_2$ e $\mathbf{c}_1 + \mathbf{c}_2$, respecticamente.

Para determinarmos o raio de empacotamento $R_P(C)$ de C, começamos calculando os raios $R_{\mathbf{c}_1}$, $R_{\mathbf{c}_2}$ e $R_{\mathbf{c}_1+\mathbf{c}_2}$:

Cálculo de $R_{\mathbf{c}_1}$: *Se $\mathbf{y} \in B_P(\mathbf{0}; 3)$, então $4 \notin supp(\mathbf{y})$. Isto implica que $d_P(\mathbf{y}, \mathbf{c}_1) \geq 4$, ou seja, $\mathbf{y} \notin B_P(\mathbf{c}_1; 3)$. Consequentemente, $B_P(\mathbf{0}; 3) \cap B_P(\mathbf{c}_1; 3) = \varnothing$. Não podemos aumentar o raio: $\{4\} \in B_P(\mathbf{0}; 4) \cap B_P(\mathbf{c}_1; 4)$. Logo,*

$$R_{\mathbf{c}_1} = 3.$$

Cálculo de $R_{\mathbf{c}_2}$: *Se $\mathbf{y} \in B_P(\mathbf{0}; 2)$, então $11 \notin supp(\mathbf{y})$ e daí que $d_P(\mathbf{y}, \mathbf{c}_2) \geq 3$. Isto mostra que $B_P(\mathbf{0}; 2) \cap B_P(\mathbf{c}_2; 2) = \varnothing$. Como $\{9, 10, 11\} \in B_P(\mathbf{0}; 3) \cap B_P(\mathbf{c}_2; 3)$, concluímos que*

$$R_{\mathbf{c}_2} = 2.$$

Cálculo de $R_{\mathbf{c}_1+\mathbf{c}_2}$: *Como $|supp(\mathbf{c}_1 + \mathbf{c}_2)| = 11$, se $\mathbf{y}$ é tal que $|supp(\mathbf{y})| \leq 5$, então $|supp(\mathbf{y} - (\mathbf{c}_1 + \mathbf{c}_2))| \geq 6$. Consequentemente, como*

$$d_P(\mathbf{y}, \mathbf{c}_1 + \mathbf{c}_2) \geq |supp(\mathbf{y} - (\mathbf{c}_1 + \mathbf{c}_2))|,$$

$d_P(\mathbf{y}, \mathbf{c}_1 + \mathbf{c}_2) \geq 6$. Estes fatos asseguram que $B_P(\mathbf{0}; 5) \cap B_P(\mathbf{c}_1 + \mathbf{c}_2; 5) = \varnothing$. Não podemos aumentar o raio: $\mathbf{c}_1 \in B_P(\mathbf{0}; 6) \cap B_P(\mathbf{c}_1 + \mathbf{c}_2; 6)$. Logo,

$$R_{\mathbf{c}_1+\mathbf{c}_2} = 5.$$

O resultado segue agora do Teorema 3.4:

$$R_P(C) = R_{\mathbf{c}_2} = 2.$$

Note agora que $R_P(C)$ não é dado em função do vetor de peso mínimo.

A descrição geral do raio de empacotamento em espaços NRT é dada em [34]. Agora vamos descrever o raio de empacotamento dos códigos coordenados C_X em um espaço poset arbitrário.

Teorema 3.5 *Se $C_X \subseteq \mathbb{F}_q^N$ é o espaço coordenado com suporte em X, então*

$$R_P(C_X) = d_P(C_X) - 1.$$

Demonstração Seja $R+1 := d_P(C_X)$. Fixe $\mathbf{0} \neq \mathbf{c} \in C_X$ e seja $\mathbf{y} \in \mathbb{F}_q^N$. Podemos escrever

$$\mathbf{y} = \mathbf{y}_{X^c} + \mathbf{y}_X$$

com $\mathbf{y}_{X^c} \in C_{X^c}$ e $\mathbf{y}_X \in C_X$. Se $\mathbf{y}_X \neq \mathbf{0}$, então $w_P(\mathbf{y}_X) \geq R+1$. Como $w_P(\mathbf{y}) \geq w_P(\mathbf{y}_X)$, concluímos que $w_P(\mathbf{y}) \geq R+1$, ou seja, $\mathbf{y} \notin B_P(\mathbf{0}; R)$. Agora, se $\mathbf{y}_X = \mathbf{0}$, então $d_P(\mathbf{y}, \mathbf{c}) \geq w_P(\mathbf{c})$, e daí segue que $d_P(\mathbf{y}, \mathbf{c}) \geq R+1$, ou seja, $\mathbf{y} \notin B_P(\mathbf{c}; R)$. Em resumo,

$$B_P(\mathbf{0}; R) \cap B_P(\mathbf{c}; R) = \varnothing$$

para todo $\mathbf{0} \neq \mathbf{c} \in C_X$. Como

$$B_P(\mathbf{0}; R+1) \cap B_P(\mathbf{c}; R+1) \neq \varnothing$$

para todo $\mathbf{c} \in \mathbf{C}$ com $w_P(\mathbf{c}) = d_P(C_X)$, concluímos que

$$R_P(C_X) = R = d_P(C_X) - 1.$$

□

Como consequência do Teorema 3.5:

Corolário 3.4 *Seja $C_{(I,J)}$ um código BGL. Então*

$$R_P\left(C_{(I,J)}\right) = d_P\left(C_{(I,J)}\right) - 1.$$

Capítulo 4

Hello World

Na introdução deste trabalho simulamos a transmissão da imagem "Hello World" (Figura 1) sobre um canal binário simétrico decodificando a imagem recebida (*imagem original + erros*) de duas maneiras diferentes, uma usando um decodificador ML e outra usando um decodificador P-NN (Figura 3) com P dado pelas relações $1 < 2 < \ldots < 7$. Neste capítulo descreveremos com detalhes o processo de codificação e os respectivos decodificadores H-NN e P-NN considerados na simulação.

4.1 O Problema

A imagem "Hello World" pertence a uma classe maior de imagens: a classe das imagens em escala de cinza, com 16 possíveis tons de cinza para cada pixel (ilustrados na Figura 4.1), tais que os erros de decodificação são avaliados pelas diferenças entre os tons recebidos e os tons enviados, ou seja, quanto maior é a diferença entre os tons, maior é o valor do erro de decodificação. Vamos considerar a seguinte escala de valores para os tons de cinza, a nossa função de valor ν relativa aos tons de cinza:

RGB	**Valor Semântico**
$(101, 101, 101)$	1.00
$(102, 102, 102)$	0.90
$(103, 103, 103)$	0.89
$(104, 104, 104)$	0.88
$(105, 105, 105)$	0.87
$(106, 106, 106)$	0.86
$(107, 107, 107)$	0.85
$(108, 108, 108)$	0.84

RGB	**Valor Semântico**
$(109, 109, 109)$	0.83
$(110, 110, 110)$	0.82
$(187, 187, 187)$	0.50
$(188, 188, 188)$	0.40
$(189, 189, 189)$	0.30
$(190, 190, 190)$	0.20
$(191, 191, 191)$	0.10
$(192, 192, 192)$	0.00

Tabela 4.1: A função de valor ν. Valores semânticos dos tons de cinza.

Para transmitir as imagens em escala de cinza por um canal binário simétrico, vamos codificar os tons de cinza dados acima com o $[7; 4]_2$ código binário de Hamming $\mathcal{H}(3)$ dado no Exemplo 2.1. Considerando o poset P dado pelas relações

$$1 < 2 < \ldots < 7,$$

nosso problema consiste em determinar um codificador de canal e um decodificador P-NN que gerem resultados melhores do que os obtidos pelo decodificador ML.

4.2 Os Decodificadores

No nosso problema estamos considerando o $[7; 4]_2$ código binário de Hamming $\mathcal{H}(3)$ dado pela matriz de verificação de paridade

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

Vamos considerar o seguinte rótulo para $\mathcal{H}(3)$:

Palavra-Código $\mathbf{c}_i$	**Palavra-Código $\mathbf{c}_i$**
$\mathbf{c}_{15} = 1111111$	$\mathbf{c}_7 = 0110110$
$\mathbf{c}_{14} = 0001111$	$\mathbf{c}_6 = 1000110$
$\mathbf{c}_{13} = 0010011$	$\mathbf{c}_5 = 1011010$
$\mathbf{c}_{12} = 1100011$	$\mathbf{c}_4 = 0101010$
$\mathbf{c}_{11} = 1010101$	$\mathbf{c}_3 = 0011100$
$\mathbf{c}_{10} = 0100101$	$\mathbf{c}_2 = 1101100$
$\mathbf{c}_9 = 0111001$	$\mathbf{c}_1 = 1110000$
$\mathbf{c}_8 = 1001001$	$\mathbf{c}_0 = 0000000$

Tabela 4.2: Palavras-código de $\mathcal{H}(3)$.

Os 16 tons de cinza representados por $\mathcal{H}(3)$ estão ilustrados Figura 4.1 abaixo (ainda não estamos preocupados com o codificador de canal):

Figura 4.1: Tons de cinza e respectivos valores de RGB.

No Exemplo 2.1 mostramos que $\mathcal{H}(3)$ admite um único decodificador ML, sendo este dado por

$$a(\mathbf{y}) = \begin{cases} \mathbf{y} & \text{se} \quad H\mathbf{y} = \mathbf{0} \\ \mathbf{y} - \mathbf{e}_i & \text{se} \quad H\mathbf{y} = i\text{-ésimo coluna de } H \end{cases}.$$

O mesmo não acontece com os decodificadores P-NN: se $\mathbf{y} = 1000111$ e $\mathbf{c}_1, \mathbf{c}_2 \in \mathcal{H}(3)$ são tais que $\mathbf{c}_1 = 1111111$ e $\mathbf{c}_2 = 0001111$, então

$$d_P(\mathbf{y}, \mathbf{c}_1) = 4 = d_P(\mathbf{y}, \mathbf{c}_2);$$

como $d_P(\mathbf{y}, \mathbf{c}) \geq 5$ para todo $\mathbf{c} \in \mathcal{H}(3)$ distinto de $\mathbf{c}_1, \mathbf{c}_2$, concluímos que $\mathbf{c}_1, \mathbf{c}_2 \in d_P(\mathbf{y}, \mathcal{H}(3))$, ou seja, temos mais de uma possibilidade para definir $a_P(\mathbf{y})$.

Para o nosso problema vamos considerar o decodificador a_P dado pelas seguintes regiões de decisão:

$a_P^{-1}(\mathbf{c}_{15})$	$a_P^{-1}(\mathbf{c}_{14})$	$a_P^{-1}(\mathbf{c}_{13})$	$a_P^{-1}(\mathbf{c}_{12})$
1111111	0001111	0010011	1100011
0111111	1001111	1010011	0100011
0011111	1101111	0110011	1000011
0010111	0101111	1110011	0000011
1011111	1010111	1111011	0001011
1110111	1100111	0111011	1001011
0000111	1000111	1011011	0101011
0110111	0100111	1101011	0011011

$a_P^{-1}(\mathbf{c}_{11})$	$a_P^{-1}(\mathbf{c}_{10})$	$a_P^{-1}(\mathbf{c}_9)$	$a_P^{-1}(\mathbf{c}_8)$
1010101	0100101	0111001	1001001
0010101	1100101	1111001	0001001
1110101	0000101	0011001	1101001
0110101	1000101	1011001	0101001
1111101	0011101	1000001	1010001
0111101	0001101	0100001	0110001
1011101	1001101	0010001	1110001
1101101	0101101	1100001	0000001

$a_P^{-1}(\mathbf{c}_7)$	$a_P^{-1}(\mathbf{c}_6)$	$a_P^{-1}(\mathbf{c}_5)$	$a_P^{-1}(\mathbf{c}_4)$
0110110	1000110	1011010	0101010
1110110	0000110	0011010	1101010
0010110	1100110	1111010	0001010
1010110	0100110	0111010	1001010
1111110	0001110	0000010	1100010
0111110	1001110	1000010	1010010
1011110	0011110	0100010	0110010
1101110	0101110	0010010	1110010

$a_P^{-1}(\mathbf{c}_3)$	$a_P^{-1}(\mathbf{c}_2)$	$a_P^{-1}(\mathbf{c}_1)$	$a_P^{-1}(\mathbf{c}_0)$
0011100	1101100	1110000	0000000
1011100	0101100	0110000	1000000
0111100	1001100	1010000	0100000
1111100	0001100	0010000	1100000
1000100	1010100	0001000	1011000
0100100	0000100	1111000	0111000
0010100	1110100	0011000	1001000
1100100	0110100	1101000	0101000

4.3 Os Codificadores Heurísticos

Agora, vamos descrever heuristicamente um possível candidato a codificador de Bayes relativo ao decodificador a_P e a função de valor ν (ver Definição 2.9). Começamos observando que a_P é menos suscetível aos erros confinados nas primeiras coordenadas: palavras contendo erros nas t primeiras coordenadas são decodificadas como sendo palavras-código contendo pelos menos as $N - t$ últimas coordenadas corretas; em particular, todos os erros confinados nas duas primeiras coordenadas são corrigidos. A última observação segue do fato de que $R_P(\mathcal{H}(3)) = 2$ (veja (3.4)). Como no nosso problema os tons de cinza mais escuros representam erros de decodificação mais severos, associamos estes às palavras-código que possuem entradas não nulas nas últimas coordenadas (7 e 6), os tons menos escuros associamos

as palavras-código que possuem entradas não nulas nas posições intermediárias (5, 4 e 3) e os tons mais claros as demais palavras-código. O codificador de canal dado na Figura 2.1 satisfaz estas condições. Como veremos, este é um bom codificador de canal.

4.4 As Perdas Esperadas e os Codificadores de Bayes

Como $\mathcal{H}(3)$ é um código binário perfeito e $0, 3, 4, 7$ são os possíveis pesos de $\mathcal{H}(3)$, segue do Teorema 2.11 que

$$G_{a_H}(\mathbf{c}_1) = G_{a_H}(\mathbf{c}_3) = G_{a_H}(\mathbf{c}_4) = G_{a_H}(\mathbf{c}_6) = G_{a_H}(\mathbf{c}_8) = G_{a_H}(\mathbf{c}_{10}) = G_{a_H}(\mathbf{c}_{13})$$

e

$$G_{a_H}(\mathbf{c}_2) = G_{a_H}(\mathbf{c}_5) = G_{a_H}(\mathbf{c}_7) = G_{a_H}(\mathbf{c}_9) = G_{a_H}(\mathbf{c}_{11}) = G_{a_H}(\mathbf{c}_{12}) = G_{a_H}(\mathbf{c}_{14}).$$

Consequentemente,

$$\mathbb{E}_{\mathcal{H}(3)}(f, a_P, \nu) = G_{a_H}(\mathbf{c}_0)\,\nu_f(\mathbf{c}_0) +$$
$$G_{a_H}(\mathbf{c}_1)\,(\nu_f(\mathbf{c}_1) + \nu_f(\mathbf{c}_3) + \nu_f(\mathbf{c}_4) + \nu_f(\mathbf{c}_6) + \nu_f(\mathbf{c}_8) + \nu_f(\mathbf{c}_{10}) + \nu_f(\mathbf{c}_{13})) +$$
$$G_{a_H}(\mathbf{c}_2)\,(\nu_f(\mathbf{c}_2) + \nu_f(\mathbf{c}_5) + \nu_f(\mathbf{c}_7) + \nu_f(\mathbf{c}_9) + \nu_f(\mathbf{c}_{11}) + \nu_f(\mathbf{c}_{12}) + \nu_f(\mathbf{c}_{14})) +$$
$$G_{a_H}(\mathbf{c}_{15})\,\nu_f(\mathbf{c}_{15})$$

com

$$\begin{aligned} G_{a_H}(\mathbf{c}_0) &= z(s)(16 + 112s), \\ G_{a_H}(\mathbf{c}_1) &= z(s)(48s^2 + 16s^3 + 64s^4), \\ G_{a_H}(\mathbf{c}_2) &= z(s)(64s^3 + 16s^4 + 48s^5), \\ G_{a_H}(\mathbf{c}_{15}) &= z(s)(112s^6 + 16s^7), \end{aligned}$$

onde

$$z(s) := \frac{1}{16\,(1+s)^7}.$$

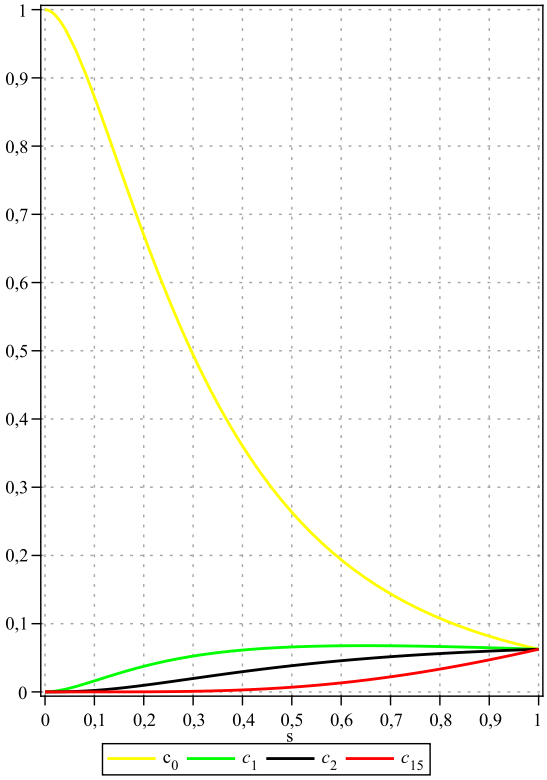

Figura 4.2: Gráficos das funções $G_{a_H}(\mathbf{c}_0)$, $G_{a_H}(\mathbf{c}_1)$, $G_{a_H}(\mathbf{c}_2)$ e $G_{a_H}(\mathbf{c}_{15})$ em função de s.

Como

$$G_{a_H}(\mathbf{c}_0) > G_{a_H}(\mathbf{c}_1) > G_{a_H}(\mathbf{c}_2) > G_{a_H}(\mathbf{c}_{15})$$

para todo $0 < s < 1$, pondo

$$C_1 = \{\mathbf{c}_1, \mathbf{c}_3, \mathbf{c}_4, \mathbf{c}_6, \mathbf{c}_8, \mathbf{c}_{10}, \mathbf{c}_{13}\}$$

e

$$C_2 = \{\mathbf{c}_2, \mathbf{c}_5, \mathbf{c}_7, \mathbf{c}_9, \mathbf{c}_{11}, \mathbf{c}_{12}, \mathbf{c}_{14}\},$$

segue do Teorema 2.3 que f é um codificador de Bayes relativo ao decodificador a_H e a função de valor ν se, e somente se,

$$\nu_f(\mathbf{c}_0) < \nu_f(\mathbf{c}') < min\{\nu_f(\mathbf{c}) : \mathbf{c} \in C_2\}$$

para todo $\mathbf{c}' \in C_1$ e

$$max\{\nu_f(\mathbf{c}) : \mathbf{c} \in C_1\} < \nu_f(\mathbf{c}') < \nu_f(\mathbf{c}_{15})$$

para todo $\mathbf{c}' \in C_2$.

As funções $G_{a_P}(\mathbf{c}_i)$ com $0 < i < 15$ são dados por:

$$\begin{aligned}
G_{a_P}(\mathbf{c}_0) &= z(s)(16 + 39s + 39s^2 + 25s^3 + 9s^4),\\
G_{a_P}(\mathbf{c}_1) &= z(s)(25s + 57s^2 + 39s^3 + 7s^4),\\
G_{a_P}(\mathbf{c}_2) &= z(s)(10s + 42s^2 + 54s^3 + 22s^4),\\
G_{a_P}(\mathbf{c}_3) &= z(s)(6s + 22s^2 + 42s^3 + 42s^4 + 16s^5),\\
G_{a_P}(\mathbf{c}_4) &= z(s)(7s + 39s^2 + 57s^3 + 25s^4),\\
G_{a_P}(\mathbf{c}_5) &= z(s)(9s + 25s^2 + 39s^3 + 39s^4 + 16s^5),\\
G_{a_P}(\mathbf{c}_6) &= z(s)(16s^2 + 42s^3 + 42s^4 + 22s^5 + 6s^6),\\
G_{a_P}(\mathbf{c}_7) &= z(s)(22s^3 + 54s^4 + 42s^5 + 10s^6),\\
G_{a_P}(\mathbf{c}_8) &= G_{a_P}(\mathbf{c}_2),\\
G_{a_P}(\mathbf{c}_9) &= G_{a_P}(\mathbf{c}_3),\\
G_{a_P}(\mathbf{c}_{10}) &= z(s)(16s^2 + 39s^3 + 39s^4 + 25s^5 + 9s^6),\\
G_{a_P}(\mathbf{c}_{11}) &= z(s)(25s^3 + 57s^4 + 39s^5 + 7s^6),\\
G_{a_P}(\mathbf{c}_{12}) &= G_{a_P}(\mathbf{c}_6),\\
G_{a_P}(\mathbf{c}_{13}) &= G_{a_P}(\mathbf{c}_7),\\
G_{a_P}(\mathbf{c}_{14}) &= z(s)(7s^3 + 39s^4 + 57s^5 + 25s^6),\\
G_{a_P}(\mathbf{c}_{15}) &= z(s)(9s^3 + 25s^4 + 39s^5 + 39s^6 + 16s^7).
\end{aligned}$$

É possível demonstrar que $G_{a_P}(\mathbf{c}_i) - G_{a_P}(\mathbf{c}_j) = 0$ para algum $0 < s < 1$ se, e somente se, $(i, j) = (4, 5)$ ou $(i, j) = (14, 15)$. Para ambos os casos $s = \frac{1}{8}$ é a única solução. Segue da Figura 4.3 abaixo que

$$G_{a_P}(\mathbf{c}_0) > G_{a_P}(\mathbf{c}_1) > G_{a_P}(\mathbf{c}_2) = G_{a_P}(\mathbf{c}_8) >$$

$$G_{a_P}(\mathbf{c}_5) > G_{a_P}(\mathbf{c}_4) > G_{a_P}(\mathbf{c}_9) = G_{a_P}(\mathbf{c}_3) >$$

$$G_{a_P}(\mathbf{c}_{12}) = G_{a_P}(\mathbf{c}_6) > G_{a_P}(\mathbf{c}_{10}) > G_{a_P}(\mathbf{c}_{11}) >$$
$$G_{a_P}(\mathbf{c}_{13}) = G_{a_P}(\mathbf{c}_7) > G_{a_P}(\mathbf{c}_{15}) > G_{a_P}(\mathbf{c}_{14})$$

se $0 < s < \frac{1}{8}$,

$$G_{a_P}(\mathbf{c}_0) > G_{a_P}(\mathbf{c}_1) > G_{a_P}(\mathbf{c}_2) = G_{a_P}(\mathbf{c}_8) >$$
$$G_{a_P}(\mathbf{c}_4) = G_{a_P}(\mathbf{c}_5) > G_{a_P}(\mathbf{c}_9) = G_{a_P}(\mathbf{c}_3) >$$
$$G_{a_P}(\mathbf{c}_{12}) = G_{a_P}(\mathbf{c}_6) > G_{a_P}(\mathbf{c}_{10}) > G_{a_P}(\mathbf{c}_{11}) >$$
$$G_{a_P}(\mathbf{c}_{13}) = G_{a_P}(\mathbf{c}_7) > G_{a_P}(\mathbf{c}_{14}) = G_{a_P}(\mathbf{c}_{15})$$

se $s = \frac{1}{8}$ e

$$G_{a_P}(\mathbf{c}_0) > G_{a_P}(\mathbf{c}_1) > G_{a_P}(\mathbf{c}_2) = G_{a_P}(\mathbf{c}_8) >$$
$$G_{a_P}(\mathbf{c}_4) > G_{a_P}(\mathbf{c}_5) > G_{a_P}(\mathbf{c}_9) = G_{a_P}(\mathbf{c}_3) >$$
$$G_{a_P}(\mathbf{c}_{12}) = G_{a_P}(\mathbf{c}_6) > G_{a_P}(\mathbf{c}_{10}) > G_{a_P}(\mathbf{c}_{11}) >$$
$$G_{a_P}(\mathbf{c}_{13}) = G_{a_P}(\mathbf{c}_7) > G_{a_P}(\mathbf{c}_{14}) > G_{a_P}(\mathbf{c}_{15})$$

se $\frac{1}{8} < s < 1$. Considerando somente o último caso, concluímos do Teorema 2.3 que f é um codificador de Bayes relativo ao decodificador a_P para todo $\frac{1}{8} < s < 1$ se, e somente se,

$$\nu_f(\mathbf{c}_1) < \nu_f(\mathbf{c}_2), \nu_f(\mathbf{c}_8) < \nu_f(\mathbf{c}_4),$$
$$\nu_f(\mathbf{c}_5) < \nu_f(\mathbf{c}_3), \nu_f(\mathbf{c}_9) < min\{\nu_f(\mathbf{c}_6), \nu_f(\mathbf{c}_{12})\},$$
$$max\{\nu_f(\mathbf{c}_3), \nu_f(\mathbf{c}_9)\} < \nu_f(\mathbf{c}_6), \nu_f(\mathbf{c}_{12}) < \nu_f(\mathbf{c}_{10}),$$
$$\nu_f(\mathbf{c}_{11}) < \nu_f(\mathbf{c}_7), \nu_f(\mathbf{c}_{13}) < \nu_f(\mathbf{c}_{14}),$$

e

$$\nu_f(\mathbf{c}_0) < \nu_f(\mathbf{c}_1) < \nu_f(\mathbf{c}_4) < \nu_f(\mathbf{c}_5) < \nu_f(\mathbf{c}_{10}) < \nu_f(\mathbf{c}_{11}) < \nu_f(\mathbf{c}_{14}) < \nu_f(\mathbf{c}_{15}).$$

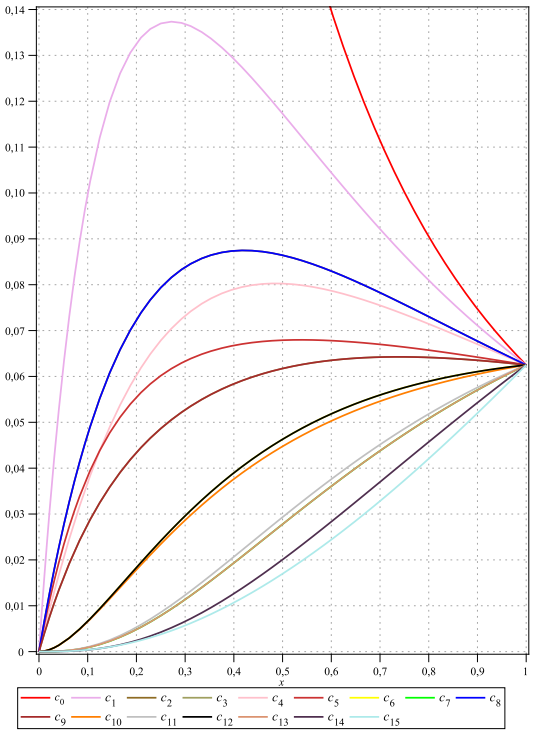

Figura 4.3: Polinômios $G_{a_P}(\mathbf{c}_i)$ com $1 < i < 15$.

4.5 Os Codificadores Heurísticos são Codificadores de Bayes?

O codificador de canal f_h dado na Figura 2.1 satisfaz as condições da heurística estabelecida na Seção 4.3. Note agora que f_h não é um codificador de Bayes relativo ao decodificador a_P e a função de valor ν: por exemplo, não é verdade que

$$\nu_{f_h}(\mathbf{c}_1) < \nu_{f_h}(\mathbf{c}_8) < \nu_{f_h}(\mathbf{c}_4).$$

Isto mostra que nem todo codificador heurístico é um codificador de Bayes para a_P. Também é verdade que f_h não é um codificador de Bayes para a_H e ν: por exemplo, não vale que

$$\nu_{f_h}(\mathbf{c}_0) < \nu_{f_h}(\mathbf{c}_3) < \nu_{f_h}(\mathbf{c}_2).$$

Relativo ao codificador de canal f_h,

$$\mathbb{E}_{\mathcal{H}(3)}\left(f_h, a_H, \nu\right) = z\left(s\right)\left(16s^7 + 112s^6 + 241.44s^5 + 349.92s^4 + 389.28s^3 + 202.08s^2\right)$$

e

$$\mathbb{E}_{\mathcal{H}(3)}\left(f_h, a_P, \nu\right) = z\left(s\right)\left(16s^7 + 102.73s^6 + 281.77s^5 + 408.89s^4 + \right.$$
$$\left. 330.49s^3 + 143.74s^2 + 27.10s\right).$$

Consequentemente,

$$\mathbb{E}_{\mathcal{H}(3)}\left(f_h, a_H, a_P, \nu\right) = z\left(s\right)\left(9.27s^6 + 58.79s^3 - 58.97s^4 - \right.$$
$$\left. 40.33s^5 + 58.34s^2 - 27.10s\right).$$

Podemos concluír da Figura 4.4 abaixo que para todo $s > 0.39$ (equivalentemente, para todo $p > 0.28$) o decodificador a_P é melhor do que o decodificador a_H, mesmo não sendo f_h um codificador de Bayes relativo a a_P.

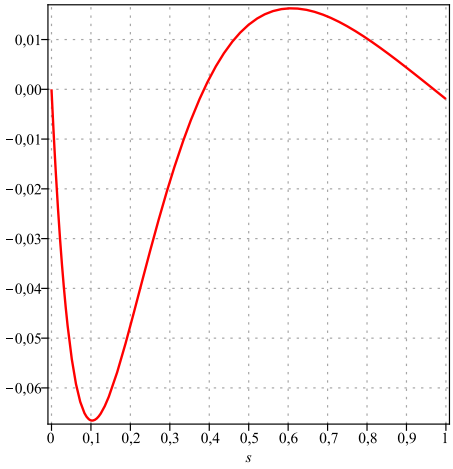

Figura 4.4: $\mathbb{E}_{\mathcal{H}(3)}\left(f_h, a_H, a_P, \nu\right)$ em função de s.

Considere agora o codificador de canal f_b dado pela Figura 4.5 abaixo:

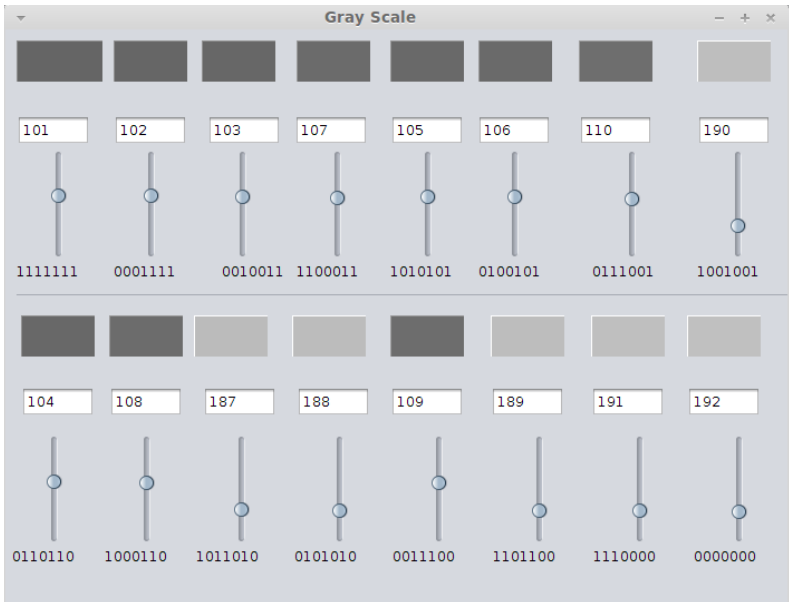

Figura 4.5: Codificador de canal f_b.

Para $\frac{1}{8} < s < 1$ temos que f_b satisfaz as condições para ser um codificador de Bayes relativo ao decodificador a_P:

$$\nu_f(\mathbf{c}_1) = 0.10 < \nu_f(\mathbf{c}_2) = 0.30, \nu_f(\mathbf{c}_8) = 0.20 < \nu_f(\mathbf{c}_4) = 0.40,$$

$$\nu_f(\mathbf{c}_5) = 0.50 < \nu_f(\mathbf{c}_3) = 0.83, \nu_f(\mathbf{c}_9) = 0.82 < min\{\nu_f(\mathbf{c}_6), \nu_f(\mathbf{c}_{12})\} = 0.84,$$

$$max\{\nu_f(\mathbf{c}_3), \nu_f(\mathbf{c}_9)\} = 0.83 < \nu_f(\mathbf{c}_6) = 0.84, \nu_f(\mathbf{c}_{12}) = 0.85 < \nu_f(\mathbf{c}_{10}) = 0.86,$$

$$\nu_f(\mathbf{c}_{11}) = 0.87 < \nu_f(\mathbf{c}_7) = 0.88, \nu_f(\mathbf{c}_{13}) = 0.89 < \nu_f(\mathbf{c}_{14}) = 0.90,$$

e

$$\nu_f(\mathbf{c}_0) = 0 < \nu_f(\mathbf{c}_1) = 0.10 < \nu_f(\mathbf{c}_4) = 0.40 < \nu_f(\mathbf{c}_5) = 0.50 <$$

$$\nu_f(\mathbf{c}_{10}) = 0.86 < \nu_f(\mathbf{c}_{11}) = 0.87 < \nu_f(\mathbf{c}_{14}) = 0.90 < \nu_f(\mathbf{c}_{15}) = 1.$$

Note agora que f_b não é um codificador de Bayes relativo ao decodificador a_H e a função de valor ν: por exemplo, não é verdade que

$$\nu_{f_h}(\mathbf{c}_0) < \nu_{f_h}(\mathbf{c}_3) < \nu_{f_h}(\mathbf{c}_2).$$

Relativo ao codificador de canal f_b,

$$\mathbb{E}_{\mathcal{H}(3)}\left(f_b, a_H, \nu\right) = z\left(s\right)\left(16s^7 + 112s^6 + 245.76s^5 + 345.60s^4 + 393.60s^3 + 197.76s^2\right)$$

e

$$\mathbb{E}_{\mathcal{H}(3)}\left(f_b, a_P, \nu\right) = z\left(s\right)\left(16s^7 + 103.17s^6 + 291.65s^5 + 420.29s^4 + \right.$$
$$\left. 323.01s^3 + 131.90s^2 + 24.70s\right).$$

Consequentemente, a diferença das perdas esperadas é dada por

$$\mathbb{E}_{\mathcal{H}(3)}\left(f_b, a_H, a_P, \nu\right) = z\left(s\right)\left(8.83s^6 - 45.89s^5 - 74.69s^4 + 70.59s^3 + 65.86s^2 - 24.70s\right).$$

A Figura 4.6 abaixo mostra que para todo $s > 0.31$ (equivalentemente, para todo $p > 0.237$) o decodificador a_P é melhor do que o decodificador a_H.

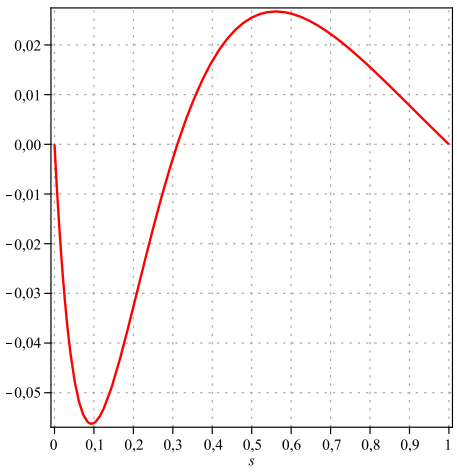

Figura 4.6: $\mathbb{E}_{\mathcal{H}(3)}\left(f_b, a_H, a_P, \nu\right)$ em função de s.

As perdas esperadas totais

$$\mathbb{E}_{\mathcal{H}(3)}\left(f_h, a_H, \nu\right), \mathbb{E}_{\mathcal{H}(3)}\left(f_h, a_P, \nu\right)$$

(Figura 4.7) e

$$\mathbb{E}_{\mathcal{H}(3)}\left(f_b, a_H, \nu\right), \mathbb{E}_{\mathcal{H}(3)}\left(f_b, a_P, \nu\right)$$

(Figura 4.8) estão ilustradas nas figuras abaixo: em preto para a_H e em vermelho para a_P. Praticamente não há diferença entre $\mathbb{E}_{\mathcal{H}(3)}(f_h, a_H, \nu)$ e $\mathbb{E}_{\mathcal{H}(3)}(f_b, a_H, \nu)$ (gráficos em preto). Em azul as probabilidades de erro de decodificações: (—) com a_H; (···) com a_P.

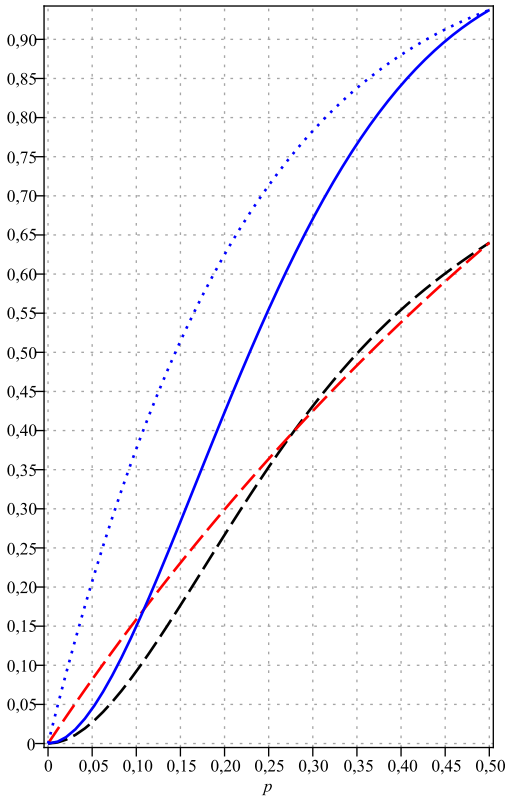

Figura 4.7: Perdas esperadas com f_h e probabilidades de erro de decodificação.

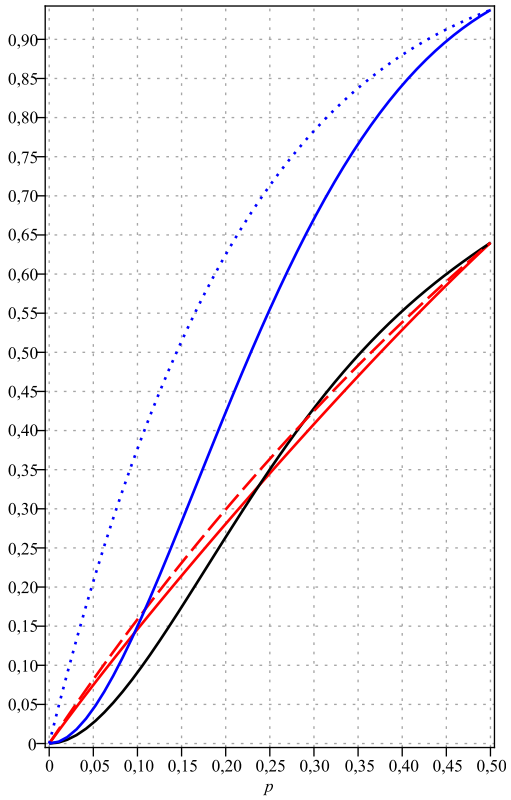

Figura 4.8: Perdas esperadas com f_b e probabilidades de erro de decodificação.

Ainda nos resta exibir um codificador de Bayes relativo ao decodificador a_H e a função de valor ν. Para tanto considere o codificador de canal $f_{\tilde{b}}$ dado pela Figura 4.9 abaixo:

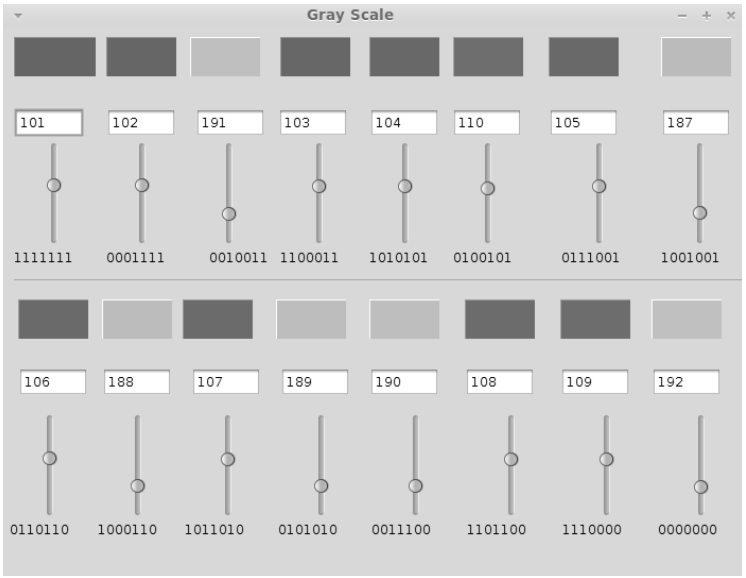

Figura 4.9: Codificador de canal $f_{\tilde{b}}$.

Temos que

$$\nu_{f_{\tilde{b}}}(\mathbf{c}_1) = 0.83, \nu_{f_{\tilde{b}}}(\mathbf{c}_3) = 0.20, \nu_{f_{\tilde{b}}}(\mathbf{c}_4) = 0.30, \nu_{f_{\tilde{b}}}(\mathbf{c}_6) = 0.40,$$

$$\nu_{f_{\tilde{b}}}(\mathbf{c}_8) = 0.50, \nu_{f_{\tilde{b}}}(\mathbf{c}_{10}) = 0.82, \nu_{f_{\tilde{b}}}(\mathbf{c}_{13}) = 0.10$$

e

$$\nu_{f_{\tilde{b}}}(\mathbf{c}_2) = 0.84, \nu_{f_{\tilde{b}}}(\mathbf{c}_5) = 0.85, \nu_{f_{\tilde{b}}}(\mathbf{c}_7) = 0.86, \nu_{f_{\tilde{b}}}(\mathbf{c}_9) = 0.87,$$

$$\nu_{f_{\tilde{b}}}(\mathbf{c}_{11}) = 0.88, \nu_{f_{\tilde{b}}}(\mathbf{c}_{12}) = 0.89, \nu_{f_{\tilde{b}}}(\mathbf{c}_{14}) = 0.90.$$

Daí que

$$\nu_{f_{\tilde{b}}}(\mathbf{c}_0) = 0 < \nu_{f_{\tilde{b}}}(\mathbf{c}') < min\left\{\nu_{f_{\tilde{b}}}(\mathbf{c}) : \mathbf{c} \in C_2\right\} = 0.84$$

para todo $\mathbf{c}' \in C_1$ e

$$max\left\{\nu_{f_{\tilde{b}}}(\mathbf{c}) : \mathbf{c} \in C_1\right\} = 0.83 < \nu_{f_{\tilde{b}}}(\mathbf{c}') < \nu_{f_{\tilde{b}}}(\mathbf{c}_{15}) = 1$$

para todo $\mathbf{c}' \in C_2$. Consequentemente, $f_{\tilde{b}}$ é um codificador de Bayes para a_H e ν. Temos que $f_{\tilde{b}}$ não é um codificador de Bayes para a_P e ν: não é verdade que

$$\nu_{f_{\tilde{b}}}(\mathbf{c}_1) < \nu_{f_{\tilde{b}}}(\mathbf{c}_4).$$

As perdas esperadas para a_H e a_P referentes a função de valor ν e ao codificador de canal $f_{\tilde{b}}$ são dadas abaixo:

$$\mathbb{E}_{\mathcal{H}(3)}\left(f_{\tilde{b}}, a_H, \nu\right) = z(s)\,(16s^7 + 112s^6 + 292.32s^5 + 299.04s^4 +$$

$$440.16s^3 + 151.20s^2)$$

e

$$\mathbb{E}_{\mathcal{H}(3)}\left(f_{\tilde{b}}, a_P, \nu\right) = z(s)\,(16s^7 + 99.30s^6 + 273.86s^5 + 401.65s^4 +$$

$$330.93s^3 + 151.85s^2 + 32.01s).$$

Neste caso temos que

$$\mathbb{E}_{\mathcal{H}(3)}\left(f_{\tilde{b}}, a_H, a_P, \nu\right) < 0$$

para todo $0 < s < 1$.

Na Figura 4.10 abaixo é possível comparar as perdas esperadas $\mathbb{E}_{\mathcal{H}(3)}(f_h, a_H, \nu)$ (- - -) e $\mathbb{E}_{\mathcal{H}(3)}(f_h, a_P, \nu)$ (- - -) com as perdas esperadas $\mathbb{E}_{\mathcal{H}(3)}\left(f_{\tilde{b}}, a_H, \nu\right)$ (• • •) e $\mathbb{E}_{\mathcal{H}(3)}\left(f_{\tilde{b}}, a_P, \nu\right)$ (○ ○ ○). Temos que

$$\mathbb{E}_{\mathcal{H}(3)}\left(f_{\tilde{b}}, a_H, \nu\right) - \mathbb{E}_{\mathcal{H}(3)}(f_h, a_P, \nu) < 0$$

para todo $0 < p < 0.5$. Já na Figura 4.11 abaixo é possível comparar as perdas esperadas $\mathbb{E}_{\mathcal{H}(3)}(f_b, a_H, \nu)$ (- - -) e $\mathbb{E}_{\mathcal{H}(3)}(f_b, a_P, \nu)$ (- - -) com as perdas esperadas $\mathbb{E}_{\mathcal{H}(3)}\left(f_{\tilde{b}}, a_H, \nu\right)$ (• • •) e $\mathbb{E}_{\mathcal{H}(3)}\left(f_{\tilde{b}}, a_P, \nu\right)$ (○ ○ ○). Neste caso temos que

$$\mathbb{E}_{\mathcal{H}(3)}\left(f_{\tilde{b}}, a_H, \nu\right) - \mathbb{E}_{\mathcal{H}(3)}(f_b, a_P, \nu) > 0$$

somente se $0.345 < p < 0.456$ (valores aproximados).

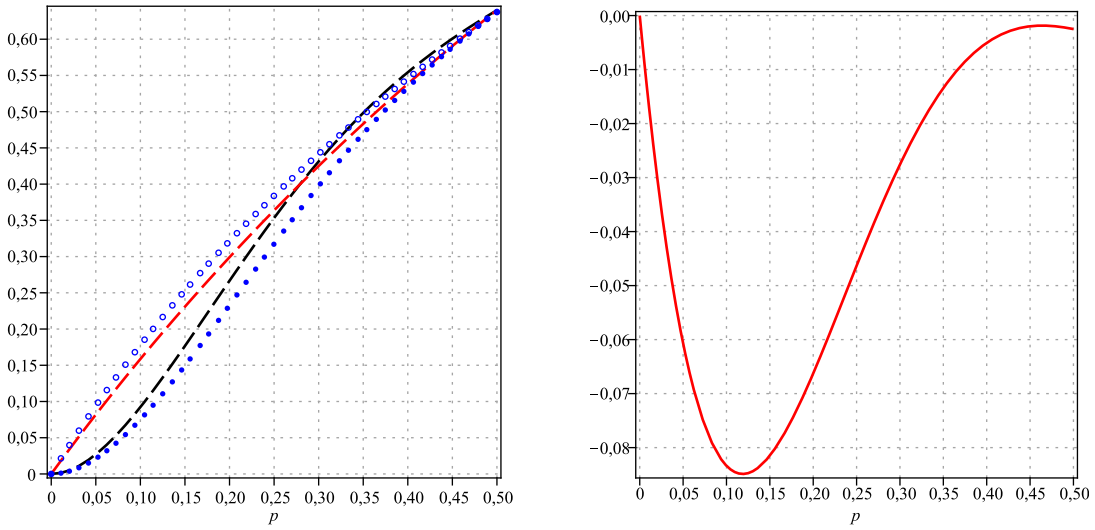

Figura 4.10: À esquerda, perdas esperadas com f_h e $f_{\widetilde{b}}$. À direita, a diferença $\mathbb{E}_{\mathcal{H}(3)}\left(f_{\widetilde{b}}, a_H, \nu\right) - \mathbb{E}_{\mathcal{H}(3)}\left(f_h, a_P, \nu\right)$.

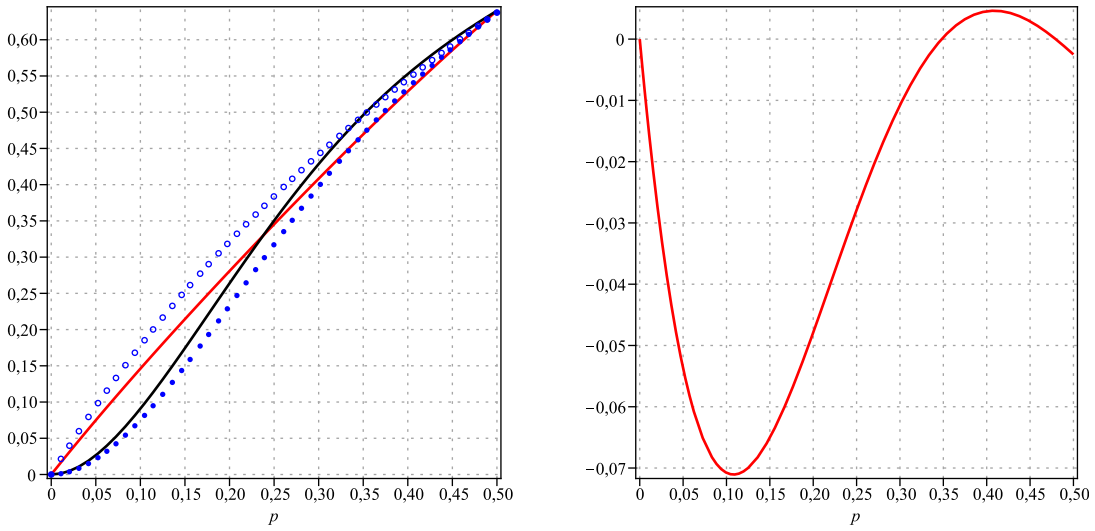

Figura 4.11: À esquerda, perdas esperadas com f_b e $f_{\widetilde{b}}$. À direita, a diferença $\mathbb{E}_{\mathcal{H}(3)}\left(f_{\widetilde{b}}, a_H, \nu\right) - \mathbb{E}_{\mathcal{H}(3)}\left(f_b, a_P, \nu\right)$.

4.6 Simulações

Encerramos este capítulo simulando diversas transmissões da imagem "Hello World" por um canal binário simétrico variando a probabilidade de erro, os codificadores de canais e os decodificadores.

As imagens serão apresentadas em blocos de seis imagens, cada bloco relativo a uma probabilidade de erro p (Figuras 4.12 a 4.16): no topo temos as imagens codificadas com f_h; no centro temos as imagens codificadas com f_b; embaixo temos as imagens codificadas com $f_{\tilde{b}}$. À esquerda estão as imagens decodificadas com a_H. À direita estão as imagens decodificadas com a_P. Na sequência são apresentadas as *imagens dos erros de decodificação* (Figuras 4.17 a 4.21) referentes a cada uma das simulações contidas nas Figuras 4.12 a 4.16: os pixels decodificados corretamente estarão representados pela cor púrpura; os pixels decodificados incorretamente aparecerão na cor decodificada.

Como veremos, o melhor desempenho do codificador f_b em relação ao codificador f_h, referente ao decodificador a_P, será nitidamente evidenciado pelas imagens geradas com p pequeno: com f_h apenas o tom escuro é bem protegido; com f_b ambos os tons estão bem protegidos. Mesmo não sendo f_h um codificador de Bayes para a_P, as imagens geradas com $p = 0.30, 0.40, 0.43$ apresentam uma melhor percepção quando comparadas com as imagens dadas por a_H. Já as imagens produzidas com o codificador de canal $f_{\tilde{b}}$ e decodificadas com a_H são superiores em qualidade as demais imagens até $p = 0.30$. Para $p = 0.40, 0.43$ as imagens produzidas com o codificador de canal f_b e decodificadas com a_P apresentam uma ligeira superioridade na qualidade.

Como o decodificador ML minimiza a quantidade de erros, já é previsível que para este decodificador as imagens dos erros estejam mais pinceladas pela cor púrpura. Com os codificadores de Bayes somos capazes de prever os pixels representados pela cor púrpura, já que os erros de decodificação geram uma imagem próxima da original. As imagens também evidenciam nossa percepção de que os decodificadores P-NN juntamente com os codificadores heurísticos são compatíveis com a idéia de informação com valor, protegendo as informações de maior valor com vizinhanças de informações com valores similares.

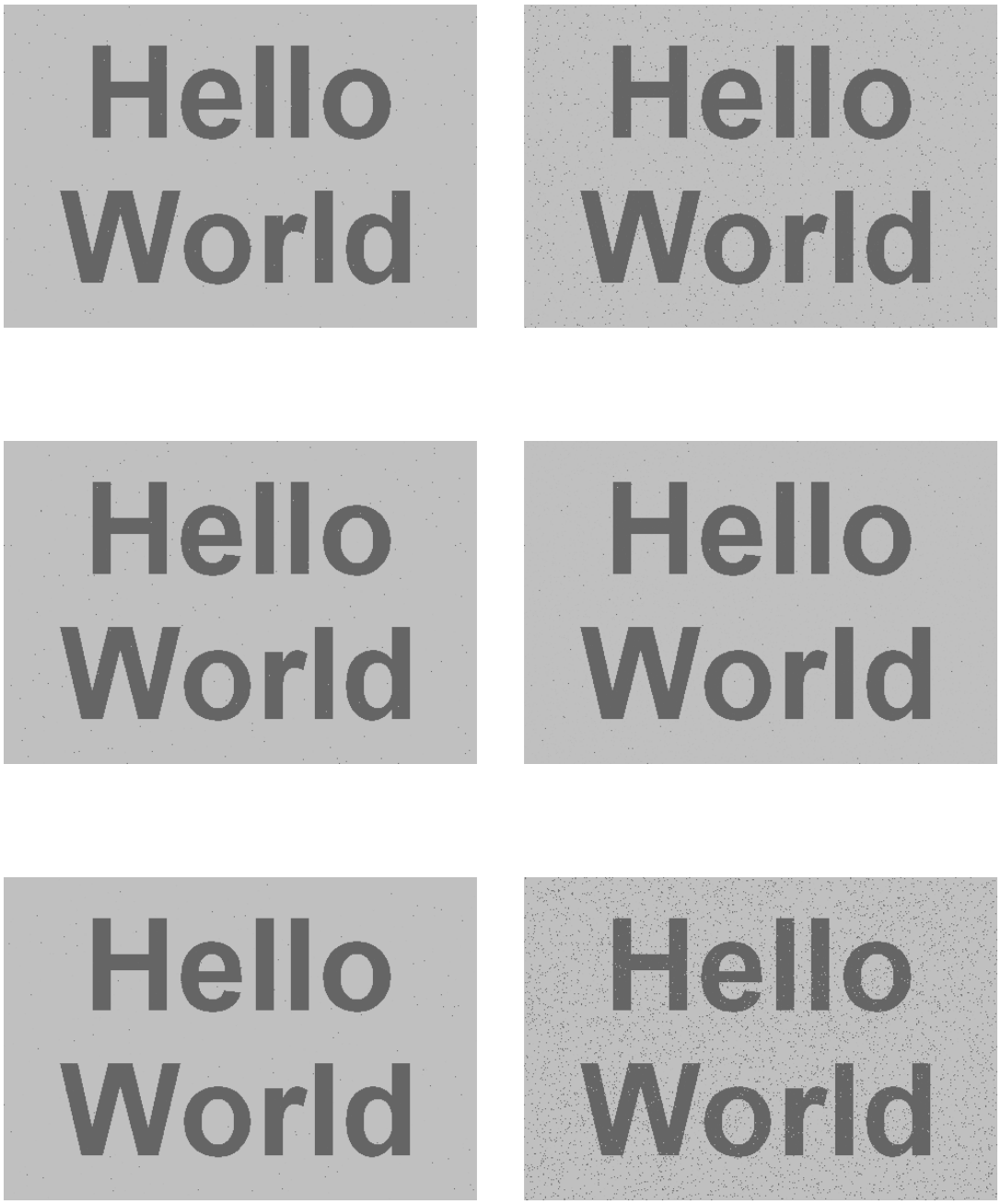

Figura 4.12: Simulações com probabilidade de erro $p = 0.01$ usando os codificadores de canal f_h (acima), f_b (no centro) e $f_{\tilde{b}}$ (embaixo): à esquerda imagens decodificadas com a_H; à direita imagens decodificadas com a_P.

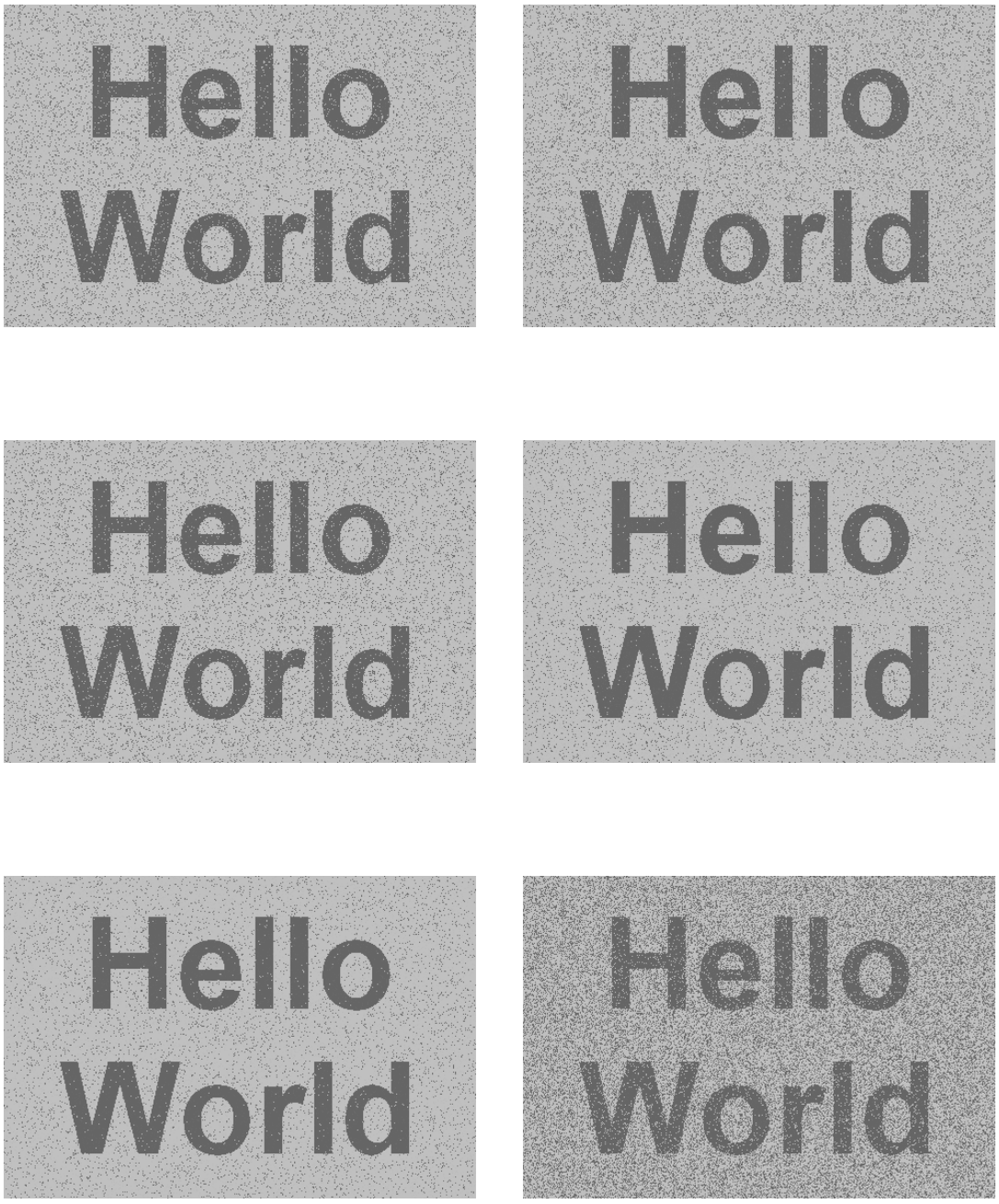

Figura 4.13: Simulações com probabilidade de erro $p = 0.1$ usando os codificadores de canal f_h (acima), f_b (no centro) e $f_{\tilde{b}}$ (embaixo): à esquerda imagens decodificadas com a_H; à direita imagens decodificadas com a_P.

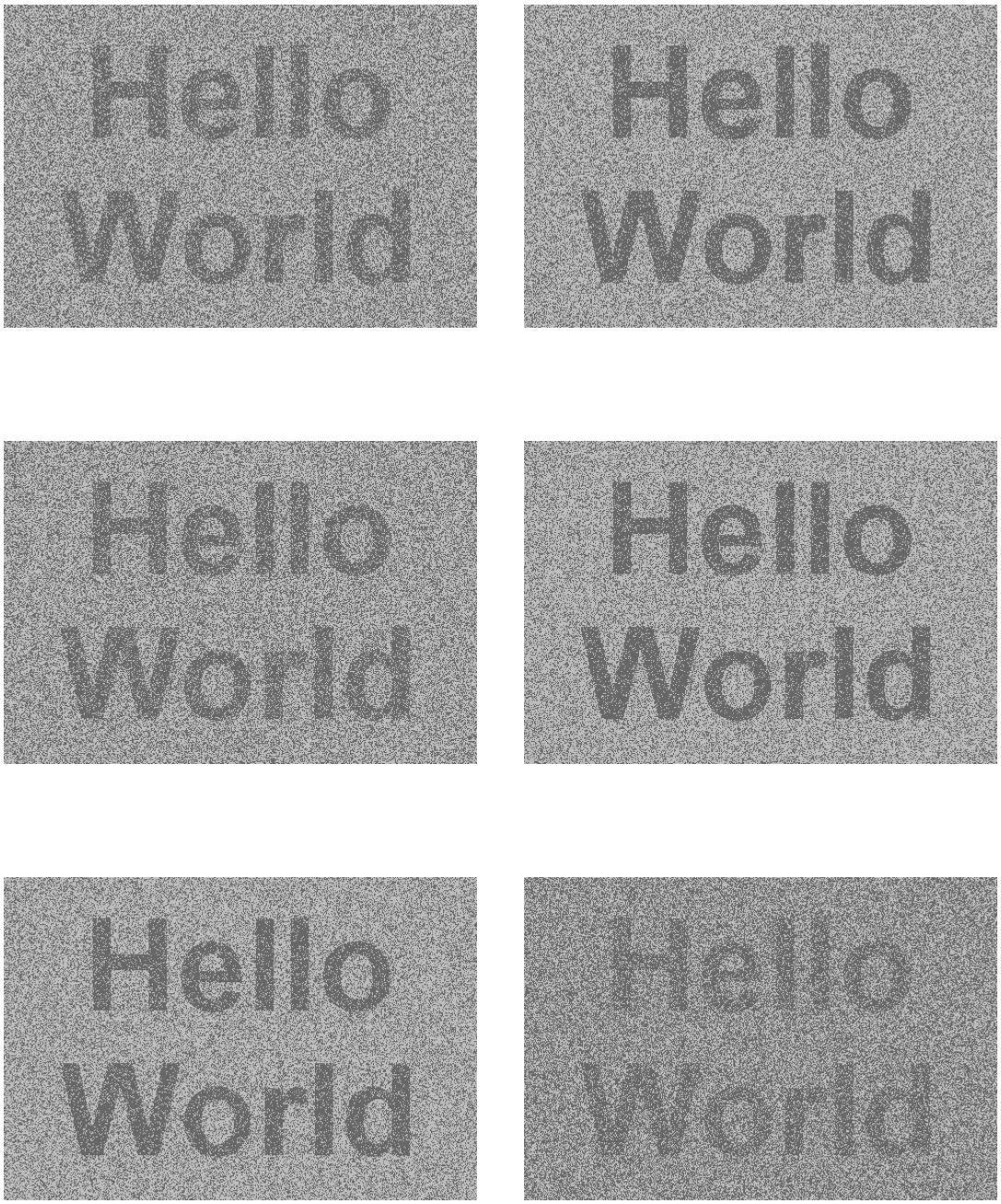

Figura 4.14: Simulações com probabilidade de erro $p = 0.3$ usando os codificadores de canal f_h (acima), f_b (no centro) e $f_{\tilde{b}}$ (embaixo): à esquerda imagens decodificadas com a_H; à direita imagens decodificadas com a_P.

Figura 4.15: Simulações com probabilidade de erro $p = 0.4$ usando os codificadores de canal f_h (acima), f_b (no centro) e $f_{\tilde{b}}$ (embaixo): à esquerda imagens decodificadas com a_H; à direita imagens decodificadas com a_P.

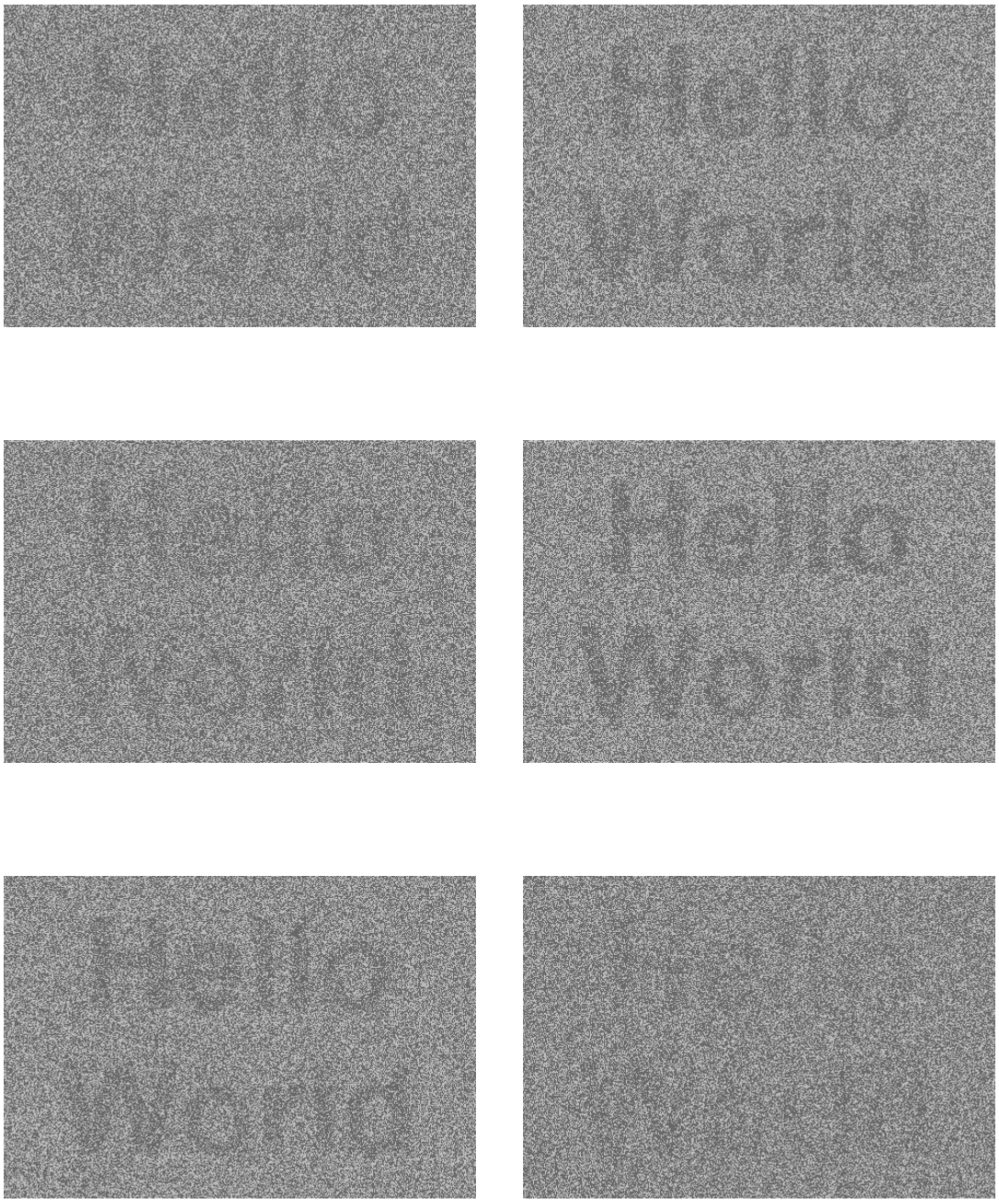

Figura 4.16: Simulações com probabilidade de erro $p = 0.43$ usando os codificadores de canal f_h (acima), f_b (no centro) e $f_{\tilde{b}}$ (embaixo): à esquerda imagens decodificadas com a_H; à direita imagens decodificadas com a_P.

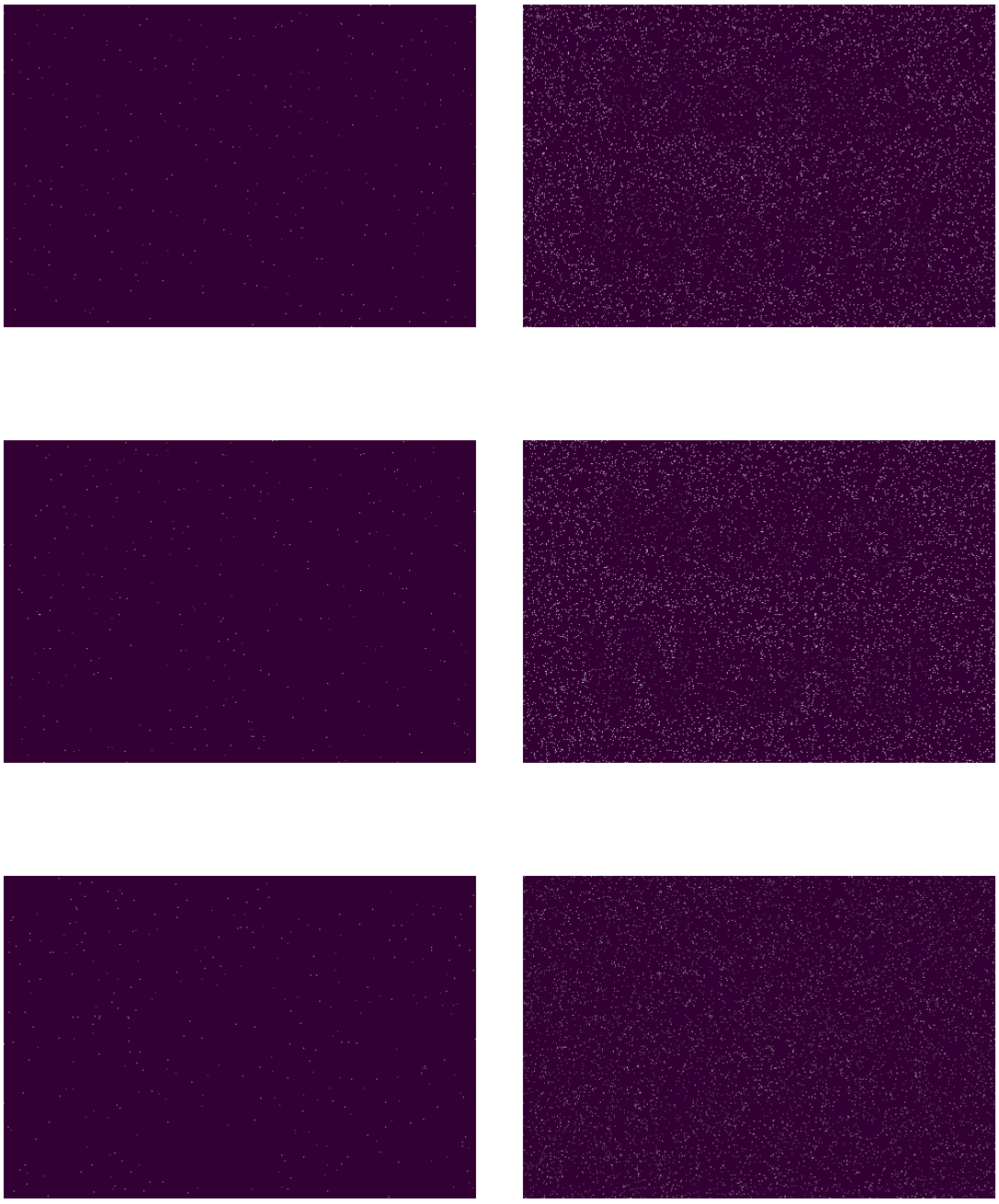

Figura 4.17: Imagens dos erros. Simulações com probabilidade de erro $p = 0.01$ usando os codificadores de canal f_h (acima), f_b (no centro) e $f_{\widetilde{b}}$ (embaixo): à esquerda imagens decodificadas com a_H; à direita imagens decodificadas com a_P.

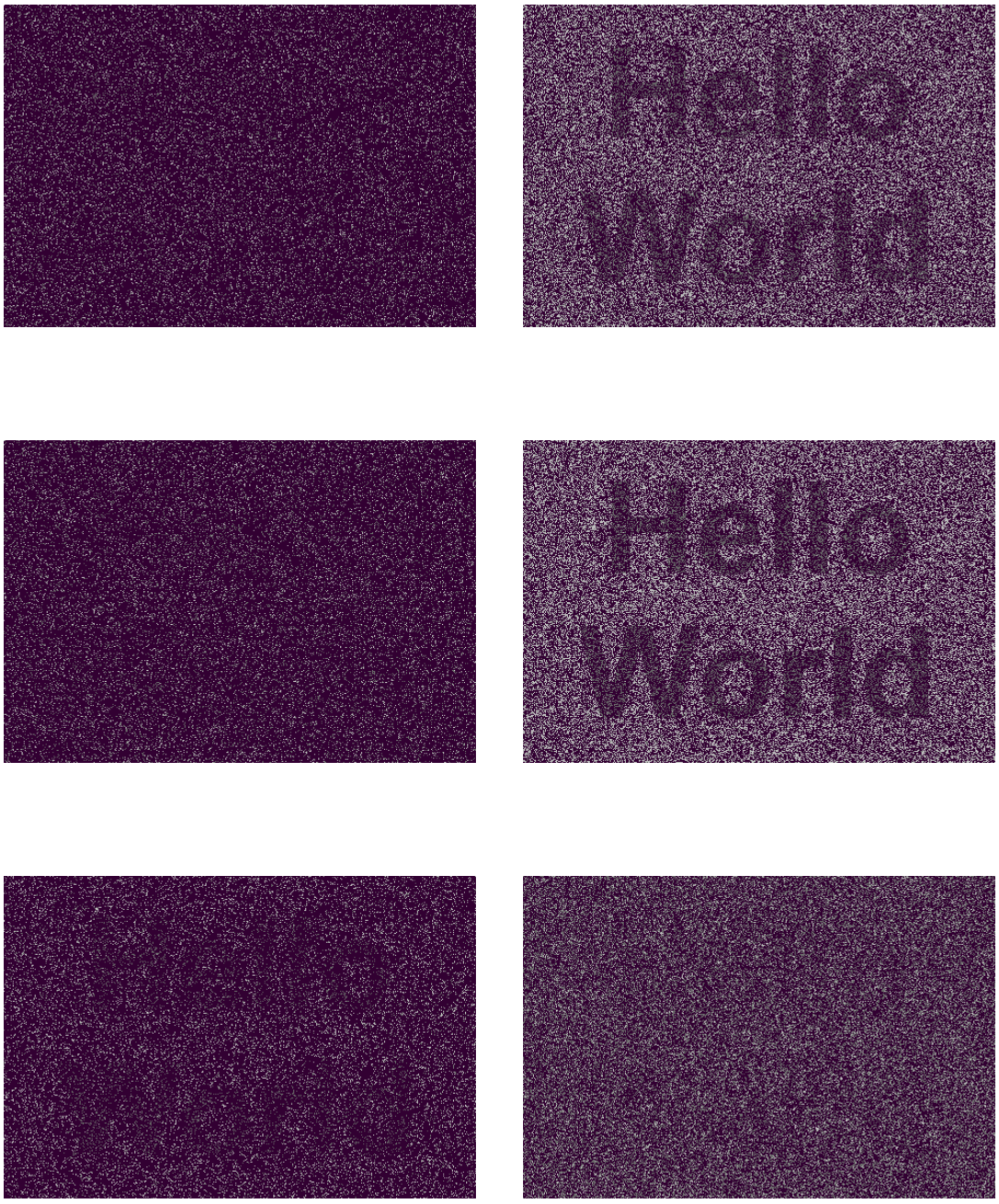

Figura 4.18: Imagens dos erros. Simulações com probabilidade de erro $p = 0.1$ usando os codificadores de canal f_h (acima), f_b (no centro) e $f_{\widetilde{b}}$ (embaixo): à esquerda imagens decodificadas com a_H; à direita imagens decodificadas com a_P.

Figura 4.19: Imagens dos erros. Simulações com probabilidade de erro $p = 0.3$ usando os codificadores de canal f_h (acima), f_b (no centro) e $f_{\widetilde{b}}$ (embaixo): à esquerda imagens decodificadas com a_H; à direita imagens decodificadas com a_P.

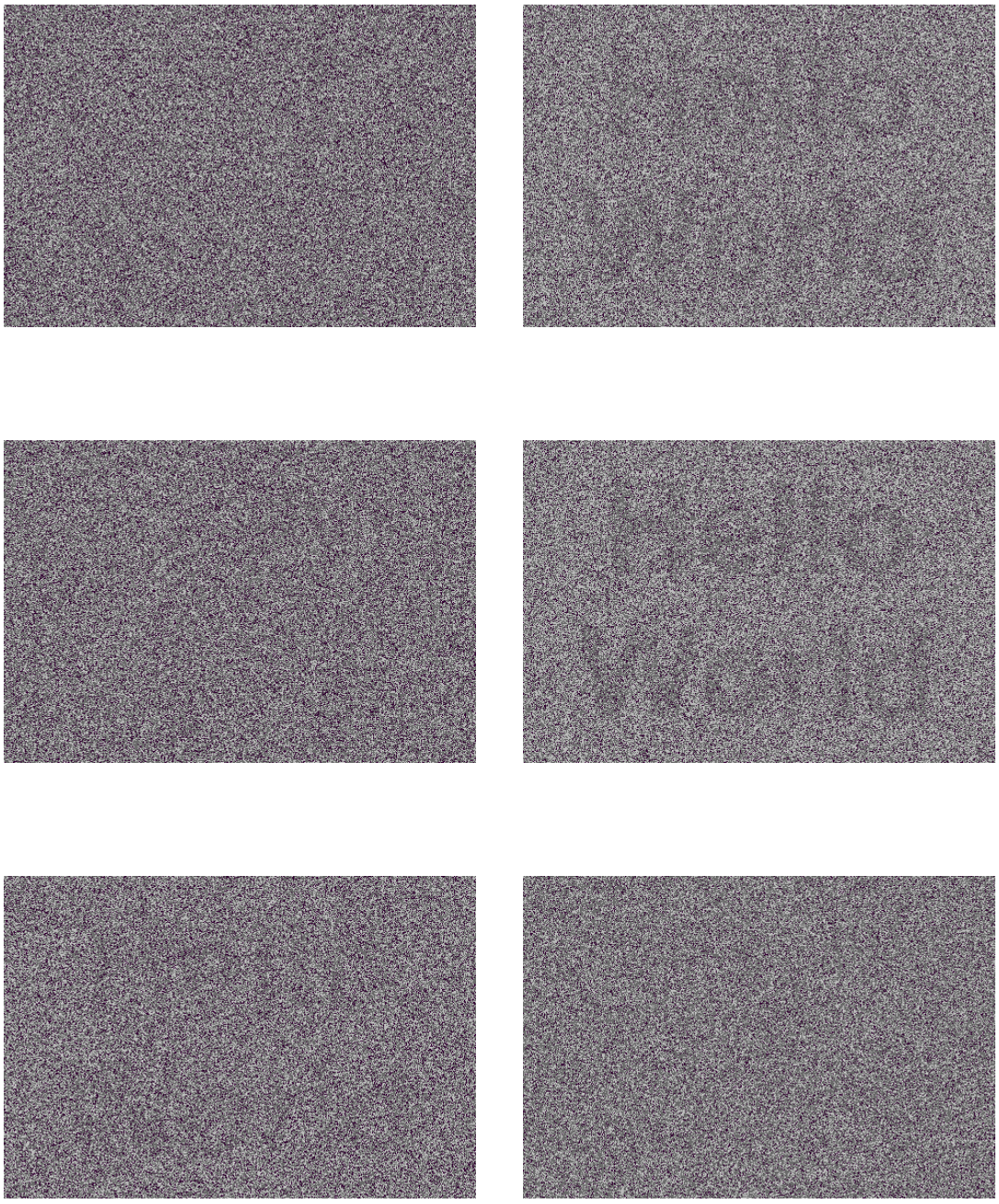

Figura 4.20: Imagens dos erros. Simulações com probabilidade de erro $p = 0.4$ usando os codificadores de canal f_h (acima), f_b (no centro) e $f_{\widetilde{b}}$ (embaixo): à esquerda imagens decodificadas com a_H; à direita imagens decodificadas com a_P.

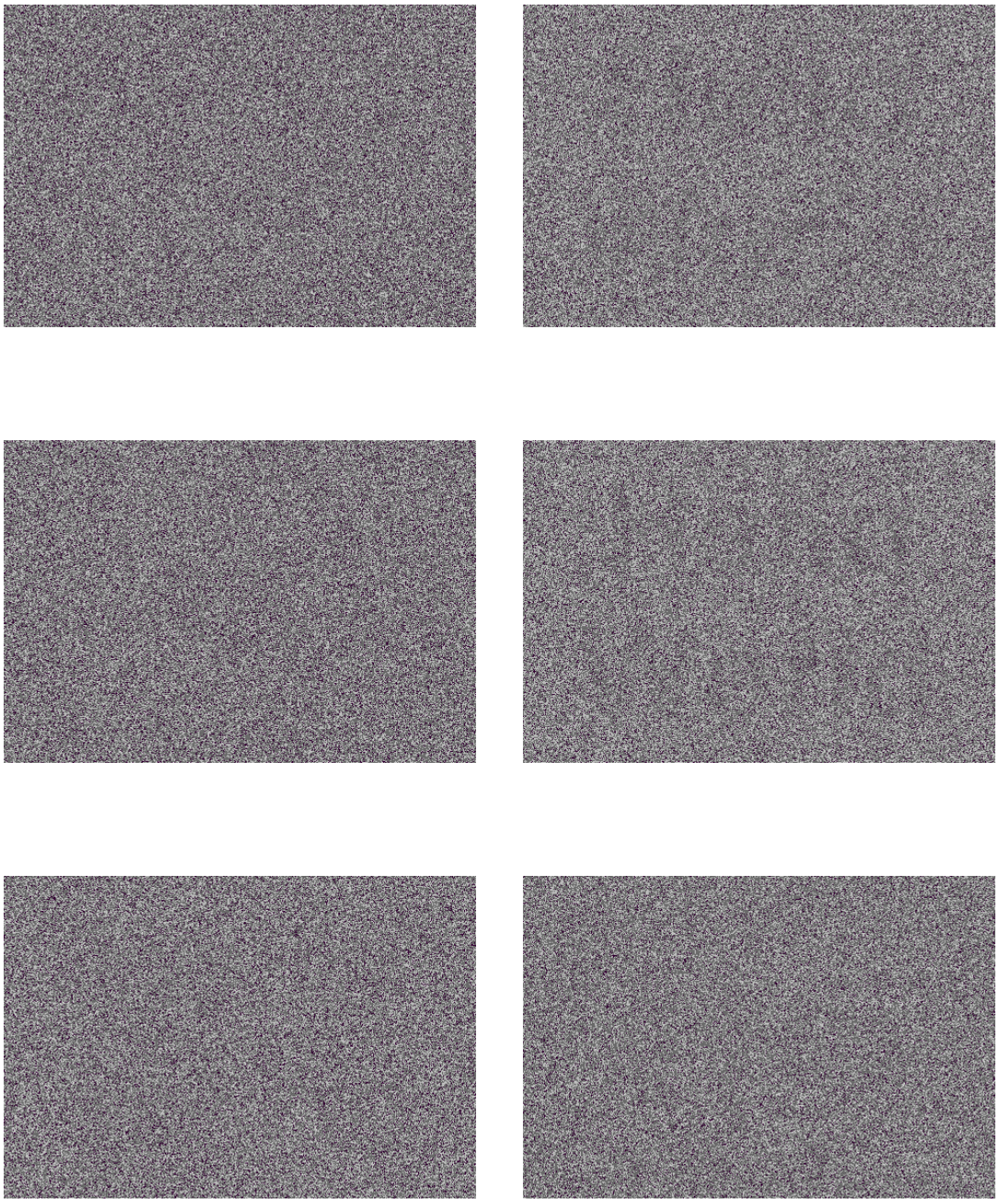

Figura 4.21: Imagens dos erros. Simulações com probabilidade de erro $p = 0.43$ usando os codificadores de canal f_h (acima), f_b (no centro) e $f_{\widetilde{b}}$ (embaixo): à esquerda imagens decodificadas com a_H; à direita imagens decodificadas com a_P.

Referências Bibliográficas

[1] J. A. Pinheiro and M. Firer - *Classification of poset-block spaces admitting MacWilliams-type identity* - IEEE-ITW, Paraty - Brasil, 2011.

[2] S. B. Z. Azami, P. Duhamel and O. Rioul - *Joint Source-Channel Coding: Panorama of Methods* - CNES Workshop on Data Compression, Toulouse - France (1996).

[3] A. Barg and P. Purkayastha - *Bounds on ordered codes and orthogonal arrays* - Moscow Mathematical Journal (2009), vol. 9, No. 2, 211-243.

[4] E. R. Berlekamp, R. J. McEliece and H. C. A. van Tilborg - *On the Inherent Intractability of Certain Coding Problems* - IEEE Transactions on Information Theory (1978), vol. 24, n. 3, 384-386.

[5] S. Borade, B Nakiboğlu and L. Zheng - *Unequal Error Protection: An Information-Theoretic Perspective* - IEEE Transactions on Information Theory (2009), vol. 55, n. 12, 5511-5539.

[6] R. Brualdi, J. S. Graves and M. Lawrence - *Codes with a poset metric* - Discrete Mathematics 147 (1995) 57-72.

[7] G. Cohen, I. Honkala, S. Litsyn and A. Lobstein - *Covering Codes* - Elsevier (1997).

[8] T. Cover - *Broadcast Channel* - IEEE Transactions on Information Theory (1972), vol. 18, 2-14.

[9] T. Cover and J. A. Thomas - *Elements of Information Theory* - John Wiley & Sons (2006).

[10] I. Csiszár - *Joint Source-Channel Error Expoent* - Problems of Control and Information Theory (1980), vol. 9, n. 5, 315-328.

[11] I. Csiszár and J. Körner - *Information Theory: Coding Theorems for Discrete Memoryless Systems* - Academic Press, New York (1981).

[12] L. V. Felix and M. Firer - *Canonical-Systematic Form of Hierarchical Codes* - Pre-print (2011).

[13] M. F. Flanagan, V. Skachek, E. Byrne and M. Greferath - *Linear-Programming Decoding of Nonbinary Linear Codes* - IEEE Transactions on Information Theory (2009), vol. 55, No. 9, 4134-4154.

[14] R. G. Gallager - *Information Theory and Reliable Communication* - Jonh Wiley and Sons (1968).

[15] M. J. E. Golay - *Notes on Digital Coding* - Proceeding of the IRE (1949).

[16] R. W. Hamming - *Coding and Information Theory* - Prentice-Hall (1980).

[17] R. W. Hamming - *Error Detecting and Error Correcting Codes* - The Bell System Technical Journal (1950), 147-160.

[18] K. Po Ho and J. M. Kahn - *Imagen Transmission over Noisy Channels Using Multicarrier Modulation* - Signal Processing: Image Comunication 9 (1997) 159-169.

[19] W. C. Huffman and V. Pless - *Fundamentals of Error-Correcting Codes* - Cambridge University Press (2003).

[20] J. Y. Hyun and H. K. Kim - *The poset structures admitting the extended binary Hamming code to be a perfect code* - Discrete Mathematics 288 (2004) 37-47.

[21] J. Y. Hyun and H. K. Kim - *Maximum distance separable poset codes* - Designs, Codes and Cryptography (2008), vol. 48, No. 3, 247-261.

[22] C. Jang, H. K. Kim, D. Y. Oh and Y. Rho - *The poset structures admitting the extended binary Golay code to be a perfect code* - Discrete Mathematics 308 (2008) 4057-4068.

[23] B. Juba and M. Sudan - *Universal semantics of Communication I* - STOC'08 Proceedings of the 40th annual ACM symposium on Theory of computing, (2008), 123-132.

[24] H. K. Kim and D. Y. Oh - *A Classification of Poset Admitting the MacWilliams Identity* - IEEE Transactions on Information Theory (2005), vol. 51, No. 4, 1424-1431.

[25] F. J. MacWilliams and N. J. A. Sloane - *The Theory of Error-Correcting Codes* - North-Holland (1977).

[26] B. Masnick and J. Wolf - *On linear unequal error protection codes* - IEEE Transactions on Information Theory (1967), vol. 3, No. 4, 600-607.

[27] R. H. Morelos-Zaragova and Shu Lin - *On a Class of Optimal Nonbinary Linear Unequal-Error-Protection Codes for Two Sets of Messages*- IEEE Transactions on Information Theory (1994), vol. 40, No. 1, 196-200.

[28] A. de Oliveira Moura and M. Firer - *Duality for poset codes* - IEEE Transactions on Information Theory (2010), vol. 56, No. 7, 3180-3186.

[29] H. Niederreiter - *Point sets and sequences with small discrepancy* - Monatshefte für Mathematik 104 (1987), 273-337.

[30] L. Panek, M. Firer - *Códigos e Métricas* - IV Bienal da Sociedade Brasileira de Matemática, Maringá-UEM (2008).

[31] L. Panek, M. Firer, H. Kwang Kim and J. Yoon Hyun - *Groups of linear isometries on poset structures* - Discrete Mathematics 308 (2008) 4116-4123.

[32] L. Panek, M. Firer and M. M. S. Alves - *Symmetry groups of Rosenbloom-Tsfasman spaces* - Discrete Mathematics 309 (2009) 763-771.

[33] L. Panek, M. Firer and M. M. S. Alves - *Classification of Niederreiter-Rosenbloom-Tsfasman Block Codes* - IEEE Transactions on Information Theory (2010), vol. 56, No. 10, 5207-5216.

[34] L. Panek, M. M. S. Alves, M. Firer - *Raio de Empacotamento e Limitante de Hamming sobre os Espaços de Rosenbloom-Tsfasman* - XXX CNMAC (2007).

[35] K. Ramchandran, A. Ortega, K. Metin Uz and M. Vetterli - *Multiresolution Broadcast for Digital HDTV Using Joint Source-Channel Coding* - IEEE Journal on Selected Areas in Communications, vol. 11, No. 1, 6-23.

[36] K. Yang, Xiaodong Wang and J. Feldman - *A New Linear Programming Approach to Decoding Linear Block Codes* - IEEE Transactions on Information Theory (2008), vol. 54, No. 3, 1061-1072.

[37] M. Yu Rosenbloom and M. A. Tsfasman - *Codes for the m-metric* - Problems of Information Transmission 33 (1997) 45-52.

[38] I. Sason and S. Shamai - *Performance Analysis of Linear Codes under Maximum-Likelihood Decoding: A Tutorial* - Foundations and Trends in Communication and Information Theory, vol. 3, No. 1-2 (2006) 1-222.

[39] C. E. Shannon - *A mathematical theory of communication* - The Bell System Technical Journal 27 (1948), 379-423.

[40] R. P. Stanley - *Enumerative Combinatorics* - Volume I, Cambridge University Press (1999).

[41] S. Tavildar and P. Viswanath - *Approximately universal codes over slow-fading channels* - IEEE Transactions on Information Theory (2006), vol. 52, No. 7, 3233-3258.

[42] Chung-Hsuan Wang, Mao-Ching Chiu and Chio-chao Chao - *On Unequal Error Protection of Convolutional Codes From an Algebraic Perspective* - IEEE Transactions on Information Theory (2010), vol. 56, No. 1, 296-315.

[43] A. D. Wyner - *Capabilities of Bounded Discrepancy Decoding* - The Bell System Technical Journal (1965), July-August.